Giant virus biology and diversity in the era of genome-resolved metagenomics

Frederik Schulz✉1,2, Chantal Abergel2 and Tanja Woyke1,3✉✉

Abstract | The discovery of giant viruses, with capsids as large as some bacteria, megabase-range genomes and a variety of traits typically found only in cellular organisms, was one of the most remarkable breakthroughs in biology. Until recently, most of our knowledge of giant viruses came from ~100 species-level isolates for which genome sequences were available. However, these isolates were primarily derived from laboratory-based co-cultivation with few cultured protists and algae and, thus, did not reflect the true diversity of giant viruses. Although virus co-cultures enabled valuable insights into giant virus biology, many questions regarding their origin, evolution and ecological importance remain unanswered. With advances in sequencing technologies and bioinformatics, our understanding of giant viruses has drastically expanded. In this Review, we summarize our understanding of giant virus diversity and biology based on viral isolates as laboratory cultivation has enabled extensive insights into viral morphology and infection strategies. We then explore how cultivation-independent approaches have heightened our understanding of the coding potential and diversity of the Nucleocytoviricota. We discuss how metagenomics has revolutionized our perspective of giant viruses by revealing their distribution across our planet’s biomes, where they impact the biology and ecology of a wide range of eukaryotic hosts and ultimately affect global nutrient cycles.

Large and giant viruses are part of a group of double-stranded DNA viruses, the nucleocyttoplasmic large DNA viruses (NCLDVs)1–4, which constitutes the viral phylum Nucleocytoviricota1. Viruses of this phylum infect a wide range of eukaryotic hosts, from the tiniest known unicellular choanoflagellates to multicellular animals1. NCLDVs typically replicate in so-called viral factories built in the host cytoplasm or use the host nucleus to replicate and sometimes assemble their progeny5,6. Hallmark features of these viruses are large genomes ranging from 70 kb to up to 2.5 Mb and virions that can reach more than 2 μm in length1. The term ‘giant virus’ was initially coined in the 1990s, when it became apparent that viruses that infect algae have unusually large genomes1 and, further, in the early 2000s, when the first virus with a genome in the megabase range was discovered; initial light microscopy observations led to the assumption that its particles corresponded to a Gram-positive bacterial pathogen of amoebae7,8. More detailed ultrastructural analyses revealed a typical icosahedral-shaped virion and genome sequencing yielded a 1.2 Mb viral genome9. This virus was named ‘minivirus’, short for ‘microbe-mimicking virus’, and represented an unexpected novelty in the virosphere, due not only to its exceptional particle and genome sizes but also to its coding potential as it includes several genes with possible roles in protein biosynthesis10. Since this discovery of giant viruses, their coding potential has been full of surprises, and the presence of hallmark genes of cellular life led to the hypothesis that these viruses might represent an enigmatic fourth domain of life11–13. Equally intriguing, much smaller viruses (so-called virophages) were found to infect some NCLDVs that have exclusively cytoplasmic infectious cycles; virophages parasitize and sometimes kill their hosts14. Also discovered was a third partner coined ‘transpoviron’, which corresponds to a 7 kb double-stranded DNA episome that is able to propagate using both the giant virus and the virophage particles as vehicles15,16.

For well over a decade, giant viruses had chiefly been studied through cultivation-based approaches until very recently, when virology followed the footsteps of microbial genomics by applying cultivation-independent metagenomics to investigate the evolutionary diversity and metabolic potential of these viruses at an unparalleled pace. In this Review, we explore a wealth of experimental data that has revealed many insights into giant virus biology, in particular their virion structure and distinctive infection strategies. We build upon this knowledge by integrating the latest sequence-based studies that expanded NCLDV diversity, biogeography, coding potential and putative host range. Furthermore,
Giant virus discovery through isolation

The earliest discovered NCLDVs were the Poxviridae, which include the causative agent of smallpox and were the first viral particles seen under a microscope more than 130 years ago\(^1\). Large viruses that infect Chlorella green algae were isolated in the 1980s. The first genomes of Vaccinia virus (a poxvirus) and Paramecium bursaria chlorella virus 1 (PBCV1) were sequenced in the early 1990s\(^1\) and 1999 (REF \(^1\)), respectively. Shortly thereafter, additional genomes of Poxviridae were sequenced (FIG. 1), with sizes ranging from 120 kb to 360 kb (REF \(^1\)). Subsequently, other viruses that infect animals, including members of the Ascoviridae, Iridoviridae and Asfarviridae families, were found and their genomes sequenced\(^2\)–\(^11\). Genomes of viruses in these groups are comparably small (up to 220 kb) and even smaller in the recently discovered shrimp-associated Mininucleoviridae (70–80 kb)\(^12\). In addition to animal-infecting NCLDVs, a wide range of NCLDVs were detected in various eukaryotic algae, including chlorophytes, haptophytes, pelagophytes, brown algae and dinoflagellates in the early 2000s\(^13\). These algae-associated NCLDVs were classified as Phycodnaviridae\(^14\) and Mesomimiviridae\(^15\) and, although most of their genomes are ~200–500 kb (REFS \(^24\), \(^25\)), the genomes of Tetrasymsa kapha virus RF01 are 668 kb (REF \(^26\)) and 1.4 Mb (REF \(^27\)), respectively.

After the discovery of mimivirus in 2003 (REF \(^28\)), other NCLDVs with larger virions and genomes above 500 kb have been found to infect heterotrophic protists\(^28\) (mainly members of the Amoebozoa). For more than a decade, Acanthamoeba strains had chiefly been used as hosts for the co-cultivation of new viruses, leading to the frequent isolation of closely related giant viruses able to infect this unicellular host\(^29\). Acanthamoeba spp. has proven to be...
a particularly suitable host for many *Meganimivirinae* and *Marseillleviridae*. Consequently, viruses from these taxonomic groups are currently among the most commonly cultivated NCLDV with more than 30 genome sequences readily available in public databases, including the novel *Meganimivirinae* lineages *tupanvirus* and *cotonovirus*. The co-cultivation approach has been widely successful and also led to the recovery of isolates from divergent NCLDV clades, facilitating the organization and naming of pithoviruses, pandoraviruses, molliviruses and medusaviruses. Recently, the use of alternative hosts, such as *Vernamoeba* spp., has led to the co-cultivation of several new faustovirus isolates, orpheovirus, pacmanvirus and kaumoebavirus—all distant relatives of pithovirus, marseilllevirus and asfarvirus. A newly developed high-throughput co-cultivation-based approach using high-content screening microscopy has proven a valuable tool for giant virus discovery and isolation. Yet, co-cultivation is limited by host specificity of giant viruses; some NCLDV lineages are able to infect only specific hosts, such as certain species of *Acanthamoeba*, whereas others may be more versatile, exhibiting a broader host range. Considering the enormous diversity of eukaryotes, and in particular of microeukaryotes, it is likely that giant viruses that have been recovered through isolation reflect only a minute fraction of NCLDV lineages extant in the wild.

**Virion structures and infection strategies**

**Viruses with nucleocytoplasmic infectious cycles.** Chloroviruses were the first viruses designated as ‘giant viruses’ owing to their largeicosahedral virions of 190 nm in diameter (T number 169) and genomes of up to 370 kb (Table 1). In particular, PBCV1 was extensively studied; its capsids have a few external fibres extending from some of the capsomeres and a spike-like structure present at one vertex to anchor onto the host cell membrane. The capsids are glycosylated with an unusual oligosaccharide synthesized by the virus-encoded glycosylation machinery; the oligosaccharide is N-linked to asparagines in atypical sequons in the major capsid protein (MCP; Vp54). The outer capsid layer covers a single lipid membrane, which is essential for infectivity. Chloroviruses deliver their genome into their algal host by creating a hole in the cell wall using a virus-encoded enzyme packaged in the virion. The viral internal membrane then fuses with the host plasma membrane, forming a channel through which the genome and some viral proteins enter the cell. Because the virus does not encode an RNA polymerase, the incoming genome must be transcribed inside the host cell’s nucleus prior to virion assembly in the cytoplasm. Virions are released after host cell lysis.

