CHAPTER FIVE

Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism

Eugene V. Koonin[1], **Natalya Yutin**
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, United States
[1]Corresponding author: e-mail address: koonin@ncbi.nlm.nih.gov

Contents

Abstract

The Nucleocytoplasmic Large DNA Viruses (NCLDV) of eukaryotes (proposed order "Megavirales") comprise an expansive group of eukaryotic viruses that consists of the families *Poxviridae*, *Asfarviridae*, *Iridoviridae*, *Ascoviridae*, *Phycodnaviridae*, *Marseilleviridae*, *Pithoviridae*, and *Mimiviridae*, as well as Pandoraviruses, Molliviruses, and Faustoviruses that so far remain unaccounted by the official virus taxonomy. All these viruses have double-stranded DNA genomes that range in size from about 100 kilobases (kb) to more than 2.5 megabases. The viruses with genomes larger than 500 kb are informally considered "giant," and the largest giant viruses surpass numerous bacteria and archaea in both particle and genome size. The discovery of giant viruses has been highly unexpected and has changed the perception of viral size and complexity, and even, arguably, the entire concept of a virus. Given that giant viruses encode multiple proteins that are universal among cellular life forms and are components of the translation system, the

Advances in Virus Research, Volume 103
ISSN 0065-3527
https://doi.org/10.1016/bs.aivir.2018.09.002

quintessential cellular molecular machinery, attempts have been made to incorporate these viruses in the evolutionary tree of cellular life. Moreover, evolutionary scenarios of the origin of giant viruses from a fourth, supposedly extinct domain of cellular life have been proposed. However, despite all the differences in the genome size and gene repertoire, the NCLDV can be confidently defined as monophyletic group, on the strength of the presence of about 40 genes that can be traced back to their last common ancestor. Using several most strongly conserved genes from this ancestral set, a well-resolved phylogenetic tree of the NCLDV was built and employed as the scaffold to reconstruct the history of gene gain and loss throughout the course of the evolution of this group of viruses. This reconstruction reveals extremely dynamic evolution that involved extensive gene gain and loss in many groups of viruses and indicates that giant viruses emerged independently in several clades of the NCLDV. Thus, these giants of the virus world evolved repeatedly from smaller and simpler viruses, rather than from a fourth domain of cellular life, and captured numerous genes, including those for translation system components, from eukaryotes, along with some bacterial genes. Even deeper evolutionary reconstructions reveal apparent links between the NCLDV and smaller viruses of eukaryotes, such as adenoviruses, and ultimately, derive all these viruses from tailless bacteriophages.

1. INTRODUCTION

It is now commonly recognized that viruses are the most abundant biological entities on our planet (Suttle, 2005, 2007). The powerful and accurate microscopic methods developed during the last 2 decades show that, in many environments, the number of virus particles exceeds the number of cells by one to two orders of magnitude (Bergh et al., 1989; Weinbauer, 2004; Wommack and Colwell, 2000) although the virus to cell ratio varies extensively (Wigington et al., 2016). The total number of virus particles that exist on the planet at any given moment is conservatively estimated at the staggering 10^{31} (Suttle, 2007; Weinbauer, 2004, p. 1856). Closer to home, a single human body contains up to 10^{16} virus particles, which is two orders of magnitude more than the number of human cells themselves; primarily, these particles represent bacteriophages infecting the bacteria in our gut microbiome (Abeles and Pride, 2014; Barr, 2017; Ogilvie and Jones, 2015).

Viruses come in a multitude of shapes and forms, with respect to the structure and size of the virions, the nature of the genomic nucleic acid and the genome size or the number of genes. Unlike cellular organisms that all have double-stranded (ds) DNA genomes that are transcribed into structural and messenger RNAs, with the latter translated into proteins, viruses

employ all known forms of nucleic acids as their genetic material, and thus, greatly differ in their replication and expression strategies (Baltimore, 1971a, b; Koonin, 1991; Koonin et al., 2015). The two most common among the seven so-called Baltimore classes of viruses that are distinguished by the form of the genomic nucleic acids that is incorporated into virions are positive sense RNA viruses (in these simple viruses, the genome also serves as an mRNA) and dsDNA viruses (these use the same genomic strategy as cellular organisms). The RNA viruses dominate the eukaryotic part of the virosphere, whereas the dsDNA viruses are most prevalent in the prokaryotic part (Koonin et al., 2015).

