



# Molecular architecture of giant viruses infecting microbial eukaryotes (protists)

ANHELINA KYRYCHENKO<sup>1,2</sup> 

<sup>1</sup>Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw, Poland

<sup>2</sup>D.K. Zabolotny Institute of Microbiology and Virology of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

Received: 12 March 2025, Revised: 12 July 2025, Accepted: 24 July 2025

## Abstract

In this review, I describe recent findings on the molecular architecture and genomic characterization of giant viruses that infect microbial eukaryotes (protists) across diverse ecosystems and ecological niches. Giant viruses are distinguished by their large and complex genomes, which encode a wide range of functions, including protein translation, carbohydrate and lipid metabolism, nitrogen cycling, light assimilation, and key metabolic pathways such as glycolysis and the tricarboxylic acid cycle. Additionally, these genomes feature unique genes, often acquired through horizontal gene transfer, that are not found in other viruses and contribute to the viruses' ability to manipulate host metabolism and evade host defenses. A core set of genes conserved across different families of giant viruses is highlighted, serving as essential components for key life-cycle processes and providing valuable phylogenetic markers. The review also discusses the role of ORFans and virophages in contributing to the genetic diversity and evolutionary adaptation of these viruses. These findings are crucial for understanding the diversity, evolutionary mechanisms, and complex virus–host interactions of giant viruses, as well as for developing more advanced classification systems. Furthermore, the potential biotechnological applications of unique viral genes and pathways are explored, underscoring the importance of ongoing research in this field.

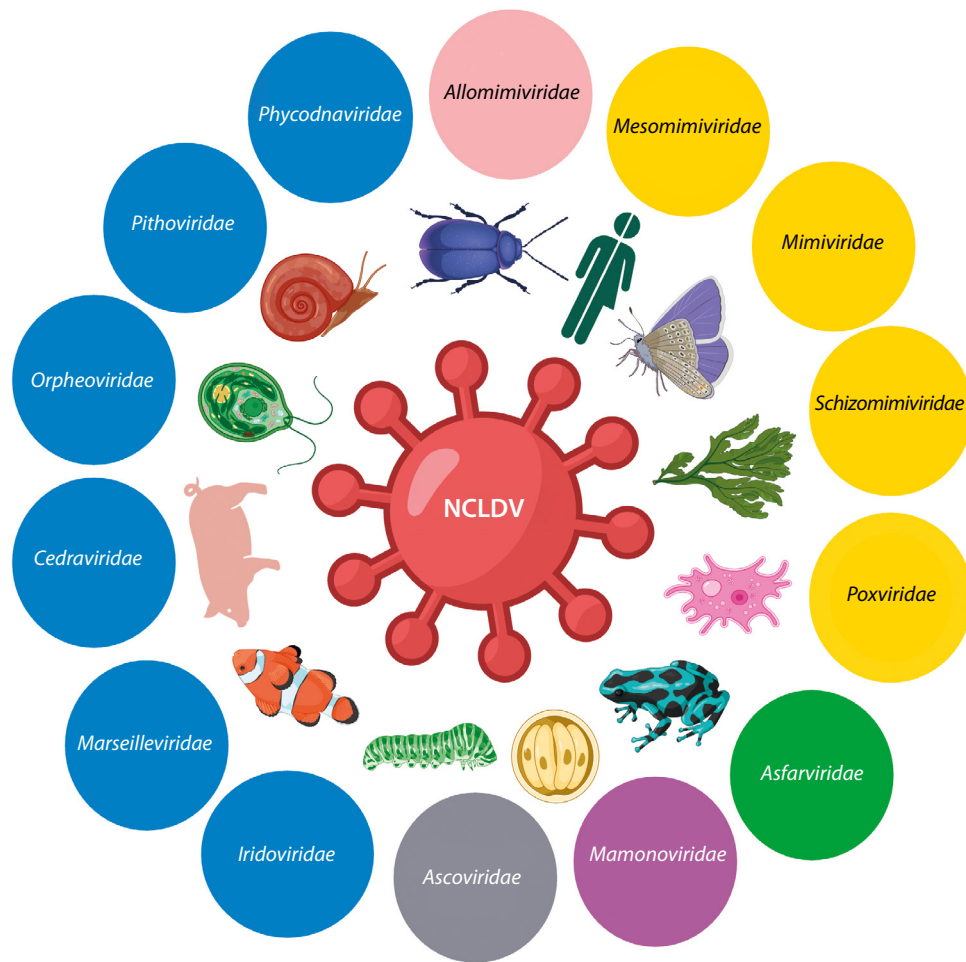
**Key words:** *Imitervirales*, metagenomic analysis, nucleocytoplasmic large DNA viruses, protists, virus–host interactions

## Introduction

Nucleocytoplasmic large DNA viruses (NCLDVs) are large double-stranded DNA eukaryotic viruses that infect a variety of hosts, including protozoa, invertebrates, and eukaryotic algae (Figure 1). Viruses in this monophyletic group – also known as giant viruses – replicate either fully or partially in the cytoplasm of eukaryotic cells and are characterized by huge genomes and virion sizes, which can reach 2.5 Mb and 1.5 µm, respectively. The discovery of *Acanthamoeba polyphaga* mimivirus (APMV) in amoebae in 2003 surprised the scientific community due to its enormous genome size, unprecedented genetic complexity, and the presence of genes previously thought to be exclusive to cellular life, including those related to translation, DNA repair, and chaperone

production (La Scola et al. 2003; Raoult et al. 2004). This discovery spurred more active searches for giant viruses in diverse environments, leading to a deeper understanding of their genome organization, replication mechanisms, transcriptional processes, virulence factors, genetic diversity, and phylogenetic relationships. Moreover, the discovery of giant viruses has challenged earlier assumptions about the structural and genomic simplicity of viruses and has led to a refinement of classical virus demarcation criteria. This has prompted renewed interest in viruses, their evolution, and their potential roles in ecosystems (Bosmon et al. 2025).

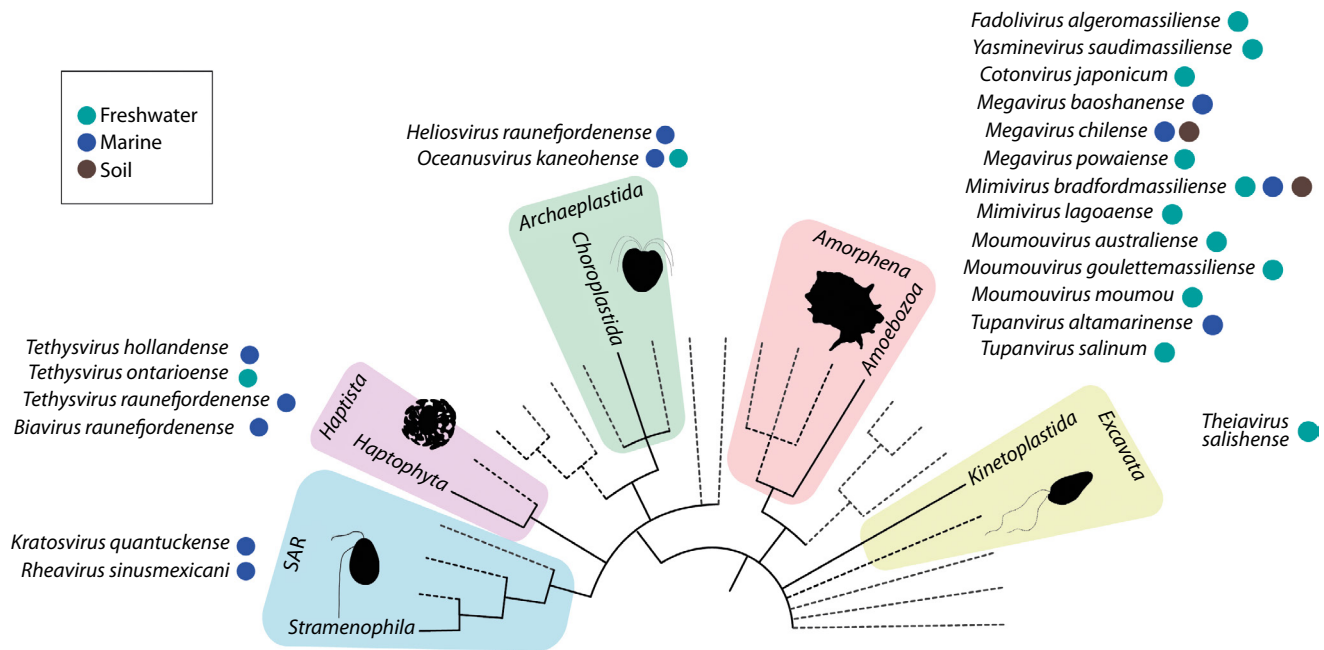
Giant viruses are primarily known to infect protists, a diverse group of eukaryotic microorganisms that includes algae, amoebae, ciliates, and other single-celled



**Figure 1.** Schematic representation of nucleocytoplasmic large DNA viruses families, illustrating the diversity of giant viruses. Created in <https://BioRender.com>

organisms (Fischer 2016; Adl et al. 2018; Kostygov et al. 2021; Queiroz et al. 2024). These viruses have been identified across various protist clades, reflecting their broad distribution and diverse host range (Figure 2). Protists play crucial roles in aquatic and terrestrial ecosystems, contributing to vital processes such as nutrient and carbon cycling, primary production, and energy flow. They inhabit a range of environments – including freshwater and marine systems, terrestrial habitats, and symbiotic or parasitic relationships within other organisms – and have even been discovered in extreme habitats (Rappaport and Oliverio 2023). The ubiquity and ecological significance of protists make them valuable model systems for in-depth studies of cell biology and for understanding complex cellular processes, functions, and responses to viral infections at the molecular and structural levels (Queiroz et al. 2024).