Other *Nucleocytoviricota* viruses that infect algae constitute small virions. Among the smallest members of the *Nucleocytoviricota* are prasinoviruses with virion diameters of ~120 nm and genomes up to 410 kb. Being small is crucial for infecting and replicating within *Ostreococcus tauri*, which is one of the smallest free-living eukaryotes with cells only 0.8 μm in size. Following viral infection, the genome is released into the nucleus and its replication begins almost immediately. Within hours, new virions assemble in the cytoplasm and, in less than 24 h, host lysis occurs. The host cell nucleus, mitochondrion and chloroplast remain intact throughout this period.

**Viruses with exclusively cytoplasmic infectious cycles.** The second most studied virus after PBCV1 is that of the amoeba-infecting mimivirus. The ~700 nm virions are made of an icosahedral capsid ~500 nm in diameter, covered by spherical-headed spikes extending from each capsomer, and have a lipid membrane that surrounds the capsid interior. A low-resolution structure was determined by cryo-EM, which returned a T number of 277. The mechanism of entry and egress of the mimivirus virion from its host has yet to be determined. After uptake into the host cytoplasm, its DNA is replicated in the host nucleus and virions assemble in the cytoplasm (Fig. 2).

**ORFans**

Predicted genes without detectable homologues in public databases.

**Molliviruses**

Viruses with exclusively cytoplasmic infectious cycles. The second most studied virus after PBCV1 is that of the amoeba-infecting mimivirus. The ~700 nm virions are made of an icosahedral capsid ~500 nm in diameter, covered by spherical-headed spikes extending from each capsomer, and have a lipid membrane that surrounds the capsid interior. A low-resolution structure was determined by cryo-EM, which returned a T number of 277. The mechanism of entry and egress of the mimivirus virion from its host has yet to be determined. After uptake into the host cytoplasm, its DNA is replicated in the host nucleus and virions assemble in the cytoplasm (Fig. 2).
microscopy provided additional insights into virion composition, further underlining the complexity of the capsid. There are two internal lipid membranes, one lining the capsid and the other in the nucleoid compartment, which contains the genome and hundreds of proteins, including RNA polymerase and transcript maturation machinery. It has been proposed that the non-structural proteins in the nucleoid are required.

**Fig. 2 | Giant virus infection mechanisms and virion structures.** A | Giant viruses enter the host by attachment to the host cell envelope followed either by endocytosis uptake (part a) or membrane fusion after capsid opening (part b). Giant virus transcription is then initiated in the cytoplasm or viral factory (part c; purple arrow) or the host nucleus (part d; green arrow). In the periphery of the cytoplasmic viral factory, genome replication and assembly of new virions then occurs (part e) or newly synthesized virions are scattered in a large cytoplasmic viral factory (part f). Finally, virions are released after host cell lysis (part g), fusion of virion-containing vacuoles with host cell membrane (part h) or exocytosis of membrane-bound virions (part i). Small coloured circles indicate viral genome and viral proteins. B | Infection strategies of selected giant viruses. C | Transmission electron micrographs of ultrathin sections of non-icosahedral viruses embedded in resin. D | Structures of isolated giant viruses resolved by cryo-electron microscopy. Note the blue coloured stargate structure on mimivirus. The scale bars in parts C and D are 100 nm. AaV, *Aureococcus anophagefferens* virus; CroV, *Cafeteria roenbergensis* virus; OtV, *Ostreococcus tauri* virus; PBCV1, *Paramecium bursaria chlorella* virus 1. Part D reprinted from REFS 67,176,177, CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). Part D reprinted with permission from REF 43, PNAS. Part D, image courtesy of M. Kazuyoshi. Part D, image courtesy of T. Klose.
Table 1 | Virion structures, genome characteristics, infection strategies and hosts of giant virus isolates

| Viral lineage | Virion type | Virion size (μm) | Genome size (Mb) | GC (%) | Specific feature | Infectious cycle | Host |
|---------------|-------------|------------------|------------------|--------|-----------------|-----------------|------|
| **Phycodnaviridae** | | | | | | | |
| Chlorovirus PBCV1 [REF.156] | T = 169 | 0.19 | 0.33<sup>a</sup> | 40 | Protruding spike, glycosylated | Assembly in host cytoplasm | Inland water green microalgae |
| Heterosigma akashiwo virus<sup>157</sup> | Icos | 0.22 | 0.27<sup>b</sup> | 30 | No RNA polymerase | Assembly in host cytoplasm | Marine microalgae |
| **Prasinoviridae** | | | | | | | |
| Micromonas pusilla virus SP1 [REF.158] | Icos | 0.12 | 0.17<sup>b</sup> | 41 | ND | Assembly in host cytoplasm | Marine green microalgae |
| **Coccolithoviridae** | | | | | | | |
| Ectocarpus siliculosus virus 1 [REF.159] | Icos | 0.15 | 0.33<sup>b</sup> | 52 | Polysaccharide synthesis genes | Assembly in host cytoplasm | Marine brown algae |
| Emiliania huxleyi virus 86 [REF.159] | T = 169 | 0.25 | 0.41<sup>b</sup> | 40 | Can be enveloped, sphingolipid metabolism | Assembly in host cytoplasm | Emiliania huxleyi |
| **Medusaviridae** | | | | | | | |
| Medusavirus<sup>13,160</sup> | T = 277 | 0.26 | 0.38<sup>b</sup> | 62 | Spikes, histones | Assembly in host cytoplasm | Acanthamoeba |
| **Pandoraviridae** | | | | | | | |
| Pandoravirus salinus<sup>161</sup> | Amp | 1 × 0.5 | 2.77<sup>c</sup> | 60 | Ostial, cellulose, hAT transposase | Nuclear | Acanthamoeba |
| Pandoravirus neocaledonia<sup>161</sup> | Amp | 1 × 0.5 | 2.0<sup>c</sup> | 58 | Ostial, cellulose, hAT transposase | Nuclear | Acanthamoeba |
| **Molliviridae** | | | | | | | |
| Mollivirus sibericum<sup>162</sup> | Sph | 0.65 | 0.65<sup>c</sup> | 62 | Ostial, cellulose | Nuclear | Acanthamoeba |
| **Megamivirinae** | | | | | | | |
| Mimivirus<sup>11</sup> | Icos | 0.5 + 0.25 | 1.18<sup>e</sup> | 28 | Stargate, 4 aaRS, glycosylated, virophage | Cytoplasmic | Acanthamoeba |
| Moumouivirus<sup>162</sup> | Icos | 0.42 + 0.2 | 1.02<sup>e</sup> | 25 | Stargate, 5 aaRS, glycosylated | Cytoplasmic | Acanthamoeba |
| Megavirus chilensis<sup>163</sup> | Icos | 0.52 + 0.15 | 1.26<sup>e</sup> | 25 | Stargate, 7 aaRS, virophage glycosylated | Cytoplasmic | Acanthamoeba |
| Tupanvirus soda lake<sup>163</sup> | Icos + tail | 0.52 + 0.15 | 1.51<sup>e</sup> | 28 | Stargate, 20 aaRS | Cytoplasmic | Acanthamoeba and Vermamoeba |
| Cotonvirus<sup>12</sup> | Icos | 0.4 + 0.2 | 1.48<sup>e</sup> | 25 | Stargate, 5 aaRS | Cytoplasmic | Acanthamoeba |
| **Klosneuvirinae** | | | | | | | |
| Bodo saltans virus<sup>164</sup> | Icos | 0.3 | 1.38<sup>e</sup> | 25 | Stargate, 2 aaRS, 2 capsid layers, fibrils | Cytoplasmic | Bodo saltans |
| Yasminevirus<sup>145</sup> | Icos | 0.3 | 2.1<sup>e</sup> | 40 | 20 aaRS | Cytoplasmic | Vermamoeba |
| Fadolivirus<sup>164</sup> | Icos | 0.3 | 1.59 | 30 | 66 tRNAs, 23 aaRS | Cytoplasmic | Vermamoeba |
| **Other Mimiviridae** | | | | | | | |
| Cafeteria roenbergensis virus<sup>154</sup> | T = 499 | 0.3 | 0.69<sup>d</sup> | 23 | 1 aaRS, virophage | Cytoplasmic | Cafeteria roenbergensis |
| **Mesomiviridae** | | | | | | | |
| Phaeocystis globosa virus<sup>165,166</sup> | Icos | 0.15 | 0.46<sup>a</sup> | 32 | UV-DNA-repair nuclease, 2 RPB2, 1 rhodopsin, episomal virophage | Cytoplasmic | Phaeocystis |
| Chrysochromulina ericina virus<sup>157</sup> | Icos | 0.16 | 0.47<sup>a</sup> | 25 | UV-DNA-repair nuclease, 2 RPB2 | Cytoplasmic | Haptolina |
| Chrysochromulina parva virus<sup>166</sup> | Icos | 0.15 | 0.43<sup>a</sup> | 25 | UV-DNA-repair nuclease, 2 RPB2, 3 virophages | Cytoplasmic | Haptolina |
| Pyramimonas orientalis virus<sup>167</sup> | Icos | 0.22 | 0.55<sup>a</sup> | 34 | UV-DNA-repair nuclease, 2 RPB2 | Cytoplasmic | Prasinophyceae |
| Pyrmenesiump kappa virus<sup>168</sup> | Icos | 0.31 | 1.42<sup>b</sup> | 23 | UV-DNA-repair nuclease, 2 RPB2, 2 aaRS, energy metabolism | Cytoplasmic | Pyrmenesiophyceae |
to initiate the viral infectious cycle, protect the virion from oxidative stress and perform early transcription55,56. Preliminary data suggest that the genome is organized in a 30-nm diameter helical nucleocapsid comprising GMC oxidoreductases, which also constitute the glycosylated fibrils of the capsid. The folded genome lines the shell of the nucleocapsid, leaving a central channel that can accommodate large proteins, including RNA polymerase59. Mimivirus enters its host by triggering phagocytosis upon adhering to the host cell membrane with its glycosylated fibrils. Once in the vacuole, a specific structure at one vertex of the icosaedron (the stargate) opens, and the membrane under the capsid is pulled out and fuses with the vacuole membrane60, allowing the transfer of the nucleoid into the host cytoplasm61,62. Early transcription begins using the virus-encoded transcription machinery, which, at first, remains confined in the nucleoid63. The accumulation of nucleic acids due to active transcription and replication leads to the size of the viral factory increasing and newly synthesized virions start budding at its periphery, recycling host cell membranes derived from the endoplasmic reticulum64 or Golgi apparatus65. The last step of virion maturation, after genome loading into the nucleoid, is the addition of the fibril layer to the capsids66, with hundreds of newly synthesized virions released after cell lysis.