The origin of this dichotomy between the prokaryotic and eukaryotic viromes remains an enigma. One can speculate that the nucleus and intranuclear defense mechanisms severely restrict the access to the host replication and transcription machineries to DNA viruses, whereas the eukaryotic cytosol, which is compartmentalized by intracellular membranes provides a relatively hospitable environment to RNA viruses (Koonin et al., 2006, 2015). Several groups of dsDNA viruses, however, have cleared these barriers and spread among eukaryotes. Some, such as papilloma, polyoma, adeno and herpesviruses, reproduce in the nucleus, whereas others adapted to a (mostly) cytoplasmic life style.

Among the dsDNA viruses that reproduce in the eukaryotic cytoplasm, by far the best characterized are poxviruses which include vaccinia virus (VACV), one of the favorite models of molecular virology, and variola virus, a devastating human pathogen (Moss, 2001). Another cytoplasmic dsDNA virus with some resemblance to poxviruses is African Swine Fever Virus (ASFV), the founding member of the family Asfarviridae and an important livestock pathogen (Tulman et al., 2009). Once the sequences of multiple complete genomes of large dsDNA viruses of eukaryotes have accumulated, a comparative genomic study has led to a seminal discovery: it has been shown that poxviruses, ASFV, and three families of viruses whose reproduction sites have not been unambiguously determined at the time, Iridoviridae, Ascoviridae, and Phycodnaviridae, shared many homologous genes and likely descended from a common viral ancestor (Iyer et al., 2001). More specifically, a crude reconstruction of genome evolution using the parsimony principle has indicated that about 50 genes in these 5 groups of viruses likely originated from the putative common ancestor. It has been accordingly concluded that these viruses comprised a monophyletic group, to the exclusion of other large dsDNA viruses of eukaryotes, such as herpesviruses and baculoviruses, which shared only a small fraction of the inferred ancestral

set of genes. The identified group of five virus families has been denoted Nucleocytoplasmic Large Eukaryotic DNA Viruses (NCLDV) (Iyer et al., 2001), a somewhat unwieldy name and acronym that were adopted to indicate that, unlike poxviruses and ASFV whose reproduction is confined to the cytoplasmic virus "factories" (Moss, 2001), there have been indications that iridoviruses and phycodnaviruses might go through a nuclear phase (Allen et al., 2006; Chinchar et al., 2017).

Back in 2001, when the NCLDV were recognized as a monophyletic group, the interest in the evolution of large dsDNA viruses had been, arguably, limited to a narrow group of virologists specialized in the study of the respective viruses (primarily, poxviruses, in practice). However, the status of these viruses changed abruptly with the isolation, in 2003 (La Scola et al., 2003), and genome sequencing, in 2004 (Raoult et al., 2004), of the mimivirus, the first giant virus to be discovered. The name Mimivirus comes from MImicking MIcrobe virus (La Scola et al., 2003), and the name is appropriate because this virus indeed resembles a bacterium, at least, in terms of the virion and genome size. In fact, the story of the Mimivirus started with years of futile attempts to grow on various laboratory media what appeared to be an unusual, "Chlamydia-like" bacterium (Birtles et al., 1997). Only after typical, although enormously large, virus-like particles have been discerned in those samples by electron microscopy, the breakthrough became possible, and the virus has been grown by co-cultivation with Acanthamoeba (La Scola et al., 2003). The Mimivirus broke away from almost all conventional notions of a virus (Raoult et al., 2004). Most strikingly, the Mimivirus defines the definition of a virus as a filterable infectious agent, a property that, since the days viruses were discovered in the late 19th century, had been thought to provide for easy separation of viruses from bacteria using 0.3 μm "sterilizing" filters (Beijerinck, 1898; Iwanowski, 1892; Loeffler and Frosch, 1897). However, Mimivirus particles are 0.7 μm (La Scola et al., 2003) and thus fail to pass these filters which imply that giant viruses would have been systematically missed in many experimental settings. The genome of the Mimivirus is commensurately huge, by the virosphere standards, and at 1.1 Mb is larger than the genomes of numerous parasitic bacteria and about the same size as the smallest genomes of free-living bacteria and archaea (Koonin, 2009). Last but not least, the Mimivirus encodes proteins that are not normally found in other viruses but are universal in cellular life forms, namely, key components of the translation system, in particular, seven aminoacyl-tRNA synthetases (aaRS) (Raoult et al., 2004). In terms of the gene content, the Mimivirus was found to