Understanding the prevalence and diversity of giant viruses infecting protists is essential for several reasons. First, it sheds light on their impact on ecosystems and their potential roles in protist biology and evolution, as giant viruses can significantly alter protist population dynamics through widespread genetic and metabolic reprogramming (Hurwitz et al. 2013; Ku 2021; Queiroz et al. 2024). Second, studying NCLDV genomes helps elucidate the mechanisms of virus evolution and diversification, revealing how these entities have acquired complex metabolic capabilities, developed sophisticated functional features, and adapted to their hosts (Schulz et al. 2022; Tee and Ku 2025). Furthermore, the impact of giant viruses on human health remains unclear and represents an emerging area of research. Current data on the identification of viral sequences in the human gut microbiome, the ability of some giant



**Figure 2.** Giant viruses found in major protist lineages. Created in <https://BioRender.com>

viruses to replicate in human peripheral blood cells, and their capacity to activate the interferon system suggest potential implications. However, many aspects of the giant virus–human relationship – such as pathogenicity, transmission dynamics, and overall influence on human health – remain to be elucidated (Brandes and Linial 2019; Athira and Antony 2023). Additionally, the unique biochemical pathways and enzymes found in giant viruses offer potential applications in biotechnology and medicine, as their complex enzymatic machinery presents new opportunities for innovative uses (Brahim Belhaouari et al. 2022; de Oliveira et al. 2022).

Recent taxonomic updates classify NCLDV within the phylum *Nucleocytoviricota*, which encompasses several approved viral orders (*Algalvirales*, *Asfuvirales*, *Chitovirales*, *Imitevirales*, and *Pimascovirales*, plus one unassigned order) and 11 taxonomic families (Figure 1) (Aylward et al. 2023; ICTV 2025). Viruses within this phylum possess large double-stranded DNA (dsDNA) genomes, replicate partially or entirely in the host cytoplasm, and share genetic similarities that support a hypothesized common origin. The latest classifications are based on comprehensive analyses of gene content, morphology, and phylogenomics, with a strong emphasis on phylogenetic relationships and an evolutionary framework (Koonin et al. 2020; Simmonds et al. 2023).

Giant viruses have highly complex and chimeric genomes that reflect dynamic gene exchange with their host cells and other viruses. In addition to the basic machinery for virion structure and DNA replication, these viruses often encode genes involved in translation, glycolysis, the tricarboxylic acid cycle, cytoskeletal dynamics, light harvesting, nutrient transport, and other pathways essential for nutrient homeostasis (Iyer et al. 2006; Endo et al. 2020; Farzad et al. 2022; Schulz et al. 2022). These findings collectively suggest that giant viruses employ a diverse array of functional genes to manipulate host physiology and modify the intracellular environment to facilitate virion propagation.

Recent advances in metagenomic techniques, bioinformatics tools, and high-throughput sequencing technologies have enabled the discovery of previously unidentified NCLDVs from environmental samples in a cultivation-independent manner (Tang 2020). These tools allow for the reconstruction of individual NCLDV genome sequences from metagenomic datasets and the exploration of predicted functions encoded in their genomes (Tang 2020; Farzad et al. 2022; Schulz et al. 2022). By applying sequence-based homology searches, computational predictions from bioinformatic analyses, domain and motif analyses, and functional annotations, researchers can discover, identify, and classify giant virus

genes and proteins, assess their diversity and functional potential, and better understand how they influence host metabolism and drive cellular reprogramming.

This review presents current data on the genomic organization of giant viruses infecting protists, with a focus on the functional features of newly identified genes in NCLDV genomes and the roles they may play in virus–host interactions. Giant viruses infect a broad spectrum of eukaryotic hosts; members of the *Asfuvirales*, *Chitovirales*, and *Pimascovirales* orders infect various metazoan and protist hosts, while viruses within the *Algavirales* and *Imitervirales* orders typically infect algae, amoebae, and other protists (Weynberg et al. 2017; Wilhelm et al. 2017; Claverie and Abergel 2018; Karki et al. 2021; Koonin and Yutin 2019; Ha and Aylward 2024). The present discussion is limited to viruses from the *Imitervirales* order, as they are especially abundant and widely distributed across diverse environments. These viruses possess complex gene structures, and their genomes contain genes with a wide range of functions, including light-sensitive proteins (rhodopsins), structural proteins of the cytoskeleton, and enzymes involved in central metabolic pathways such as the TCA cycle and glycolysis (Abrahão et al. 2018).

Recent changes in virus taxonomy – particularly the recognition of *Allomimiviridae*, *Mesomimiviridae*, and *Schizomimiviridae* as distinct families within the *Imitervirales* – have expanded our understanding of the diversity among viruses closely related to *Mimivirus*, often described as the “extended *Mimiviridae*” group (Aylward et al. 2023; ICTV 2025). This overview may be useful for researchers studying viral evolution and virus–host interactions, as well as for molecular biologists and biotechnologists exploring the potential applications of enzymes encoded by giant viruses.

## Genes and proteins of Imitervirales

### General genome architecture of Mimiviridae

The *Mimiviridae* is one of the largest and most diverse families of eukaryotic viruses. It includes unusually large viruses that infect a range of protists, particularly amoebae. These viruses are notable for their exceptionally large virion sizes (> 400 nm) and enormous genomes, which encode more than 900 proteins. *Mimivirus*, a prominent member and the prototype of the *Mimiviridae* family, carries a linear double-stranded DNA genome of approximately 1.2 Mbp – exceeding

the genome sizes of most known viruses and even some bacteria, archaea, and lower eukaryotes (Abrahão et al. 2018; Koonin et al. 2015). More than 60% of the genes found in *Mimiviridae* viruses have no currently assigned function. The remaining genes are primarily involved in processes essential for viral survival, such as nucleotide synthesis, recombination, DNA repair, protein translation, and other pathways typically found in cellular organisms (Claverie and Abergel 2018; Dos Santos Silva et al. 2020). Given their genome size and unexpected complexity, these viruses have drawn considerable scientific interest, particularly in studies of their molecular organization and the functions of the proteins they encode.

The gene and protein arsenal of *Mimiviridae* can be categorized into several functional groups: structural proteins (e.g., capsid proteins, membrane proteins); enzymes involved in replication and transcription (e.g., DNA polymerases, helicases); translation components (e.g., aminoacyl-tRNA synthetases); host interaction and modulation proteins (involved in host defense evasion and manipulation); metabolic enzymes (e.g., those involved in nucleotide and lipid metabolism); and accessory proteins, such as protein kinases and ubiquitin-related factors (Claverie et al. 2009). Unique to these viruses are genes canonically involved in signal transduction (e.g., serine/threonine kinases and phosphoinositide 3-kinase), DNA processing, and assembly, as well as virally encoded DNA repair systems like base excision repair (BER) enzymes (Claverie et al. 2009; Colson et al. 2018; Kalafati et al. 2022; Lad et al. 2023).

In addition to the core gene sets common to NCLDVs, mimiviruses possess several unexpected genes, such as components of the translation system that are unable to initiate transcription of early genes in the host cytoplasm immediately after infection (Kalafati et al. 2022). Notably, *Mimivirus* and related viruses carry genes involved in DNA repair, including those encoding a mismatch-recognition protein (MutS) homolog, NEIL-1 and NEIL-2 proteins, and PrimPol – enzymes that function within the BER pathway. The *Mimivirus* polymerase X protein (mvPolX) also contributes to the repair process by binding gapped DNA substrates and performing single-nucleotide gap filling, followed by downstream strand displacement (Lad et al. 2023). These genes suggest a high degree of replication autonomy and the presence of complex genome maintenance mechanisms (Bandaru et al. 2007; Lad et al. 2023).