Several viruses related to mimivirus have similar infectious cycles but smaller virions. Among them is *Cafeteria roenbergensis virus*, which has an icosahedral capsid of 300 nm in diameter (FIG. 2) with a lipid membrane underneath the capsid shell. Its mode of infection is not fully understood but, similar to mimivirus, a nucleoid structure in the cytoplasm and extracellular empty capsids have been observed, supporting an external opening of the capsids followed by fusion of the internal membrane with that of the cell, thus allowing the transfer of the nucleoid into the host cytoplasm. Virions contain ~150 proteins, which either make up the icosahedral capsid or are necessary to initiate the infectious cycle61. Nascent virions assemble during the late stage of infection and are released through cell lysis. The structure of the complex capsid, determined by cryo-EM, corresponds to a T number of 499 and has a possible stargate-like structure at one vertex of the icosahedron (the stargate) opens, and the membrane under the capsid is pulled out and fuses with the vacuole membrane60, allowing the transfer of the nucleoid into the host cytoplasm61,62. Similar to other known members of the *Mimiviridae*, Mimivirus replicates in its host's cytoplasm63,64. Early transcription begins using the virus-encoded transcription machinery, which, at first, remains confined in the nucleoid63. The accumulation of nucleic acids due to active transcription and replication leads to the size of the viral factory increasing and newly synthesized virions start budding at its periphery, recycling host cell membranes derived from the endoplasmic reticulum61,62 or Golgi apparatus65. The last step of virion maturation, after genome loading into the nucleoid, is the addition of the fibril layer to the capsids66, with hundreds of newly synthesized virions released after cell lysis.

Another member of the *Mimiviridae*, with a similar icosahedral capsid of 300 nm in diameter, is *Bodo saltans virus*. Its capsid appears to be made of two proteinaceous layers surrounded by 40 nm-long fibrils. A possible stargate-like structure is present at one vertex of the capsid and there are two membranes, one

---

**Table 1 (cont.)** | Virion structures, genome characteristics, infection strategies and hosts of giant virus isolates

| Viral lineage | Virion type | Virion size (μm) | Genome size (Mb) | GC (%) | Specific feature | Infectious cycle | Host |
|---------------|-------------|-----------------|-----------------|--------|-----------------|-----------------|------|
| *Mesomimiviridae* (cont.) | | | | | | | |
| *Aureococcus anophagefferens virus* | Amp | 1.5 × 0.5 | 0.71 | 35 | Cork, tegument, membrane | Cytoplasmic | *Aureococcus* |
| *Cedratvirus A11* (REF.69) | Amp | 1 × 0.5 | 2 × 4 | 43 | Two corks, tegument, membrane | Cytoplasmic | *Acanthamoeba* |
| *Orpheovirus LCC2* (REF.70) | Amp | 1 × 0.5 | 1.47 | 25 | Pandora-like ostiole, tegument, membrane | Cytoplasmic | *Vermamoeba* |
| *Asfarviridae* | | | | | | | |
| *Faustovirus E12* (REFS.41,42) | T = 277 | 2.2 | 2.6 | 36 | 2 capsids, MCP with introns | Cytoplasmic | *Vermamoeba* |
| *Pacmanivirus* | T = 64 | 0.25 | 0.39 | 34 | Internal membrane | Cytoplasmic | *Acanthamoeba* |
| *Kaumoebavirus Sc* | T = 309 | 0.26 | 0.35 | 44 | MCP with introns, gene strand bias, internal membrane | Cytoplasmic | *Vermamoeba* |

---

aaRS, aminoacyl-tRNA synthetase; Amp, Amphora; Icos, Icosahedral; MCP, major capsid protein; ND, not determined; PBCV1, *Paramaecium bursaria chlorella virus* 1; RPβ2, DNA-directed RNA polymerase II subunit 2; Sph, Spherical; T, triangulation number. aLinear, covalently closed hairpin termini. bLinear. cLinear DNA, terminal repeats. dCircular.
lining the external protein shell and one internal to the nucleoid compartment containing the genome. The infectious cycle is similar to that of mimivirus except that the host’s nuclear genome appears to be degraded. The viral factory develops at the posterior pole of the cell to fill two-thirds of the cell space, pushing aside the nucleus and organelles. Lipid vesicles are recruited for virion assembly, which takes place at one side of the viral factory, and mature virions detach after genome loading and migrate to the posterior pole of the cell. Virions are released by budding in vesicles from the host membrane after cell lysis\(^{46}\) (FIG. 2).