be a strange agglomeration of typical viral genes, in particular, all the core genes of the NCLDV, and genes that, at least traditionally, have been thought of as signatures of cellular life (Raoult et al., 2004). The conservation of the core NCLDV genes squarely placed the Mimivirus in the midst of this group (Iyer et al., 2006). However, the discovery of the signature cellular genes immediately fueled attempts to graft the Mimivirus onto the tree of (cellular) life. This initial phylogenetic analysis of the Mimivirus aaRS failed to conclusively identify affinity with either of the three domains of cellular life, leading to provocative ideas on the origin of giant viruses. It has been proposed that giant viruses descended, through reductive evolution, from a cellular ancestor that belonged to a fourth, still unidentified, but most likely, extinct domain of cellular life (Claverie and Abergel, 2009; Claverie et al., 2006, 2009; Raoult et al., 2004). How to reconcile the presence of the complete genetic core of the NCLDV and the genes for universal cellular genes in the giant virus genomes has not been clarified in the initial presentations of the fourth domain scenario. Nevertheless, this hypothesis, to a large extent, has shaped the discourse on giant viruses for the decade to follow. The concept of the fourth domain has been strongly promoted (Colson et al., 2011, 2012; Desnues et al., 2012; Nasir et al., 2012, 2017) but also hotly contested, both on technical and on more general biological grounds (Forterre et al., 2014; Moreira and Lopez-Garcia, 2005, 2015; Williams et al., 2011; Yutin et al., 2014). This discussion has even evolved into a quasi-philosophical debate on the nature of viruses (giant and, by extension, all), whether they should be considered "alive" or not, and accordingly, whether or not viruses could, in principle, belong in the tree of life (Koonin and Starokadomskyy, 2016; Lopez-Garcia, 2012; Moreira and Lopez-Garcia, 2009).

Following the seminal discovery of the Mimivirus, its numerous relatives as well as several new groups of the NCLDV, and in particular, giant viruses have been identified, primarily, by co-cultivation with Acanthamoeba, or in some case, other amoeba, but additionally, by occasional virus isolation from marine protists and by metagenomic methods (Abergel et al., 2015; Aherfi et al., 2016; Colson et al., 2017a, b; Fischer, 2016; Khalil et al., 2016). The size records set by the Mimivirus have been mercilessly crushed by Pandoraviruses that have virions of about 1 by 0.5 μm and genomes of more than 2.5 Mb (Legendre et al., 2018; Philippe et al., 2013), and by Pithoviruses that hold the current record of virion size (about 1.5 by 0.5 μm) albeit with considerably smaller genomes (Legendre et al., 2014). To add to the weirdness, the virions of both Pandoraviruses and Pithoviruses