In addition, the presence of histone-like proteins (homologs of eukaryotic histones) encoded in the viral genome and involved in organizing genomic DNA within the capsid – found in *Mimiviridae* and certain other viruses of the NCLDV group – represents a unique feature that reflects the complex evolutionary history of these viruses and their possible connections to early eukaryotic hosts (Sharma et al. 2016; Talbert et al. 2022; Sharma et al. 2024). Genome analyses of these viruses suggest that many viral genes were acquired through horizontal gene transfer from their hosts, co-infecting viruses, or other intracellular parasites existing within the same cell (Kalafati et al. 2022).

*Mimiviridae* genomes also exhibit high proportions of duplicated genes, ORFans (genes with no known homologs), and transpovirons – mobile genetic elements resembling transposons (Colson et al. 2018; Dos Santos Silva et al. 2020; Kalafati et al. 2022). These features contribute to the unique structural and functional properties of the viruses, possibly aiding in host interactions and adaptation. Furthermore, the presence of sequences resembling intronic regions of 18S ribosomal RNA genes further underscores the complexity of the viral genome (Abrahão et al. 2018). While the exact functional roles of these intronic sequences remain under investigation, their presence suggests potential regulatory functions in viral gene expression or interference with host cellular processes.

The large genomes of giant viruses, along with multiple packaged proteins, require genome condensation. In viruses belonging to the phylum *Nucleocytoviricota*, this condensation is believed to be facilitated by histone-like or nonhistone proteins. The discovery of proteins such as gp275 – an architectural protein involved in organizing genomic DNA within the *Mimivirus* capsid and similar to the histone-like methanogen chromosomal protein 1 (MC1) – highlights the sophisticated mechanisms these viruses employ for genome condensation and protection. Remarkably, no DNA-condensing protein responsible for assembling the enormous genome had been reported until this discovery (Sharma et al. 2024).

Another notable feature of the *Mimiviridae* is the presence of cyclophilins (CYNs), prolyl isomerases not typically found in viruses (Barik 2018). CYNs are known for their peptidyl-prolyl cis-trans isomerase activity and important roles in protein folding and chaperone function. Viral CYNs may participate in viral replication

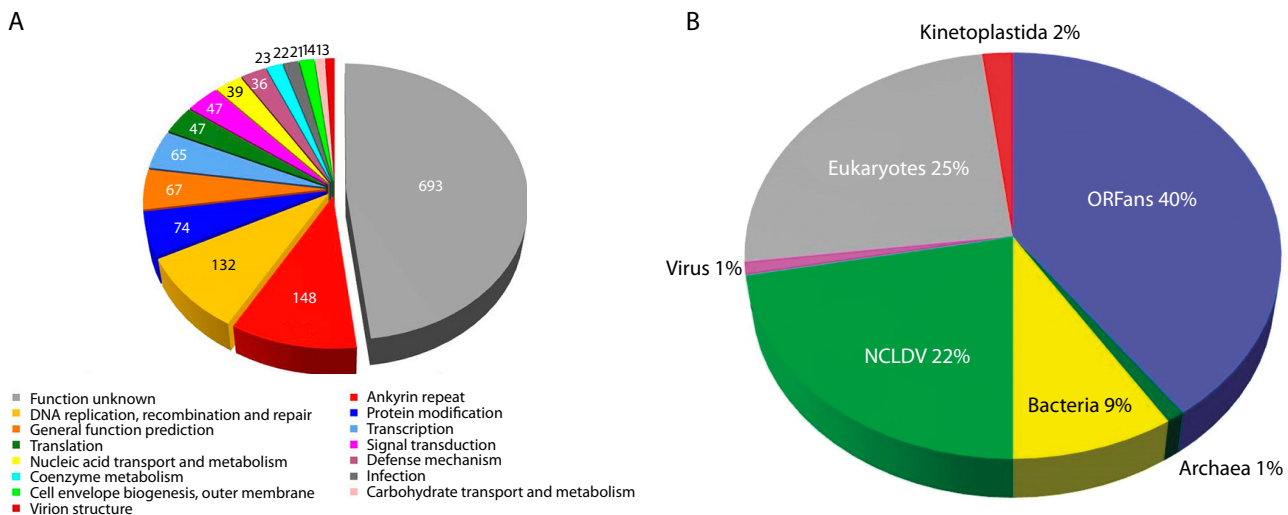
and host cell manipulation, indicating a complex interplay between virus and host. The origin of giant virus CYNs remains unclear and is hypothesized to involve lateral gene transfer among mimiviruses, potentially reflecting adaptation to the specific needs of individual viruses (Barik 2018). The study of viral CYNs is a growing field that promises to deepen our understanding of evolutionary mechanisms, functional specificity, and their potential roles in the life cycle of giant viruses and host–virus interactions.

An example of the surprising genomic plasticity of giant viruses is *Theiavirus salishense* (formerly Bodo saltans virus, BsV), which possesses one of the largest sequenced genomes within the *Mimiviridae* family and is among the most abundant members of this family in aquatic ecosystems (Figure 3) (Deeg et al. 2018). Its 1.39 Mb genome encodes 1,227 predicted ORFs, including complex replication machinery. Notably, BsV has lost much of its translational apparatus, including all tRNAs. Essential genes are invaded by homing endonuclease – encoding self-splicing introns, which may function as a defense mechanism against competing viruses. The presence of putative antihost factors, accompanied by extensive gene duplication via a “genomic accordion” mechanism, suggests an ongoing evolutionary arms race. This rapid evolution and genomic plasticity have contributed to the phenomenon of genome gigantism and highlight the unique nature of giant viruses (Deeg et al. 2018).

### General genome architecture of *Allomimiviridae*

The *Allomimiviridae* family, part of the recently classified *Imitervirales* order, shares many characteristics with other giant viruses, particularly those in the *Mimiviridae* family. This family includes species such as *Heliosvirus raunefjordenense*, formerly known as Pyramimonas orientalis virus 01b (PoV-01b), and the recently cultivated *Oceanusvirus kaneohense*, formerly known as Tetraselmis virus (TetV). Both are marine viruses that infect green algae – specifically *P. orientalis* and *Tetraselmis* sp., respectively (Aylward et al. 2023).

The genomic structure of *Allomimiviridae* is consistent with other mimiviruses in terms of gene content and organization. It is complex and contains a wide array of functional genes. For instance, PoV-01b has a genome size of 560 kbp and encodes structural proteins, DNA and RNA polymerases, helicases, tRNA synthetases, and



**Figure 3.** BsV genome content: **A)** Functional assignment of BsV genome content. **B)** Domain of best BLASTp hits (adapted from Deeg et al. 2018)

proteins involved in host interactions. The virus particles measure approximately  $220 \times 180$  nm, with the major polypeptide being 44 kDa in size (Sandaa et al. 2001).

Among the notable genome features of viruses infecting *P. orientalis* are the channelrhodopsin (ChR) genes. These genes, homologous to those found in the host algae and other organisms, suggest recent acquisition from a close relative within the prasinophyte green algal classes: *Pyramimonadophyceae*, *Mamiellophyceae*, and *Nephroselmidophyceae* (Gallot-Lavallée and Archibald 2020; Rozenberg et al. 2020). These viral ChRs, identified through bioinformatic analysis of microbial metagenomic data and their algal homologs, act as anion-conducting channels. They may enhance viral replication and spread by manipulating the host's cellular processes, such as membrane potential and ion flow, potentially affecting the host's swimming behavior (Rozenberg et al. 2020). The growing number of viral rhodopsins identified among NCLDVs raises questions about their role during infection, including whether they modify host phototaxis or disrupt membrane potential – possibly leading to cell death (Oppermann et al. 2023).

Although the structure, function, and role of viral rhodopsins in host protist infection remain unclear, the distinct functional properties of these ChRs – such as their light activation spectra and ion selectivity – present intriguing targets for further biochemical and biophysical studies. The presence of ChR genes in *P. orientalis*

viruses highlights how viruses can acquire and potentially repurpose host genes to manipulate cellular processes. These viral ChRs not only provide insights into virus–host interactions but also hold promise for developing new tools in optogenetics. Further research into their functional properties and potential applications could deepen our understanding of their roles in viral biology and their technological uses.

Tetraselmis virus 1 (TetV) infects the cosmopolitan green alga *Tetraselmis* (*Chlorodendrophyceae*), with virions measuring on average between  $226 \pm 9$  and  $257 \pm 9$  nm. Its circular dsDNA genome spans 668 kb and contains 663 predicted genes, including 653 coding DNA sequences (CDSs) and 10 tRNA genes within an 879-bp region (Schvarcz and Steward 2018). In addition to genes involved in replication, recombination, transcription, and translation, TetV encodes six serine/threonine protein kinases associated with post-translational modification and signal transduction (Schvarcz and Steward 2018). The transcription-related and translation-related genes are highly similar to those found in other algal-infecting mimiviruses. The genome also features a functional potassium ( $K^+$ ) channel – similar to other representatives of *Nucleocytoviricota* – which may confer an advantage during infection of algal hosts (Kukovetz et al. 2020).