Some of the largest viruses that infect algae belong to the **Mimiviridae**, all of which have icosaheiral capsids with sizes ranging from 150 nm in the case of *Aureococcus anophagefferens virus* (FIG. 2, TABLE 1) to 370 nm in the recently described *Pyrenesiunm kappa virus*\(^{49}\). These viruses also build a viral factory in the host cytoplasm, but it is unknown if the transcription machinery is loaded into the capsids, allowing an entirely cytoplasmic infectious cycle.

The largest virions found in the **Nucleocytoviricota** are those of pithovirus and cedravirus (FIG. 2), which have very large amphora-shaped capsids that can be up to 2-µm long and 600-nm wide encapsidating genomes of up to 685 kb (TABLE 1). The capsids are closed by corks — one cork for pithovirus\(^{48,60}\) (FIG. 2) and two for cedravirus\(^{48}\) — that are made by proteins organized in a honeycomb array. Despite a virion morphology that closely resembles that of pandoravirus, the external tegument is different and appears to be made of parallel strips and no cellulose; the capsids appear to be coated with short sparse fibrils\(^{48,60}\). The infectious cycle proceeds, as for other amoeba-infecting viruses, by phagocytosis followed by capsid opening and membrane fusion with the phagosome\(^{48}\). For pithovirus and cedravirus, the RNA polymerase loaded in the virion starts early transcription in the cytoplasm and the host nucleus remains intact during the entire infectious cycle. During maturation, reservoirs of tegument and corks accumulate in the host cytoplasm and are used to build the new amphora-shaped virions. The nascent virions then exit the host cell either by exocytosis or upon cell lysis\(^{48,60}\).

Outside of the **Mimiviridae**, there are smaller amoeba-infecting viruses such as members of the **Marseilleviridae**, which have icosaheiral virions of ∼250 nm in diameter (FIG. 2). A recent publication and two preprints showed the cryo-EM structure of the capsid for two members of the family at various resolutions, revealing a T number of 309 and a complex capsid structure\(^{41,71,72}\) with many minor capsid proteins. Melbourneivirus and other members of the family **Marseilleviridae** are taken up by phagocytosis and then lose their icosaheiral appearance to become spherical after the disappearance of the vacuole membrane. Similar to **Megavirusviricidae**, their genome remains in the cytoplasm; however, RNA polymerase is not loaded into the virion. Instead, the nuclear proteins are recruited to the early viral factory, including the host RNA polymerase that performs early transcription\(^{19}\). The appearance of the cell nucleus changes early in infection and becomes leaky through a still-unknown mechanism triggered by viral infection. After 1 h of infection, the nucleus integrity is restored and the virus-encoded RNA polymerase performs intermediate and late transcription\(^{35}\), and icosaheiral particles assemble inside the viral factory (FIG. 2A). **Marseilleviridae** viruses encode histone doublets that form nucleaseome to pack the genome into virions\(^{35,72}\). Mature capsids can gather in large vesicles\(^{35}\) and cell lysis leads to the release of both individual virions and filled vacuoles.

As these examples illustrate, there is no shared blueprint for the structure of giant viruses and their infection mechanisms; these characteristics vary between giant virus lineages and are likely shaped by the host organisms. The host range of the experimentally characterized giant viruses is limited to a few amoeba and algae lineages representing only a minute fraction of eukaryotic diversity. Thus, we expect that many more unusual virions and infection strategies will be revealed when new viruses will be captured together with their native hosts.

**Cultivation-independent genomics**

**Sequence-inferred prevalence and diversity of giant viruses**. Many important discoveries in giant virus biology and diversity have been made through giant virus isolation and cultivation. However, such approaches are constrained by the need to satisfy optimal growth requirements in a laboratory setting and are often restricted to lytic viruses. Cultivation-independent methods have proven to be an indispensable tool to discover the genetic make-up of giant viruses from environmental samples.

In the earlier days of metagenomics, single-marker gene-based surveys (BOX 1) revealed that several viruses of the **Phycodnaviridae** and **Mimiviridae** were present in a wide range of marine metagenomes collected during the Tara Oceans and the Sargasso Sea expeditions\(^{59,60}\) and that these viruses were more abundant in the photic layer than eukaryotes\(^{58}\). In a follow-up study, data from these surveys gave rise to the hypothesis that giant viruses are more diverse in the oceans than any cellular organism\(^{51}\). Subsequently, a large-scale analysis of the NCLDV major capsid protein (MCP), in which more than 50,000 of these proteins were found across Earth’s microbiomes, revealed the global dispersal of giant viruses, including in terrestrial ecosystems\(^{52}\).

Other approaches that enabled the discovery of novel NCLDV’s are single-virus or single-cell genomics and mini-metagenomics (BOX 1). First, sorting viral particles from marine samples enabled the detection of viruses that had previously been found to be associated with the algae *Ostreococcus* spp. and *Phaeocystis globosa*\(^{51}\). This approach led to the sequencing of several so-called giant virus single amplified genomes, of which the largest was a 813 kb genome belonging to the **Mimiviridae** that encoded a metacaspase, which potentially enables autocatalytic cell death of the host cell\(^{64}\). Single-cell methods, including sorting and genome amplification of single eukaryotic cells, were also used to identify and genome sequence five giant viruses associated with marine chaoanollagellates\(^{65,66}\); comparative genomics together with all other NCLDV genomes revealed that viruses that infect hosts with similar trophic modes,

---

**Corks**

The distinctive structures of some virions, in the case of pithovirus, the cork is located at the apex of the viral particle and made of 15 nm-spaced stripes organized in a hexagonal honeycomb-like array.

**Nucleosomes**

Compact structural forms of DNA packed through binding at positively charged proteins.

**Mini-metagenomics**

Low complexity metagenomes generated from generally tens to hundreds of cell-sized particles.

**Metacaspase**

A multifunctional cysteine-dependent protease that, for example, plays a role in programmed cell death in eukaryotes.
Box 1 | Toolkits for giant virus discovery: cultivation-independent genomic approaches

Read mapping-based approaches
Mapping metagenomic reads to giant virus reference genomes has been successfully applied to detect giant viruses and estimate their abundances in the environment1,10,11,17,18, and several tools have been published13,15. Read mapping-based approaches are advantageous because they are sensitive enough to detect giant viruses at low levels180; however, they typically do not lead to the recovery of viral genomes published178,181. read mapping-based approaches are advantageous because giant viruses are often present in environmental samples that have been horizontally acquired from bacteria or eukaryotes may lead to viral sequence read mapping to cellular genomes, resulting in false positive hits.

Marker gene surveys
Detection and phylogenetic analysis of signature genes in complex environmental datasets is a commonly used approach to assess viral diversity in metagenome data. For Nucleocytoviricota genes that encode the major capsid protein, DNA Polymerase B or viral packaging ATPase have been used as marker genes. The approach is less error-prone than read mapping as it can be coupled with phylogenetic analysis to confirm the monophyly of the respective gene homologues found in known viral genomes. This approach has been successfully applied in several studies15,42,100 and, although less sensitive than read mapping, it can detect viruses that were not abundant enough in a metagenome to be successfully assembled and binned.