Many of TetV's predicted proteins are novel, showing no similarity to known proteins, and include a significant number of eukaryote-derived proteins as well as proteins similar to those found in viruses and bacteria



(Schvarcz and Steward 2018). Of the 653 predicted protein-coding genes in the TetV genome, only 192 have been functionally annotated. Among these are genes encoding two key fermentation enzymes – pyruvate formate-lyase and pyruvate formate-lyase activating enzyme – proteins with unique functions not previously identified in viruses. These enzymes are involved in cellular fermentation pathways that enable energy production in the absence of oxygen. This is particularly relevant because *Tetraselmis*, the only known host of TetV, utilizes anaerobic energy metabolism under low-oxygen conditions, suggesting a possible case of horizontal gene transfer (HGT) (Brahim Belhaouari et al. 2022). Alternatively, this gene distribution may reflect differential gene loss in other eukaryotic lineages (Ku et al. 2015; Bremer et al. 2025). Notably, *Tetraselmis* and other green algae possess some of the most complete, vertically inherited sets of genes for anaerobic energy metabolism among eukaryotes (Gould et al. 2019), suggesting that these genes may have been retained rather than acquired. Furthermore, the presence of these enzymes implies that TetV can manipulate the fermentation pathway of infected algal cells, alter host metabolic dynamics, and enhance viral replication and spread – potentially leading to significant ecological consequences.

Other unique genes identified include those encoding  $\alpha$ -galactosidase and mannitol 1-phosphate dehydrogenase (M1PHD), which are involved in saccharide degradation and mannitol metabolism, respectively (Schvarcz and Steward 2018). Additionally, TetV possesses genes related to ubiquitination, methylation, DNA repair, and a unique eukaryotic initiation factor, eIF-1A, not observed in other algal-infecting mimiviruses (Iyer et al. 2006; Schvarcz and Steward 2018).

Overall, the genomes of *Allomimiviridae*, similar to those of other giant viruses in the phylum *Nucleocytoviricota*, are mosaic in nature and contain genes from a variety of sources – including other viruses, bacteria, archaea, and eukaryotes. This evidence of HGT underscores a long history of co-evolution between these viruses and their eukaryotic hosts, contributing to the large and diverse gene repertoires seen in giant viruses.

### General genome architecture of *Mesomimiviridae*

*Mesomimiviridae*, a recently recognized group within the order *Imitervirales*, is particularly prevalent in aquatic systems (Aylward et al. 2023; Farzad et al. 2022).

This family includes viruses such as *Tethysvirus hollandense* (formerly *Phaeocystis globosa* virus isolates, PgV), *Tethysvirus ontarioense* (formerly *Chrysochromulina parva* virus BQ2, CpV-BQ2), and *Tethysvirus raunefjordenense* (formerly *Chrysochromulina ericina* virus CeV-01B, CeV), which infect haptophytes – a group of unicellular algae (Aylward et al. 2023).

The genomes of *Mesomimiviridae* are characterized by their moderate size (~500 kb) compared to other members of *Imitervirales*. Approximately 70% of the linear dsDNA genome of PgV is conserved with other large dsDNA viruses and includes core genes essential for viral replication and repair, such as methyltransferases and transposases (de Oliveira et al. 2022). These genomes also encode capsid proteins and other structural components necessary for virus assembly. Of the 434 putative CDSs identified in the PgV genome, only 150 (34.6%) have predicted functions based on homology – reflecting the typical proportion of uncharacterized genes in large eukaryotic viruses. Unique enzymes, such as phospholipase and asparagine synthetase homologs, are also encoded within these genomes (Santini et al. 2013).

Genes encoding proteorhodopsin, identified in the genomes of PgV, provide another example of viral manipulation of host primary metabolism. The origins of viral proteorhodopsin genes remain unknown, but they may have been captured from their unicellular eukaryotic hosts (Kukovetz et al. 2020). In addition to DNA-manipulating enzymes common to most giant viruses, PgV possesses DNA repair enzymes such as an apurinic/apyrimidinic exonuclease and a mismatch repair enzyme, both involved in correcting errors that occur during DNA replication (Endo et al. 2020; Santini et al. 2013). The PgV genome is also rich in putative ribonucleases, including ribonuclease HII, an RNase T-like enzyme, and a PIN-domain-containing RNase H. These enzymes contribute to the removal of RNA primers during DNA replication, the repair of RNA–DNA hybrids, and possibly RNA processing and degradation. The specific role of the asparagine synthase genes identified in *Phaeocystis globosa* virus during infection remains unclear (Farzad et al. 2022).

The remaining members of *Mesomimiviridae*, such as *Tethysvirus ontarioense* (CpV-BQ2) and *Tethysvirus raunefjordenense* (CeV) – marine viruses infecting *Chrysochromulina* species – share similarly large dsDNA

genomes and complex gene content but differ in their specific host ranges. CeV, for example, has a 474-kb genome containing 512 putative protein-coding genes and 12 tRNA genes (Gallot-Lavallée et al. 2017; Kyrychenko et al. 2023). Many of these proteins have no significant matches in the NCBI nonredundant sequence database, with the majority of identified homologs being viral. Among the 293 predicted proteins with database homologs, 75.4% matched best to eukaryote-infecting large dsDNA viruses. The best nonviral matches were distributed among bacteria, eukaryotes, and host haptophytes, suggesting possible HGT and the inheritance of genes from common ancestors of eukaryote-infecting large dsDNA viruses (Kyrychenko et al. 2023). Among the 305 proteins unique to CeV (i.e., without recognizable homologs in the other *Mimiviridae*, including 68% ORFans), only 106 (35%) have a database homolog, of which only 52 are associated with functional attributes. Unique features of the CeV genome include the presence of DNA repair nucleases, such as ERCC4, useful for repairing UV-induced genome damage (Gallot-Lavallée et al. 2017).

CeV and PgV share homologs of the cold shock protein, which is known to act as an RNA chaperone. Additionally, CeV encodes inteins – mobile genetic elements that function as protein introns, removing themselves from host proteins through autocatalytic excision and protein splicing. It is unusual not only that these genes are present in the viral genome, but also that there are so many (eight inteins), which are strongly biased toward DNA-processing enzymes (Gallot-Lavallée et al. 2017). The presence of unique elements such as ERCC4, a DNA repair nuclease, and multiple inteins involved in protein splicing highlights the complexity and distinct adaptations of these viruses.

CpV-BQ2, like other members of the *Mesomimiviridae* family, possesses a mosaic gene composition derived from various sources – including viruses, bacteria, and eukaryotes. As with other giant viruses, a significant fraction of its genes have no readily detectable homologs. The 437-kb genome encodes 503 ORFs, which include typical NCLDV proteins involved in core viral functions, as well as unique gene clusters such as those encoding E3 ubiquitin ligases. These ligases may inhibit host defense mechanisms, highlighting the virus's complex strategies for evading host immunity. Additionally, CpV-BQ2's gene repertoire includes DNA and chromatin modification enzymes, such as DNA methyltransfe-

rases and histone modification enzymes, which likely target host chromatin (Stough et al. 2019).

In addition to a conserved set of genes common to NCLDVs, the CpV-BQ2 genome contains genes involved in central carbon metabolism and light harvesting. Notably, the CpV genome hosts the genomes of three distinct virophages, which co-occur with the virus and likely exploit CpV for their replication (Tokarz-Deptuła et al. 2023). These virophages encode a variety of genes, including both functionally characterized proteins and others with uncharacterized or unexpected functions. The identification of chlorophyll-binding proteins and genes related to the TCA cycle in *Mesomimiviridae* members suggests that these viruses may manipulate host metabolic pathways, particularly those involved in energy production and photosynthesis (Farzad et al. 2022).

For a detailed description of the predicted genes of viruses infecting *T. hollandense* and *Chrysochromulina* species (haptophytes), refer to the NCLDV metagenome-assembled genomes (MAGs) list, which compiles data from environmental metagenomic datasets collected globally. These MAGs provide the best matches to transcripts of chrysophytes, further elucidating the genomic complexity and diversity within the *Mesomimiviridae* family (Endo et al. 2020).

Like other viruses in the order *Imitervirales*, members of the *Mesomimiviridae* family generally possess genomes with complex and diverse gene content. In addition to their primary functions, these viruses contain genes encoding structural proteins and enzymes involved in amino acid metabolism, DNA processing, central carbon metabolism, and light harvesting. These features suggest that *Mesomimiviridae* may employ various strategies to manipulate host cellular physiology during infection (Farzad et al. 2022).

### General genome architecture of *Schizomimiviridae*

The *Schizomimiviridae* family, previously referred to as the *Aureococcusvirus* group, includes two representative species: *Biavirus raunefjordenense* and *Kratosvirus quantuckense*, formerly known as Pymnesium kappa virus RF01 (PkV-RF01) and *Aureococcus anophagefferens* virus BtV-01 (AaV), respectively (Aylward et al. 2023). These marine giant viruses infect haptophyte and heterokont hosts and exhibit genomic features and host interactions that distinguish them from other members of *Imitervirales*. The presence of hallmark genes such



as DNA polymerase B (PolB) and the major capsid protein (MCP) places these viruses within the NCLDV group and reflects their conserved roles in replication and virion structure (Yutin and Koonin 2012; Johannesen et al. 2014).

*Biavirus raunefjordenense* (PkV-RF01) has a linear DNA genome of approximately 1.42 Mb, comprising 1,161 genes, including 1,121 coding sequences and 40 tRNA genes. Notably, PkV-RF01 contains a complete base excision repair (BER) pathway – a feature also observed in some *Mimiviridae* infecting heterotrophs – suggesting enhanced capabilities for DNA repair and protein synthesis (Blanc-Mathieu et al. 2021).

PkV-RF01 is distinguished by an extensive array of genes involved in energy metabolism, including those associated with the tricarboxylic acid (TCA) cycle and the  $\beta$ -oxidation pathway (Blanc-Mathieu et al. 2021). Among these, the gene encoding succinate dehydrogenase subunit A is actively transcribed during infection, highlighting the virus's potential to regulate host energy metabolism. Furthermore, PkV-RF01 encodes a substantial number of glycosyltransferases and other carbohydrate-active enzymes – surpassing the numbers found in many other viruses (Blanc-Mathieu et al. 2021).

Initially identified as *Aureococcus anophagefferens* virus (AaV), *Kratosvirus quantuckense* was isolated in 2002 and sequenced twice – in 2014 and 2022 – to better unravel its functionality by determining its genetic potential (Truchon et al. 2023). According to the first sequencing, *Kratosvirus quantuckense* has a relatively small dsDNA genome of approximately 370.92 kb, encoding 377 putative coding sequences and exhibiting a genome coding density of 88.3%, typical of large dsDNA viruses. The genome harbors NCLDV-specific core genes and a substantial proportion (71.35%) of unique genes, many of which appear to have been acquired horizontally from the host and other sources. These include genes involved in DNA replication, transcription, and translation (Truchon et al. 2023).

AaV, along with two other photosynthetic protozoan viruses – *Phaeocystis globosa* virus (PgV) and *Chrysochromulina ericina* virus (CeV) – shares a small set of seven genes that distinguish them from other members of the *Mimiviridae* family. Among these shared genes, only one has a predicted function: an ERCC4-type DNA repair nuclease. This enzyme is typically associated with the cellular response to UV-induced DNA

damage, suggesting a potential role in protecting the viral genome from such damage. The presence of this specific gene among these viruses may indicate an evolutionary adaptation to environments with high UV exposure (Gallot-Lavallée et al. 2017).

The limited number of shared genes among AaV, PgV, and CeV – despite their common infection of photosynthetic protozoan hosts – suggests that their genomic features are not primarily dictated by lifestyle or host type. This points to a significant degree of genomic diversity and specialization within these viruses, possibly driven by factors such as ecological niche, evolutionary history, or specific host–virus interactions.

According to genomic sequencing results from 2022, the AaV genome is 381,717 bp long, with 384 CDSs and an additional 5-kb region located between the previously predicted polar ends of the reference genome (Truchon et al. 2022). Other observed changes include apparent duplications, gene elongations, and gene fusions missed during the initial assembly (Truchon et al. 2022). A notable feature of the AaV genome is the presence of two subunits of DNA-directed RNA polymerase 2 (Rpb2), likely resulting from an ancient gene duplication event. This duplication may contribute to the virus's functional diversification and adaptation (Truchon et al. 2022). Although the role and frequency of gene duplication events remain unclear, their occurrence suggests functional significance that may enhance viral fitness (Truchon et al. 2023). Overall, the genomic organization of AaV appears to reflect genome expansion driven by gene duplication and the acquisition of host-derived genes (Truchon et al. 2022, 2023).

AaV also contains unique genes for pectate lyases – typically involved in plant tissue degradation – and concanavalin A-like lectin/glucanase genes, suggesting roles in carbohydrate metabolism during infection (Truchon et al. 2022, 2023). These nonglycosyltransferase carbohydrate metabolism genes are hypothesized to play important roles during infection. The packaging of one pectate lyase and three lectin/glucanases within the viral particle underscores their importance, particularly for metabolic processes initiated immediately upon host invasion (Truchon et al. 2022, 2023). Additionally, AaV includes a viral tRNA<sup>Tyr</sup> gene with a 20 bp intron in the canonical position between nucleotides 37 and 38 of the tRNA precursor – a characteristic shared with

**Table 1.** Basic characteristics of NCLDV infecting protists

Family	Genome size (kb)	Capsid: symmetry/size (nm)/ features	Natural or experimental host	Unique genes involved in (related to)
<i>Allomimiviridae</i>	500–600	Ico/220–260/ glycosylated fibrils, protrusions	Chlorophyta	host interaction and metabolism, fermentation, anion-conducting channelrhodopsins
<i>Mimiviridae</i>	1020–1260	Ico/300–500/ heavily glycosylated fibrils	Amoebae SAR supergroup Kinetoplastida Stramenopiles	translation, DNA repair, carbohydrate metabolism
<i>Schizomimiviridae</i>	370–1420	Ico/250–300/ long surface fibrils	Haptophyta SAR supergroup Pelagophytes Heterokonts	glycosyltransferases, enzymes for energy metabolism
<i>Mesomimiviridae</i>	430–600	Ico/300–400/ glycoproteins, fibrils	Haptophyta ( <i>Chrysochromulina</i> )	DNA and chromatin modification enzymes, energy metabolism
<i>Phycodnaviridae</i>	160–560	Ico/130–150/ lipid envelope in some members	Alveolata Chlorophyta Haptophyta Stramenopiles	photosynthesis, glycosyltransferases, dinoflagellate/ viral nucleoproteins, host cell apoptosis, signaling pathways, transcription, DNA polymerase
<i>Mamonoviridae</i>	360–382	Ico/~260/ surface spikes	Amoebae ( <i>Acanthamoeba</i> )	full set of histone proteins, lack two of the core genes – RNA polymerase and DNA topoisomerase II
<i>Marseilleviridae</i>	300–400	Ico/200–400/ surface protrusions and glycoproteins	Amoebae ( <i>Acanthamoeba</i> )	cell signaling, metabolism
<i>Cedraviridae</i>	460–600	Ico/150–200/ surface proteins, two apex corks	Amoebae ( <i>Acanthamoeba</i> )	DNA replication, repair
<i>Orpheoviridae</i>	1474	Ico/900–1300/ ostiole at the apex	Amoebae ( <i>Acanthamoeba</i> )	metabolic and regulatory genes
<i>Pithoviridae</i>	600–700	Ico/1500/fibers, single apex cork projections	Amoebae ( <i>Acanthamoeba</i> )	transcription and translation

\*Ico – icosahedral symmetry

most bacterial, plastid, eukaryotic, and archaeal tRNA-intronic RNAs (Bhattacharjee et al. 2023).

The AaV genome is characterized by a significantly lower GC content (28.7%) compared to its host's (69.9%), suggesting adaptations that support nucleotide salvage pathways (Truchon et al. 2023). Moreover, compared to other *Nucleocytoviricota*, the virus encodes an unusually large number of methyltransferases – eight, according to the initial annotation of *Kratosvirus quantuckense* – which may function in mRNA capping and DNA methylation, highlighting the complexity of its

host-virus regulatory mechanisms (Truchon et al. 2022). Understanding the specific genetic traits and functional capabilities of *K. quantuckense* is essential for developing innovative molecular tools and bioinformatics methods to define similar uncultured viruses and study viruses across diverse environments.

## Conclusions

The genomes of giant viruses are remarkably diverse, encompassing genes responsible for a wide array of functions, including protein translation, carbohydrate

and lipid metabolism, nitrogen cycling, and light assimilation. These capabilities suggest that giant viruses can significantly alter the metabolic processes of their hosts, affecting pathways such as glycolysis, gluconeogenesis, and the tricarboxylic acid cycle. Such modifications highlight the complex interactions these viruses have with their host organisms.

Table 1 presents the main characteristics of NCLDV infecting protists, including genome size, capsid properties, host range, and unique viral genes. A core set of genes, conserved across various families of giant viruses, encodes essential proteins for key life-cycle processes. The selection of markers used to assign a virus to the NCLDV group and classify it within the group remains an active area of research. In addition to the diverse set of genes used to determine evolutionary relationships, researchers employ different analytical strategies. The use of single-gene markers – such as the major capsid protein or DNA polymerase – provides only moderate resolution, particularly at higher taxonomic levels, and may be affected by horizontal gene transfer or recombination events (Larsen et al. 2008; Chen and Suttle 1996; Rowe et al. 2011).