Genome-resolved metagenomics
The reconstruction of MAGs through metagenome binning is an established approach to recover microbial genomes. Owing to their virion sizes, giant viruses are often present in environmental samples that have been selectively filtered to target microorganisms, although individual viral species are often found at low abundance within a high genetic diversity background. In contrast to smaller viruses such as most bacteriophages, the large genomes of most members of the Nucleocytoviricota typically require metagenomic binning to increase genome completeness111. However, in most microorganism-centric metagenome projects, giant virus genome bins were frequently neglected since tools that estimate genome quality181 predict viral genomes to be of low completeness based on their lack of cellular marker genes151, which then leads to their exclusion from downstream analyses183. Several recent studies employed custom workflows to identify GVMAGs and to estimate completeness and contamination by, for example, identifying copy numbers of conserved giant virus genes153 or inferring deviations from lineage-specific copy numbers of long-copy orthologues151. It is important to note that GVMAGs are typically incomplete, limiting the feasibility of some sequence-based inferences (for example, gene absence analyses). As Nucleocytoviricota phylogenies generally rely on a small set of viral hallmark genes, the reconstruction of evolutionary relationships is certainly feasible using incomplete GVMAGs as is the analysis of horizontal gene transfers.

Single-virus and single-cell genomic approaches
Flow cytometry-based sorting and sequencing of single viruses can be used to detect viruses in environmental samples184,185, yet only a few such studies have discovered novel giant viruses115,44,108. Owing to large virion sizes and a bright signal using DNA stains42,123, giant viruses are a promising target for sorting. A drawback of this approach is that the subsequent whole-genome amplification, if performed on a single virus, may lead to low genome recovery185. An alternative approach to direct sorting of giant viruses from an environmental sample is targeted sorting of host cells108. Viruses actively replicating inside a host cell can produce hundreds to thousands of viromes with clonal copies of viral genomes, which would vastly improve whole-genome amplification108. Furthermore, if successful, this approach enables identification of the virus and its native host.

Similarly, mini-metagenomics uses fluorescence-activated single-cell sorting or microfluidics to collect tens to hundreds of cell-sized particles112. The presence of many identical viral particles, either through repeated sorting of single clonal viruses, an infected host cell or the sorting of vacuoles filled with giant viruses, would increase genome recovery.

including host habitat and lifestyles, express distinct genetic features106,107. Furthermore, mini-metagenomics analysis (BOX 1) of a single forest soil sample led to the enrichment and discovery of 15 diverse giant virus metagenome-assembled genomes (MAGs), including several members of the Klosneuvirinae, highlighting an untapped diversity of giant viruses in soil106.

The most successful approach for obtaining NCLDV genomes from environmental sequence data is genome-resolved metagenomics (BOX 1). Since the early 2000s, this approach has become common practice for recovering genomes of bacteria and archaea from complex environmental samples19, yet it took nearly another decade before the first giant virus MAGs (GVMAGs) appeared in public databases (FIG. 1). Yau et al. reconstructed the first GVMAGs as a by-product of their work on virophages in metagenomes from the Organic Lake in Antarctica117. Several years later, four additional potentially algae-associated GVMAGs were retrieved from environmental sequence data from Yellowstone Lake in Yellowstone National Park, United States; they were found to be related to the viral families Phycodnaviridae and Mimiviridae and shared some genes with virophages that co-occurred in the same sample126. Cultivation-independent approaches for the discovery of giant virus genome-centric sequence information gained traction when members of a Mimiviridae-affiliated subfamily, the proposed Klosneuvirinae, were recovered from metagenomic data125. The fact that these were found in metagenomes from freshwater and sewage samples originating from four different continents suggested this novel group of giant viruses is cosmopolitan125. More than 20 GVMAGs from the deep sea were subsequently discovered, including 15 affiliated with the Pithoviridae, indicating a surprisingly high prevalence of pithovirus-like viruses in the ocean126, followed by the discovery of additional, likely algae-associated freshwater giant viruses in samples collected from Dishui Lake, Shanghai, China121,128. The unique strength of cultivation-independent approaches for viral genomics and discovery became most evident when more than 2,000 GVMAGs were extracted from metagenome datasets generated from analyses of thousands of samples collected from diverse biomes127; an additional 500 GVMAGs from mainly marine systems were reconstructed shortly after128. The addition of the GVMAGs to the Nucleocytoviricota species tree led to an increase in phylogenetic diversity by more than tenfold and enabled a comprehensive update of the taxonomic framework of the Nucleocytoviricota129, in which the Mesomimiviridae makes up more than
one-third of the observed diversity [FIG. 3]. The addition of the new lineages also led to a substantial increase in the size of the Nucleocytoviricota pan-genome, which now comprises more than 900,000 proteins82. This translated to an extensively expanded repertoire of functional genes, providing not only many novel insights into how giant viruses may interact with their hosts and the environment but also generating compelling novel hypotheses about their evolutionary roles82,96–98.

Exploring the host range of giant viruses. Genome-resolved metagenomics enabled the discovery of thousands of viral genomes, of which many represented lineages divergent from viruses recovered by isolation or co-cultivation82,96 (FIG. 3). However, giant viruses recovered from metagenomes typically lack information on host organisms99. An approach to overcome this limitation is the detection of viruses and potential eukaryotic hosts co-occurring in the same sample. Furthermore,
horizontal transfer of genetic material between viruses and their hosts is a common phenomenon and can go in both directions\(^{80-102}\), and the analysis of viral genes that may have been acquired through recent horizontal gene transfer (HGT) might identify host organisms. In the early days of giant virus metagenomics, read mapping-based co-occurrence analysis (Box 1) revealed that the presence of viral sequences in some marine samples was positively correlated with those of eukaryotic oomycetes\(^80\), which have not been found to be associated with NCLDVs. In another study, co-expression analysis of metatranscriptomic data revealed a strong connection between *Aureococcus anophagefferens* virus and its algal host, and also indicated that other *Mimiviridae* present in the same sample were likely associated with *Aureococcus spp.*\(^{201}\). This approach also linked *Phycodnaviridae* and *Mimiviridae* members to a wide range of marine microeukaryotes, including choanoflagellates, stramenopiles, diatoms, dinoflagellates and cercozoan algae\(^{103}\). In a different study, virus–host relationships were implied through the co-occurrence analysis of viral and eukaryotic PolB-encoding genes and the hypervariable V9 region of the eukaryotic 18S rRNA gene\(^{84}\). This approach was then applied to a comprehensive set of marine metagenomes collected during the Tara Oceans expedition, revealing that particular microeukaryotes belonging to the Alveolata, Opisthokonta, Rhizaria and Stramenopiles co-occurred with different NCLDV lineages\(^{104}\). In a similar study, a strong co-occurrence signal was detected between a virus belonging to the *Mimiviridae* and marine chrysophytes as its potential host\(^{105}\). Subsequent detection of putative HGT events between GVMAGs and chrysophyte genomes and transcriptomes provided further support for this host–virus relationship\(^{106}\). A systematic analysis of HGT candidates present in more than 2,000 NCLDV genomes, most of which were MAGs from diverse global sampling sites, revealed thousands of genes likely introduced into host chromosomes or derived from the host through recent HGT\(^{82}\). Based on these results, it was possible to propose connections between NCLDV’s and members of all major eukaryotic phyla\(^{82}\). Although most of these predicted hosts have not yet been found to be infected by giant viruses, more than 20 previously isolated virus–host relationships were successfully predicted through recent HGT events, underlining the validity of this sequence inference-based approach to metagenome-assembled viral genomes (Fig. 4).

Although sequence-based computational host predictions provide a means to expand the range of putative NCLDV hosts, the approaches have some potential challenges and biases. For example, co-occurrence analysis is dependent on sufficient host genome coverage for detection in metagenome data, and HGT analysis requires the availability of the host genomic sequences. Furthermore, it is difficult to detect ancient HGT from previous hosts. Another limitation to the analysis of the integration of NCLDV genes into host genomes can be the quality of the database used. For example, GVMAGs have been found mis-annotated as bacteria, archaea or eukaryotes in public databases, which hampers the use of automated tools for correct HGT detection\(^{82,106}\). Despite some of these limitations, expanding the putative host range of metagenome-derived NCLDV’s provides a basis for targeted sampling of putative hosts, for the study of virus–host co-evolution and to identify viral-encoded functions for targeted modulation of host metabolism. Sequence-based inferences of viruses and their hosts may then be extrapolated to assess the impact of such interactions on global ecosystems.