In contrast, multigene approaches using sets of 5–50 conserved genes involved in major viral functions provide higher phylogenetic resolution, especially for reconstructing deep evolutionary splits and preserving the monophyly of major NCLDV clades (Koonin and Yutin 2010; Aylward et al. 2021; Sun and Ku 2021). Among these conserved markers, genes such as *PolB* (DNA polymerase family B), *RNAPL* (RNA polymerase large subunit), A32-like packaging ATPase, topoisomerase II, *VLTF3*, *TFIIB*, and superfamily II helicase are frequently used due to their length, conservation, and functional significance. Some studies have also employed whole-proteome approaches, which offer high phylogenetic resolution but require complete or near-complete genome sequences (Yu et al. 2010).

A subgroup of ten highly conserved key genes (phylogenetic markers) has been proposed as a reliable phylogenetic framework for comparative analysis, lineage classification, and evolutionary studies of NCLDVs. These include the A2L-like transcription factor, A32-like packaging ATPase, D5-like helicase-primase, elongation subunit of B-family DNA polymerase, helicase II, large subunit of the mRNA capping enzyme, myristoylated envelope protein, small subunit of ribonucleotide reduc-

tase, and the  $\alpha$ -subunit and  $\beta$ -subunit of RNA polymerase (Athira and Antony 2023; Campillo-Balderas et al. 2023).

Beyond conserved core genes and established phylogenetic markers, other genomic elements – such as ORFans and virophages – further contribute to the diversity and evolutionary potential of NCLDVs. ORFans – genes without known homologs in viruses infecting unicellular eukaryotes, particularly within the NCLDV group – add to the functional complexity and unique characteristics of these viruses. Virophages, by integrating into giant virus genomes, can influence viral functionality through mechanisms such as horizontal gene transfer, modulation of viral replication, and alteration of host–virus dynamics. These interactions can drive viral evolution, contributing to the complexity and adaptability of giant viruses.

Recent studies show that gene expression in giant viruses follows a well-regulated temporal pattern and occurs in three phases: early (DNA replication, host invasion), intermediate (nucleotide metabolism), and late (morphogenesis and structural protein synthesis) (Moniruzzaman et al. 2018; Bessenay et al. 2024; Chen et al. 2024). In each phase, the virus activates specific groups of genes that control different steps of the infection cycle: early genes initiate DNA replication and host invasion, intermediate genes support nucleotide synthesis, and late genes control proper virion assembly. This stepwise expression program illustrates the complex interplay between the virus and host machinery, as well as the virus's ability to manipulate host processes throughout infection (Moniruzzaman et al. 2018; Chen et al. 2025).

During the infectious cycle, the virus not only reprograms the expression of its own genes but also induces large-scale changes in the host cell transcriptome. According to Nuri et al. (2022), *Mimivirus* infection in *Acanthamoeba polyphaga* triggers cell cycle arrest and complex intracellular modifications that support viral replication. In addition to interfering with cell cycle progression and redirecting host metabolic pathways, the virus induces morphological changes – such as cell rounding and loss of pseudopodia – by downregulating cytoskeleton-related genes and upregulating genes involved in secretory pathways, peroxisomes, and the ubiquitin–proteasome system. These changes facilitate viral replication and may contribute to viral factory formation through translocation of endoplasmic

reticulum membranes, reorganization of the cytoskeletal network, and sustained cell cycle arrest.

This already complex expression program may be further influenced by additional factors. In the Megamimivirinae subfamily, for instance, transcription is modulated by hyperparasitic virophage co-infections and transpovirons – plasmid-like linear dsDNA molecules – all of which share common transcriptional regulatory elements (promoter motifs) (Ku et al. 2020; Bessenay et al. 2024; Bosmon et al. 2025). Interestingly, virophage co-infection transiently alters giant virus gene expression, while transpovirons appear to play a commensal role with no significant impact, suggesting a rich network of interactions. These close transcriptional-level interactions affect each partner in this complex relationship and influence specific cellular functions.

Moreover, environmental factors such as nutrient availability also influence viral transcription, adding another layer to the complexity of gene regulation (Bessenay et al. 2024). These new findings not only extend our understanding of viral replication but also raise several important questions: the existence of hidden layers of temporal gene regulation; the intricate transcriptional interactions among viruses, virophages, and transpovirons; the phenomenon of transient parasitism; and the broader ecological significance of these interactions.

An intriguing hypothesis regarding the formation of chimeric virus–host RNA polymerase complexes – composed of both viral and host subunits – also warrants further investigation, particularly in terms of protein composition and the cellular localization of viral and host components during infection (Bessenay et al. 2024). Further research is needed to assess how widespread this phenomenon is among giant viruses and to quantify the impact of environmental factors on gene expression.

Altogether, these findings emphasize the complexity of transcriptional regulation in giant viruses and their remarkable ability to manipulate host processes during infection. They also point to a promising field for future research, particularly through transcriptomic profiling and single-cell methods (Moniruzzaman et al. 2018; Ku et al. 2020; Bessenay et al. 2024). Continued investigation is likely to reveal new regulatory mechanisms and gene functions that shape the ecological roles of giant viruses across diverse environments.

This review underscores the significance of exploring the molecular architecture of viral genomes and proteomes. Such studies are crucial for understanding the origins, evolutionary mechanisms, and intricate virus–host relationships of giant viruses. Furthermore, these investigations have the potential to uncover novel viral functions and infection strategies, which could have important biotechnological applications. As microbial metagenome databases continue to expand, sequencing, comparative genomics, and phylogenetic analyses of giant virus genomes are essential for accurately identifying viral sequences. These efforts are key to delineating the boundaries between different orders and families of giant viruses and provide a foundation for understanding their evolutionary relationships and taxonomic classification.

### Acknowledgments

This work was supported by the European Molecular Biology Organization (EMBO) Solidarity Grant (SLG-5417). I would like to express my deepest gratitude to Prof. Dr. Anna Karnkowska (Institute of Evolutionary Biology, University of Warsaw, Poland) and Dr. Matthias G. Fischer (Max Planck Institute for Medical Research, Heidelberg, Germany) for their invaluable support and guidance throughout this project. Their insightful discussions, expert advice, and unwavering encouragement were instrumental in the successful completion of this work. This project would not have been possible without their generous assistance and profound knowledge.

### Conflict of interest

The author declares that there is no conflict of interest.

### AI disclosure statement

The author declares that no artificial intelligence tools were used to generate the scientific content of this manuscript. AI-based language or grammar assistance was utilized solely to improve readability and clarity.

### References

- Abrahão J, Silva L, Silva LS, Khalil JYB, Rodrigues R, Arantes T, Assis F, Boratto P, Andrade M, Kroon EG, et al. 2018. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat Commun.* 9: 749. <https://doi.org/10.1038/s41467-018-03168-1>.
- Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, et al. 2019. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *J Eukaryot Microbiol.* 66: 4–119. <https://doi.org/10.1111/jeu.12691>.
- Athira MT, Antony SP. 2023. The Tiny Giants: Overview of giant viruses. *Ecol Genet Genom.* 29: 100210. <https://doi.org/10.1016/j.egg.2023.100210>.
- Aylward FO, Abrahão JS, Brussaard CPD, Fischer MG, Moniruzzaman M, Ogata H, Suttle CA. 2023. Taxonomic

- update for giant viruses in the order Imitervirales (phylum Nucleocytoviricota). *Arch Virol*. 168: 283. <https://doi.org/10.1007/s00705-023-05906-3>.
- Aylward FO, Moniruzzaman M, Ha AD, Koonin EV. 2021. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol*. 19(10): e3001430. <https://doi.org/10.1371/journal.pbio.3001430>.
- Bandaru V, Zhao X, Newton MR, Burrows CJ, Wallace SS. 2007. Human endonuclease VIII-like (NEIL) proteins in the giant DNA Mimivirus. *DNA Repair (Amst)*. 6: 1629–1641. <https://doi.org/10.1016/j.dnarep.2007.05.011>.
- Barik S. 2018. A Family of Novel Cyclophilins, Conserved in the Mimivirus Genus of the Giant DNA Viruses. *Comput Struct Biotechnol J*. 16: 231–236. <https://doi.org/10.1016/j.csbj.2018.07.001>.
- Bessenay A, Bisio H, Belmudes L, Couté Y, Bertaux L, Claverie J, Abergel C, Jeudy S, Legendre M. 2024. Complex transcriptional regulations of a hyperparasitic quadripartite system in giant viruses infecting protists. *Nat Commun*. 15(1): 8608. <https://doi.org/10.1038/s41467-024-52906-1>.
- Bhattacharjee AS, Schulz F, Woyke T, Orcutt BN, Martínez Martínez J. 2023. Genomics discovery of giant fungal viruses from subsurface oceanic crustal fluids. *ISME Commun*. 3: 10. <https://doi.org/10.1038/s43705-022-00210-8>.
- Blanc-Mathieu R, Dahle H, Hofgaard A, Brandt D, Ban H, Kalinowski J, Ogata H, Sandaa R. 2021. A Persistent Giant Algal Virus, with a Unique Morphology, Encodes an Unprecedented Number of Genes Involved in Energy Metabolism. *J Virol*. 95: e02446-20. <https://doi.org/10.1128/jvi.02446-20>.
- Bosmon T, Abergel C, Claverie JM. 2025. 20 years of research on giant viruses. *NPJ Viruses* 3: 9. <https://doi.org/10.1038/s44298-025-00093-1>.
- Brahim Belhaoui D, Pires De Souza GA, Lamb DC, Kelly SL, Goldstone JV, Stegeman JJ, Colson P, La Scola B, Aherfi S. 2022. Metabolic arsenal of giant viruses: Host hijack or self-use? *eLife*. 11: e78674. <https://doi.org/10.7554/elife.78674>.
- Brandes N, Linial M. 2019. Giant Viruses – Big Surprises. *Viruses*. 11(5): 404. <https://doi.org/10.3390/v11050404>.
- Bremer N, Martin WF, Steel M. 2025. Surprising effects of differential loss in genome evolution: the last-one-out. *FEMS Microbiol Lett*. 372: fnaf051. <https://doi.org/10.1093/femsle/fnaf051>.
- Campillo-Balderas JA, Lazcano A, Cottom-Salas W, Jácome R, Becerra A. 2023. Pangenomic Analysis of Nucleo-Cytoplasmic Large DNA Viruses. I: The Phylogenetic Distribution of Conserved Oxygen-Dependent Enzymes Reveals a Capture-Gene Process. *J Mol Evol*. 91: 647–668. <https://doi.org/10.1007/s00239-023-10126-z>.
- Chen F, Suttle CA. 1996. Evolutionary Relationships among Large Double-Stranded DNA Viruses That Infect Microalgae and Other Organisms as Inferred from DNA Polymerase Genes. *Virology* 219(1): 170–178. <https://doi.org/10.1006/viro.1996.0234>.
- Chen J, Ogata H, Hikida H. 2025. Sputnik virophage disrupts the transcriptional regulation of its host giant virus. *J Virol*. 99(4): e0019225. <https://doi.org/10.1128/jvi.00192-25>.
- Claverie JM, Abergel C. 2018. Mimiviridae: an expanding family of highly diverse large dsDNA viruses infecting a wide phylogenetic range of aquatic Eukaryotes. *Viruses* 10: 506. <https://doi.org/10.3390/v10090506>.
- Claverie JM, Grzela R, Lartigue A, Bernadac A, Nitsche S, Vacelet J, Ogata H, Abergel C. 2009. Mimivirus and Mimiviridae: giant viruses with an increasing number of potential hosts, including corals and sponges. *J Invertebr Pathol*. 101: 172–180. <https://doi.org/10.1016/j.jip.2009.03.011>.
- Colson P, Ominami Y, Hisada A, La Scola B, Raoult D. 2019. Giant mimiviruses escape many canonical criteria of the virus definition. *Clin Microbiol Infect*. 25: 147–154. <https://doi.org/10.1016/j.cmi.2018.09.010>.
- Deeg CM, Chow CT, Suttle CA. 2018. The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. *eLife*. 7: e33014. <https://doi.org/10.7554/eLife.33014>.
- Dos Santos Silva LK, Rodrigues RAL, Dos Santos Pereira Andrade AC, Hikida H, Andreani J, Levasseur A, La Scola B, Abrahão JS. 2020. Isolation and genomic characterization of a new mimivirus of lineage B from a Brazilian river. *Arch Virol*. 165: 853–863. <https://doi.org/10.1007/s00705-020-04542-5>.
- Endo H, Blanc-Mathieu R, Li Y, Salazar G, Henry N, Labadie K, de Vargas C, Sullivan MB, Bowler C, Wincker P, et al. 2020. Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nat Ecol Evol*. 4: 1639–1649. <https://doi.org/10.1038/s41559-020-01288-w>.
- Farzad R, Ha AD, Aylward FO. 2022. Diversity and genomics of giant viruses in the North Pacific Subtropical Gyre. *Front Microbiol*. 13: 1021923. <https://doi.org/10.3389/fmicb.2022.1021923>.
- Fischer MG. 2016. Giant viruses come of age. *Curr Opin Microbiol*. 31: 50–57. <https://doi.org/10.1016/j.mib.2016.03.001>.
- Gallot-Lavallée L, Archibald JM. 2020. Evolutionary Biology: Viral Rhodopsins Illuminate Algal Evolution. *Curr Biol*. 30: R1469–R1471. <https://doi.org/10.1016/j.cub.2020.10.080>.
- Gallot-Lavallée L, Blanc G, Claverie JM. 2017. Comparative Genomics of Chrysochromulina ericina Virus and Other Microalga-Infecting Large DNA Viruses Highlights Their Intricate Evolutionary Relationship with the Established Mimiviridae Family. *J Virol*. 91: e00230-17. <https://doi.org/10.1128/jvi.00230-17>.
- Gould SB, Garg SG, Handrich M, Nelson-Sathi S, Gruenheit N, Tielens AGM, Martin WF. 2019. Adaptation to life on land at high O<sub>2</sub> via transition from ferredoxin-to NADH-dependent redox balance. *Proc R Soc B Biol Sci*. 286(1909): 20191491. <https://doi.org/10.1098/rspb.2019.1491>.
- Ha AD, Aylward FO. 2024. Automated classification of giant virus genomes using a random forest model built on trademark protein families. *NPJ Viruses*. 2, 9. <https://doi.org/10.1038/s44298-024-00021-9>.
- Hurwitz BL, Hallam SJ, Sullivan MB. 2013. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol*. 14: R123. <https://doi.org/10.1186/gb-2013-14-11-r123>.
- ICTV 2025. Current ICTV Taxonomy Release. <https://ictv.global/taxonomy> (Accessed: 10.03.2025).
- Iyer LM, Balaji S, Koonin EV, Aravind L. 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res*. 117: 156–184. <https://doi.org/10.1016/j.virusres.2006.01.009>.