**From HGT to endogenization**

Not only is HGT between viruses and their hosts a common phenomenon but some giant viruses can even integrate their entire genomes into the host chromosome (Fig. 4). This so-called endogenization is a mechanism
observed for most eukaryotic viruses. Arrays of NCLDV genes have occasionally been found in genomes of eukaryotes, in particular in algae, plants, and amoebae. A recent survey of published eukaryotic genomes and transcriptomes revealed the presence of giant virus genes in 66 different eukaryotes, including several Acanthamoeba species, flagellates, ciliates, stramenopiles, oomycetes, fungi, arthropods and diverse unicellular and multicellular algae (Fig. 4). Yet, for many of these eukaryotes, giant virus infections have not been observed. The integration of NCLDV genes often appears to be highly host specific, with viral genes detected in one eukaryotic species being unrelated to viral genes found in closely related species. Among the integrated genes are NCLDV hallmark genes that are, in some instances, scattered throughout the host chromosome and, in others, co-localized in islands composed of more than 100 genes. The integration of complete viral genomes has been described for some members of the Mesomimiviridae, for example, Ectocarpus siliculosus virus integrated into its brown algal host more than 20 years ago likely through use of integrases. The related Phaeocystis globosa virus is a lysogenic virus that causes continuous infections, which is in stark contrast to many other known NCLDV lineages that were successfully isolated based on the fact that they lyse their amoeba host. The analysis of existing algal genomes and transcriptome data revealed other examples of whole giant virus genomes integrated into eukaryotic host chromosomes. Some regions encoded more than 1,500 viral genes, making up to 10% of the genes of the green algal host. Several of the detected viral genes were annotated as enzymes with roles in carbohydrate metabolism, chromatin remodelling, signal transduction, energy production and translation.

It remains unknown whether integrated giant viruses are dormant with no or minimal benefit to the host, or whether the host cell benefits from some viral genes that may provide or fine-tune metabolic capabilities. Another unanswered question is whether there are mechanisms encoded in the integrated viral genome that may reactivate infection after transcribing and translating some of the integrated viral genes. This would then be followed by the release of the giant virus genetic material during host replication and effective dispersal to new hosts. If there is no reactivation of viral infection, giant virus genes decay over time, leading to rearrangements and pseudogenization and making their detection more challenging or impossible. Giant virus endogenization has been found mainly through the analysis of eukaryotic isolate genomes, but we anticipate that genome-resolved metagenomics of eukaryotes will further facilitate the discovery of many additional examples of this phenomenon. Future investigation of the integration of giant virus genes is expected to provide some answers for how endogenization has shaped and continues to shape the evolution and ecology of eukaryotic organisms.

Reprogramming of the host and its impact on host populations

Upon infection, a virus reprograms its host cell and turns it into a so-called virocell that supports viral replication. Analogous to bacteriophages, which are viruses (including large ones) that infect bacteria, giant viruses seem to contribute genes to their hosts to augment and/or modulate the metabolic capabilities of the host cell. The first described example was a virus-encoded hyaluronan synthase, encoded by Chlorella virus, that enabled its algal host to synthesize hyaluronan. In addition, an active potassium channel encoded by Chlorella virus was found to be integrated into the host membrane during infection. Another example is that of a host-derived nitrogen transporter in Ostreococcus tauri virus that is expressed during the infection of its green algal host. Experimental characterization provided evidence that this transporter may increase the uptake of nitrogen by the host cell. Other studies revealed the presence of fermentation genes in the Tetradselmis virus genome with possible implications for host metabolism in nutrient-limited marine systems. A survey of giant virus isolates and MAGs revealed the widespread presence of genes for cytochrome P450 monooxygenases, potentially enabling or modulating complex metabolic processes such as the synthesis of sterols and other fatty acids. Metagenome-informed experimental characterization of the distinctive cytochrome P450 of hokovirus did not reveal any sterol metabolized by the recombinant viral cytochrome P450 (Ref. 4). Distant homologues of eukaryotic actins (‘viractins’) and myosins (‘virmyosins’) have been found in NCLDV genomes in two recent studies and a preprint, indicating that these viruses impact cell structure, motility and intracellular transport processes; however, further functional validation is needed. Furthermore, a giant virus related to Mesomimiviridae that infects heterotrophic choanoflagellates was found to encode type 1 rhodopsins together with the pathway for synthesis of the required pigment, β-carotene. Metagenome-informed experimental characterization of the NCLDV rhodopsin showed that the putative rhodopsin likely functions as a proton pump, generating energy from light. A phylogenetically distinct NCLDV rhodopsin was found in a GVMAG from Organic Lake, Antarctica, and experimental characterization of this protein revealed that it may function as a light-gated pentameric ion channel, potentially impacting ion homeostasis and phototaxis of the host cell. Furthermore, through global metagenomics, it was predicted that genes encoding various substrate transport processes, energy generation through light (rhodopsins and genes involved in photosynthesis), carbon fixation and glycolysis are commonly found in GVMAGs affiliated with diverse lineages of the Nucleocytoviricota (Fig. 5). More detailed phylogenetic analysis revealed that some auxiliary metabolic genes encoding transporters for iron, phosphate, magnesium and ammonium originated in eukaryotic hosts and were likely recently acquired by giant viruses through HGT. Other genes encoding several rhodopsins, sucrose hydrogenase, aconitate and glyceraldehyde 3-phosphate dehydrogenase showed a pattern that suggested a viral origin or a common evolutionary origin in one of the ancestral hosts. Taken together, the widespread presence of metabolic genes in diverse NCLDV lineages implies that augmenting host metabolic capacities is likely a
strategy more commonly used by NCLDVs than initially assumed. However, the current lack of experimental evidence of the functions and activities of most of these genes and pathways as well as their effects on the host cell demands further experimental investigation.

Metabolic reprogramming has direct consequences on host population structure and dynamics. One striking example is the cosmopolitan marine coccolithophore *Emiliania huxleyi*, which forms massive blooms that play key roles in global carbon and sulfur cycles. *E. huxleyi* populations are subject to persistent but ultimately lytic infections by the coccolithovirus *Emiliania huxleyi virus*. Once lysis is induced, it leads to the termination of the algal bloom and the deposition of massive amounts of calcite and nutrients into the ocean, which increases the marine pool of dissolved organic matter. Importantly, viral infections do not only lead to host lysis but also promote viral replication by rewiring host physiology, in particular the turnover of sugars and synthesis of fatty acids and lipids. Comparably little is known about how host populations are impacted by giant viruses that were recovered through genome-resolved metagenomics but, considering the predicted host range of these viruses, it is conceivable that similar principles are omnipresent and are actively shaping the biomes and biogeochemical cycles of Earth.

**Giant virus genomes encode hallmark genes of cellular life**

Among the most intriguing features found in giant virus genomes are hallmark genes of cellular life such as tRNAs and genes involved in protein biosynthesis. This phenomenon was first described upon sequencing the mimivirus genome. Subsequent analyses revealed the phylogenetic placement of virus-encoded cellular genes between bacteria and eukaryotes, suggesting an ancient origin. Other cellular hallmark genes with similarly deep branching patterns were found in other giant virus genomes and led to the hypotheses that giant viruses may either represent a fourth domain of life or are remnants of a highly degraded eukaryotic cell derived by reductive evolution. The subsequent use of more complex phylogenetic models revealed that many of these genes had most likely been acquired from different eukaryotic hosts.
Some of these genes might represent ancient transfers from undiscovered eukaryotic hosts. This finding provided evidence for the hypothesis that giant viruses may have evolved from smaller viruses. Yet, other studies have reported alternative topologies for some housekeeping and other metabolic genes of cellular organisms, including rhodopsins and cytochrome P450 (REF. 98). It has also been proposed that some genes may have been transferred from ancestral giant viruses to past eukaryotic hosts, or even to a proto-eukaryote, highlighting a potentially integral role of giant viruses in the evolution of the eukaryotic cell (REFs. 142,143). Furthermore, it is possible that some genes that may function as part of the eukaryotic core metabolism were introduced upon integration of giant virus genetic material into the genome of an ancient eukaryotic cell, further shaping eukaryotic evolution (REFs. 142,143). The presence of genes for aminoacyl tRNA synthetases (aaRS) and eukaryotic translation factors has been recorded multiple times in newly recovered giant virus genomes. Indeed, a nearly complete set of 20 aaRS has been reported in klosneuvirus from metagenomic data (REF. 100). Shortly after, two tupsanviruses were isolated with genomes that contain a full set of aaRS and tRNAs, and subsequently the first Klosneuvirinae isolates were described, of which one also contained a complete set of aaRS (REF. 101). Especially in the Klosneuvirinae, the presence of aaRS with lineage-specific evolutionary histories provided additional support that these genes derived from different eukaryotic hosts (REF. 102). The presence of genes for a complete set of aaRS is currently constrained to members of the Mimiviridae and information on the role of giant virus aaRS in host interactions is limited; however, some have been experimentally studied and were indeed functional (REF. 103). There is even some experimental evidence for the potential roles of these genes in making giant viruses less dependent on host machinery, for example, during shutdown of host translation in response to viral infection or other adverse conditions (REF. 104). On the other hand, a suspected role in enhancing viral translation by providing additional copies of aaRS to support host translation has not yet been confirmed. Additional hallmark genes of cellular life include those encoding for the four core histones (REFs. 105,106,148,149) and giant virus genes predicted to be involved in energy generation (REFs. 150,151). A recent study reported an active membrane potential in Pandoravirus mollivirus virions together with the expression of several remote homologues of tricarboxylic acid cycle genes (REF. 152). Despite encoding functions that were recently thought to be exclusively present in cellular organisms, there is currently no evidence that giant viruses perform protein translation without host-derived ribosomes or host-independent energy generation.

Conclusions
Nearly 20 years of giant virus isolation has yielded viral isolates representing highly diverse lineages. Complementary detailed research on the biology of these viruses has revealed many important details of virion structures and infection strategies. It has become clear that there are stark differences in virion size and structure and, although there are some similarities in how these viruses enter and exit the host cell, most giant viruses employ contrasting strategies for replicating within and exploiting their host cells. Sequencing of viral isolates has led to the discovery of the largest and smallest known genomes of viruses of the Nucleocytoviricota.

Cultivation-independent approaches have accelerated the discovery of genome sequences of new giant viruses and other large viruses in the Nucleocytoviricota, providing novel insights into their phylogenetic diversity and functional potential. Metagenomics also revealed that these viruses can be found nearly anywhere on Earth, are affiliated with diverse eukaryotes and are likely modifying host physiology through metabolic reprogramming, ultimately altering the structure and function of host communities in the environment. At the same time, estimates based on NCLDV hallmark genes in metagenomic datasets indicated that only a small fraction of giant virus genomes have been discovered so far and that the diversity of giant viruses may be far greater than that of bacteria, at least in the oceans (REF. 153). A controlled metagenomic binning experiment where giant viruses were spiked into an environmental sample showed that genome fragments of many giant viruses that are present in a given sample likely remain below the detection limit, highlighting the need for ultra-deep metagenome sequencing (REF. 154) or targeted isolation efforts (REF. 155). Furthermore, there is a strong bias towards detecting giant viruses that are similar to those already known, as tools used to identify viruses from metagenomes rely heavily on features observed in sequenced NCLDV genomes such as large sets of conserved genes (REFs. 156,157,158,159). However, giant virus genomes exhibit extensive plasticity, such that viruses within the same clade quickly diverge and share very few genes (REF. 160). A recent stunning example of NCLDV diversity is yaravirus, which was isolated with its native amoeba host (REF. 161), yet no closely related sequences were detectable in public metagenomic datasets. Its placement within NCLDV was difficult owing to more than 90% of its genes lacking similarity to those in public databases and the paucity of most viral hallmark genes (REF. 162), and its placement within the Nucleocytoviricota is currently still under debate. Furthermore, a recent preprint described the genome-resolved metagenomic-based discovery of the Proculviricetes and Mirusviricetes from marine systems, which might be a two-class-level novel lineages within the Nucleocytoviricota that lack most of the typical viral hallmark genes (REF. 163). Taken together, the excessive gene novelty of viruses in the Nucleocytoviricota, observed through both cultivation and cultivation-independent methods, further underlines that many giant viruses are likely to be hiding in plain sight.

Published online 28 July 2022

1. Fischer, M. G. Giant viruses come of age. Curr. Opin. Microbiol. 31, 50–57 (2016).
2. Iyer, L. M., Balaji, S., Koonin, E. V. & Aravind, L. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. 117, 156–184 (2006).
3. Koonin, E. V. et al. Global organization and proposed megataxonomy of the virus world. Microbiol. Mol. Biol. Rev. 84, e00061-19 (2020).
4. Sun, T.-W. et al. Host range and coding potential of eukaryotic giant viruses. Viruses 12, 1537 (2020).
5. Abergel, C., Legendre, M. & Claverie, J.-M. The rapidly expanding universe of giant viruses: mimivirus, pandoravirus, pilovirus and mollivirus. FEMS Microbiol. Rev. 39, 779–796 (2015).
Isolation and characterization of a novel giant virus related to mimovirus together with its native host, the marine predaceous flagellate Cafeteria roenbergensis.
81. Schulz, F. et al. Schulz, F. et al. Hidden diversity of soil giant viruses. (2015).
82. Martínez Martínez, J., Swan, B. K. & Wilson, W. H. Martínez Martínez, J., Swan, B. K. & Wilson, W. H. A distinct lineage of giant DNA viruses in Tara oceans microbial metagenomes. (2011).
83. Mihara, T. et al. Mihara, T. et al. Linking virus genomes with host phylogenetic diversity, the repertoire of predicted targeted viromics. (2019).
84. Da Cunha, V., Gaia, M., Ogata, H., Jaillon, O. & Delarouque, N. Delarouque, N. & Boland, W. The genome of the brown marine virus isochilensis was isolated from a marine sample. (2019).
85. Hingamp, P. & Worthey, E. Hingamp, P. & Worthey, E. Exploring nucleocytoplasmic large DNA virus diversity in Tara oceans metagenomes. (2015).
86. Schulz, F. et al. Giant virus diversity and host interactions through global metagenomics. Nature 527, 452–458, doi:10.1038/nature15790 (2015).
87. Recovery of more than 2,000 giant virus genome-assembly-generated genomes from global metagenomes. Nature 540, 557–561, doi:10.1038/nature20749 (2017).
88. Martinez Martinez, J. Martinez Martinez, J. & Feschotte, C. Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. 13, 283–296 (2012).
89. Chiba, S. et al. Widespread endogenization of genome sequence fractions, depths and biomes in marine samples and predictions of their associations with eukaryotic communities. Nat. Commun. 5, 4268 (2014).
90. Wang, L. et al. Wang, L. et al. Viral infection with ocean viruses brings a rhodopsin photosystem to unicellular algae. Nature 529, 354–361 (2016).
91. Delaroque, N. & Boland, W. Delaroque, N. & Boland, W. The genome of the brown marine virus isochilensis was isolated from a marine sample. (2019).
92. Weinheimer, A. R. & Aylward, F. O. Weinheimer, A. R. & Aylward, F. O. Dynamic genome reprogramming. Proc. Natl Acad. Sci. USA 116, 20377–20383, doi:10.1073/pnas.1912006116 (2019).
93. Delmont, T. O. Delmont, T. O. Giant viruses encode novel types of complement of translation system components. (2020).
94. Hurwitz, B. L., Hallam, S. J. & Sullivan, M. B. Hurwitz, B. L., Hallam, S. J. & Sullivan, M. B. Dynamic genome reprogramming. Proc. Natl Acad. Sci. USA 116, 20377–20383, doi:10.1073/pnas.1912006116 (2019).
95. Delaroque, N., Maier, I., Knippers, R. & Müller, D. G. Delaroque, N., Maier, I., Knippers, R. & Müller, D. G. Widespread endogenization of genome sequence fractions, depths and biomes in marine samples and predictions of their associations with eukaryotic communities. Nat. Commun. 5, 4268 (2014).
96. Chiba, S. et al. Widespread endogenization of genome sequence fractions, depths and biomes in marine samples and predictions of their associations with eukaryotic communities. Nat. Commun. 5, 4268 (2014).
97. Weinheimer, A. R. & Aylward, F. O. Weinheimer, A. R. & Aylward, F. O. Dynamic genome reprogramming. Proc. Natl Acad. Sci. USA 116, 20377–20383, doi:10.1073/pnas.1912006116 (2019).
98. Hurwitz, B. L., Hallam, S. J. & Sullivan, M. B. Hurwitz, B. L., Hallam, S. J. & Sullivan, M. B. Dynamic genome reprogramming. Proc. Natl Acad. Sci. USA 116, 20377–20383, doi:10.1073/pnas.1912006116 (2019).
99. Delmont, T. O. Delmont, T. O. Giant viruses encode novel types of complement of translation system components. (2020).
100. Hurwitz, B. L., Hallam, S. J. & Sullivan, M. B. Hurwitz, B. L., Hallam, S. J. & Sullivan, M. B. Dynamic genome reprogramming. Proc. Natl Acad. Sci. USA 116, 20377–20383, doi:10.1073/pnas.1912006116 (2019).
101. Fiille, J. & Chandler, M. Fiille, J. & Chandler, M. Gene exchange and the origin of giant viruses. Intervirology 53, 354–361 (2010).
102. Fiille, J. & Chandler, M. Fiille, J. & Chandler, M. Gene exchange and the origin of giant viruses. Intervirology 53, 354–361 (2010).
Reviews