- Johannessen TV, Bratbak G, Larsen A, Ogata H, Egge ES, Edvardsen B, Eikrem W, Sandaa R. 2014. Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). *Virology* 476: 180–188. <https://doi.org/10.1016/j.virol.2014.12.014>.
- Kalafati E, Papanikolaou E, Marinos E, Anagnou NP, Pappa KI. 2022. Mimiviruses: Giant viruses with novel and intriguing features (Review). *Mol Med Rep.* 25: 207. <https://doi.org/10.3892/mmr.2022.12723>.
- Karki S, Moniruzzaman M, Aylward FO. 2021. Comparative genomics and environmental distribution of large dsDNA viruses in the family Asfarviridae. *Front Microbiol.* 12: 657471. <https://doi.org/10.3389/fmicb.2021.657471>.
- Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology.* 479–480: 2–25. <https://doi.org/10.1016/j.virol.2015.02.039>.
- Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn JH. 2020. Global organization and proposed megataxonomy of the virus world. *Microbiol Mol Biol Rev.* 84: e00061-19. <https://doi.org/10.1128/mmmr.00061-19>.
- Koonin EV, Yutin N. 2010. Origin and evolution of eukaryotic large nucleocytoplasmic DNA viruses. *Intervirology.* 53(5): 284–292. <https://doi.org/10.1159/000312913>.
- Koonin EV, Yutin N. 2019. Evolution of the large nucleocytoplasmic DNA viruses of Eukaryotes and convergent origins of viral gigantism. *Adv Virus Res.* 103: 167–202. <https://doi.org/10.1016/bs.aivir.2018.09.002>.
- Kostygov AY, Karnkowska A, Votýpka J, Tashyreva D, Maciszewski K, Yurchenko V, Lukeš J. 2021. Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol.* 11: 200407. <https://doi.org/10.1098/rsob.200407>.
- Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani-Covo E, McInerney JO, Landan G, Martin WF. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524(7566): 427–432. <https://doi.org/10.1038/nature14963>.
- Ku C. 2021. Giant Virus-Eukaryote Interactions as Ecological and Evolutionary Driving Forces. *mSystems* 6: 10.1128/msystems.00737-21. <https://doi.org/10.1128/msystems.00737-21>.
- Ku C, Sheyn U, Sebé-Pedrós A, Ben-Dor S, Schatz D, Tanay A, Rosenwasser S, Vardi A. 2020. A single-cell view on alga-virus interactions reveals sequential transcriptional programs and infection states. *Sci Adv.* 6(21): eaba4137. <https://doi.org/10.1126/sciadv.aba4137>.
- Kukovetz K, Hertel B, Schvarcz CR, Saponaro A, Manthey M, Burk U, Greiner T, Steward GF, Van Etten JL, Moroni A, et al. 2020. A Functional K<sup>+</sup> Channel from Tetraselmis Virus 1, a Member of the Mimiviridae. *Viruses* 12: 1107. <https://doi.org/10.3390/v12101107>.
- Kyrychenko A, Burkot V, Shcherbatenko I. 2023. Giant DNA Viruses Infecting Unicellular Protists. *Mikrobiol Zh.* 85: 72–82. <https://doi.org/10.15407/mikrobiolzh85.04.072>.
- La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie JM, Raoult D. 2003. A giant viruses in amoebae. *Science* 299: 203. <https://doi.org/10.1126/science.1081867>.
- Lad SB, Upadhyay M, Thorat P, Nair D, Moseley GW, Srivastava S, Pradeepkumar PI, Kondabagil K. 2023. Biochemical Reconstitution of the Mimiviral Base Excision Repair Pathway. *J Mol Biol.* 435: 168–188. <https://doi.org/10.1016/j.jmb.2023.168188>.
- Larsen JB, Larsen A, Bratbak G, Sandaa R. 2008. Phylogenetic Analysis of Members of the Phycodnaviridae Virus Family, Using Amplified Fragments of the Major Capsid Protein Gene. *Appl Environ Microbiol.* 74(10): 3048–3057. <https://doi.org/10.1128/aem.02548-07>.
- Moniruzzaman M, Gann ER, Wilhelm SW. 2018. Infection by a Giant Virus (AaV) Induces Widespread Physiological Reprogramming in *Aureococcus anophagefferens* CCMP1984 – A Harmful Bloom Algae. *Front Microbiol.* 9: 752. <https://doi.org/10.3389/fmicb.2018.00752>.
- Nuri R, Feldmesser E, Fridmann-Sirkis Y, Keren-Shaul H, Nevo R, Minsky A, Reich Z. 2022. *Acanthamoeba* polyphaga de novo transcriptome and its dynamics during Mimivirus infection. *bioRxiv* (Cold Spring Harbor Laboratory). <https://doi.org/10.1101/2022.07.20.500700>.
- de Oliveira EG, Carvalho JVRP, Botelho BB, da Costa Filho CA, Henriques LR, de Azevedo BL, Rodrigues RAL. 2022. Giant Viruses as a Source of Novel Enzymes for Biotechnological Application. *Pathogens* 11: 1453. <https://doi.org/10.3390/pathogens11121453>.
- Oppermann J, Rozenberg A, Fabrin T, González-Cabrera C, Béjré O, Prigge M, Hegemann P. 2023. Robust Optogenetic Inhibition with Red-light-sensitive Anion-conducting Channelrhodopsins. *eLife* 12: RP90100. <https://doi.org/10.7554/eLife.90100.2>.
- Queiroz VF, Tatara JM, Botelho BB, Rodrigues RAL, Almeida GMF, Abrahao JS. 2024. The consequences of viral infection on protists. *Commun Biol.* 7: 306. <https://doi.org/10.1038/s42003-024-06001-2>.
- Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM. 2004. The 1.2-megabase genome sequence of Mimivirus. *Science* 306: 1344–1350. <https://doi.org/10.1126/science.1101485>.
- Rappaport HB, Oliverio AM. 2023. Extreme environments offer an unprecedented opportunity to understand microbial eukaryotic ecology, evolution, and genome biology. *Nat Commun.* 14: 4959. <https://doi.org/10.1038/s41467-023-40657-4>.
- Rowe JM, Fabre M, Gobena D, Wilson WH, Wilhelm SW. 2011. Application of the major capsid protein as a marker of the phylogenetic diversity of *Emiliana huxleyi* viruses. *FEMS Microbiol Ecol.* 76(2): 373–380. <https://doi.org/10.1111/j.1574-6941.2011.01055.x>.
- Rozenberg A, Oppermann J, Wietek J, Fernandez Lahore RG, Sandaa RA, Bratbak G, Hegemann P, Béjré O. 2020. Lateral Gene Transfer of Anion-Conducting Channelrhodopsins between Green Algae and Giant Viruses. *Curr Biol.* 30: 4910–4920.e5. <https://doi.org/10.1016/j.cub.2020.09.056>.
- Sandaa RA, Heldal M, Castberg T, Thyraug R, Bratbak G. 2001. Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* 290: 272. <https://doi.org/10.1006/viro.2001.1161>.



- Santini S, Jeudy S, Bartoli J, Poirot O, Lescot M, Abergel C, Barbe V, Wommack KE, Noordeloos AA, Brussaard CP, et al. 2013. Genome of *Phaeocystis globosa* virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proc Natl Acad Sci USA* 110: 10800–10805. <https://doi.org/10.1073/pnas.1303251110>.
- Schulz F, Abergel C, Woyke T. 2022. Giant virus biology and diversity in the era of genome-resolved metagenomics. *Nat Rev Microbiol.* 20: 721–736. <https://doi.org/10.1038/s41579-022-00754-5>.
- Schvarcz CR, Steward GF. 2018. A giant virus infecting green algae encodes key fermentation genes. *Virology* 518: 423. <https://doi.org/10.1016/j.virol.2018.03.010>.
- Sharma D, Coulibaly F, Kondabagil K. 2024. Mimivirus encodes an essential MC1-like non-histone architectural protein involved in DNA condensation. *bioRxiv*. <http://dx.doi.org/10.1101/2024.02.22.580433>.
- Sharma V, Colson P, Pontarotti P, Raoult D. 2016. Mimivirus inaugurated in the 21st century the beginning of a reclassification of viruses. *Curr Opin Microbiol.* 31: 16–24. <https://doi.org/10.1016/j.mib.2015.12.010>.
- Simmonds P, Adriaenssens EM, Zerbini FM, Abrescia NGA, Aiewsakun P, Alfenas-Zerbini P, Bao Y, Barylski J, Drosten C, Duffy S, et al. 2023. Four principles to establish a universal virus taxonomy. *PLoS Biol.* 21: e3001922. <https://doi.org/10.1371/journal.pbio.3001922>.
- Stough JMA, Yutin N, Chaban YV, Moniruzzaman M, Gann ER, Pound HL, Steffen MM, Black JN, Koonin EV, Wilhelm SW, et al. 2019. Genome and Environmental Activity of a *Chrysochromulina parva* Virus and Its Virophages. *Front Microbiol.* 10: 703. <https://doi.org/10.3389/fmicb.2019.00703>.
- Sun T, Ku C. 2021. Unraveling gene content variation across eukaryotic giant viruses based on network analyses and host associations. *Virus Evol* 7(2): 1–13. <https://doi.org/10.1093/ve/veab081>.
- Talbert PB, Armache K, Henikoff S. 2022. Viral histones: pickpocket's prize or primordial progenitor? *Epigenetics Chromatin* 15(1): 21. <https://doi.org/10.1186/s13072-022-00454-7>.
- Tang L. 2020. Learning the diversity of giant viruses. *Nat Methods* 17: 253. <https://doi.org/10.1038/s41592-020-0778-z>.
- Tokarz-Deptuła B, Chrzanowska S, Gurgacz N, Stosik M, Deptuła W. 2023. Virophages – Known and Unknown Facts. *Viruses* 15: 1321. <https://doi.org/10.3390/v15061321>.
- Truchon AR, Chase EE, Gann ER, Moniruzzaman M, Creasey BA, Aylward FO, Xiao C, Gobler CJ, Wilhelm SW. 2023. Kratosvirus quantuckense: the history and novelty of an algal bloom disrupting virus and a model for giant virus research. *Front Microbiol.* 14: 1284617. <https://doi.org/10.3389/fmicb.2023.1284617>.
- Truchon AR, Gann ER, Wilhelm SW. 2022. Closed, circular genome sequence of *Aureococcus anophagefferens* Virus, a lytic virus of a brown tide-forming alga. *Microbiol Resour Announc.* 11: e0028222. <https://doi.org/10.1128/mra.00282-22>.
- Weynberg KD, Allen MJ, Wilson WH. 2017. Marine prasinoviruses and their tiny plankton hosts: a review. *Viruses* 9: 43. <https://doi.org/10.3390/v9020043>.
- Wilhelm SW, Bird JT, Bonifer KS, Calfee BC, Chen T, Coy SR, Gainer PJ, Gann ER, Heatherly HT, Lee J, et al. 2017. A student's guide to giant viruses infecting small Eukaryotes: from *Acanthamoeba* to *Zooxanthellae*. *Viruses* 9: 46. <https://doi.org/10.3390/v9020046>.
- Yu Z, Chu KH, Li CP, Anh V, Zhou L, Wang RW. 2010. Whole-proteome phylogeny of large dsDNA viruses and parvoviruses through a composition vector method related to dynamical language model. *BMC Evol Biol.* 10(1): 192. <https://doi.org/10.1186/1471-2148-10-192>.
- Yutin N, Koonin EV. 2012. Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes. *Virol J.* 9: 161. <https://doi.org/10.1186/1743-422X-9-161>.