148. Rolland, C. et al. Clandestinovirus: a giant virus with chromatin proteins and a potential to manipulate the cell cycle of its host Vermamoeba vermiciformis. Front. Microbiol. 12, 715608 (2021).

149. Boyer, M. et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc. Natl Acad. Sci. USA 106, 21848–21853 (2009).

150. Aherfi, S. et al. Incomplete tricarboxylic acid cycle and proton gradient in Pandoravirus massiliensis: is it still a virus? ISME J. https://doi.org/10.1038/s41396-021-01117-3 (2021).

151. Schulz, F. et al. Advantages and limits of metagenomic assembly and binning of a giant virus. mSystems 5, e00484-20 (2020).

152. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat. Biotechnol. https://doi.org/10.1038/s41587-020-00774-7 (2020).

153. Andreani, J. et al. Morphological and genomic features of the new kleinosvirus isolate fadovivirus IHUMI-VV54. Front. Microbiol. 12, 719703 (2021).

154. Brussaard, C. P. D., Short, S. M., Frederickson, C. M. & Suttle, C. A. Isolation and phylogenetic analysis of novel viruses infecting the phyttoplankton Phaeocystis globosa (Prymnesiophyceae). Appl. Environ. Microbiol. 70, 3700–3705 (2004).

155. Santini, S. et al. Genome of Phycodnavirus globosa virus Pgl-167 highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. Proc. Natl Acad. Sci. USA 110, 10800–10805 (2015).

156. Sandaa, R. A., Heldal, M., Castberg, T. Thyrhaug, R. & Bratkovic, G. Isolation and characterization of two viruses with large genome size infecting Chrysochromulina ericina (Prymnesiophyceae) and Pyramimonas orientalis (Prasinophyceae). Virolology 290, 272–280 (2001).

157. Stough, J. M. A. et al. Genome and environmental activity of a Chrysochromulina parva virus and its virophages. Front. Microbiol. 10, 705 (2019).

158. Gastrich, M. D., Anderson, O. R., Bennmayor, S. S. & Cooper, E. M. Ultrastructural analysis of viral infection in the brown-tide alga, Aureococcus anophagefferens (Pelagophyceae). Phycologia 37, 300–306 (1998).

159. Thomas, V. et al. Lassaanavirus, a giant amoebal virus encoding histone doublers. Environ. Microbiol. 13, 1454–1466 (2011).

160. Cottrell, M. T. & Sutton, C. A. Dynamics of lytic virus infecting the photosynthetic marine picoflagellate Micromonas pusilla. Limnol. Oceanogr. 40, 750–759 (1995).

161. Bratkovic, G., Egge, J. K. & Heldal, M. Viral mortality of the marine alga Emiliana huxleyi (Haptophyceae) and termination of algal blooms. Mar. Ecol. Prog. Ser. 95, 39–48 (1995).

162. Watanabe, R., Song, C., Kayama, Y., Takemura, M. & Murata, K. Particle morphology of medusa virus inside and outside the cells reveals a new maturation process of giant viruses. J. Virol. 96, e01853-21 (2022).

163. Legendre, M. et al. Diversity and evolution of the emerging Pandoraviridae family. Nat. Commun. 9, 2285 (2018).

164. Yosuf, N. et al. Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moomoxivirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome Biol. Evol. 4, 1532–1530 (2012).

165. Rodriguez, R. A., L. Mougari, S., Colson, P., La Scola, B. & Abraluo, J. S. ‘Tupanavirus’, a new genus in the family Mimiviridae. Arch. Viro’ https://doi.org/10.1007/s00705-018-0677-4 (2018).

166. Andreani, J. et al. Morphological and genomic features of the new kleinosvirus isolate fadovivirus IHUMI-VV54. Front. Microbiol. 12, 719703 (2021).

167. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. 35, 725–751 (2017).

168. Woyke, T., Doud, D. F. R. & Schulz, F. The trajectory of microbial single-cell sequencing. Nat. Methods 14, 1045–1054 (2017).

169. Martinez, J. M., Martinez-Hernandez, F. & Martinez-Garcia, M. Single-virus genomes and beyond. Nat. Rev. Microbiol. 18, 705–716 (2020).

170. Khali, J. Y. B. et al. High-throughput isolation of giant viruses in liquid medium using automated flow cytometry and fluorescence staining. Front. Microbiol. 7, 26 (2016).

171. Yu, F. B. et al. Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex environmental samples. Elife 6, e26580 (2017).

Acknowledgements

This work was conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, under contract no. DE-AC02–05CH11231. C.A. received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (grant agreement no. 832601). The authors thank K. R. Chuan from the Department of Chemistry and Biochemistry, University of Texas, El Paso, USA, for providing 3D reconstruction images for AaV, mimivirus and CroV. The authors acknowledge R. Watanabe and K. Murata, ExCELLS, NINS, Japan, who provided 3D reconstruction image for medusa virus, R. N. Burton-Smith and K. Murata, ExCELLS, NINS, Japan, for cryo-electron micrographs of melbournevi- rus, and T. Risse, Department of Biological Sciences, Purdue University, USA, for the 3D reconstruction image for faustovirus.

Author contributions

All authors researched data for the article, substantially contributed to discussion of content, and wrote and reviewed/edit the manuscript before submission.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Microbiology thanks Frank Aylward, James Van Etten and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s), author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature Limited 2022