have previously unseen, asymmetrical, amphora-like shapes, a far cry from typical icosahedral virions of most of the NCLDV (including the Mimiviruses), whereas Mollivirus has a spherical virion, also unique among the known viruses (Legendre et al., 2015). And then, viruses have been identified that encode not 7, like the Mimivirus, but 19 or even all 20 aaRS along with many other translation system components, almost all of it except for the ribosome (Abrahao et al., 2018; Schulz et al., 2017). However, notwithstanding their oddities, all these freaks of the virus world neatly fit within the NCLDV by virtue of the conservation of the signature core gene set (Abrahao et al., 2018; Koonin and Yutin, 2010; Schulz et al., 2017; Yutin et al., 2009, 2014). These findings buttress the monophyly of the NCLDV and have led to the proposal to codify the evolutionary coherence of this group of viruses by establishing an order "Megavirales" (Colson et al., 2012, 2013). So far, however, the proposal has not been approved by the International Committee for Taxonomy of Viruses (ICTV).

A discovery that has further boosted the interest in the giant viruses was the identification of the virophages, relatively small dsDNA viruses with genomes of about 20 kb (subsequently classified as the family *Lavidaviridae*) (Krupovic et al., 2016) that parasitize on the giant Mimiviruses, reproducing in the giant virus factories, partially or completely inhibiting their growth, and in some cases, protecting the host from the lethal giant virus infection (Claverie and Abergel, 2009; Desnues et al., 2012; Fischer and Hackl, 2016; Fischer and Suttle, 2011; La Scola et al., 2008). Although the virophages, in most respects, resemble satellite viruses that parasitize on many diverse viruses (Krupovic and Cvirkaite-Krupovic, 2011), this discovery has been widely perceived as additional evidence of a unique, cell-like character of giant viruses, and even of their status as life forms (Desnues et al., 2012; Pearson, 2008).

In this chapter, we outline the currently characterized diversity of the NCLDV and explore the phylogeny of the conserved core NCLDV genes that appear to have settled into a stable topology with the growth of the NCLDV genome collection. We trace the turbulent genome dynamics that accompanies the evolution of these viruses, and in particular, present evidence of multiple, convergent origins of giant viruses. We further summarize the observations on the evolution of the translation system components encoded by the NCLDV and show that these genes have been captured, primarily, from eukaryotic hosts, on multiple occasions during virus evolution, and not inherited from a fourth domain of cellular life. Finally, we analyze the bigger picture of the evolution of the virosphere,

with the focus on the position of the NCLDV in the global network of dsDNA viruses, and discuss the prospects for changing the taxonomic status of the NCLDV and the constituent groups of viruses.

2. THE EXPANDING DIVERSITY AND STABILIZING PHYLOGENY OF THE NCLDV

Presently, the NCLDV include eight virus families and several taxonomically unassigned groups (Table 1 and Fig. 1). The virus genomes differ in size by more than an order of magnitude, spanning the range from about 100 kb in some iridoviruses to about 2.5 Mb in Pandoraviruses. The genome size directly translates into the number of genes because the NCLDV have "wall-to-wall" genomes, with genes packed at a density of about 1.1 gene per kilobase (Table 1) (Koonin, 2009). The phylogeny of the NCLDV can be constructed using the same approach as routinely employed for cellular life forms, namely, by building phylogenetic trees for universal marker genes (Puigbo et al., 2009; Woese, 1987; Woese et al., 1990; Woese and Fox, 1977). For cellular life forms, these are rRNAs and universal protein components of the translation and transcription systems. In the case of the NCLDV, there are very few genes that are conserved in all members of the group, and that number is dwindling with the increase in diversity brought about by newly discovered viruses (Koonin and Yutin, 2010; Yutin et al., 2009). Therefore, there is little choice of markers for phylogeny construction. During the last few years, phylogenies of the NCLDV have been built for the growing collection of genomes and using either individual, nearly universal genes or concatenations of 3–7 genes (Abrahao et al., 2018; Deeg et al., 2018; Schulz et al., 2017; Yutin et al., 2009, 2013, 2014). The main features of the phylogeny have remained remarkably stable, apparently, indicating that the major branches of the NCLDV have already been robustly defined.

Fig. 1 shows a phylogeny obtained with an updated, representative set of the NCLDV, using a concatenation of five genes of which three are present in all NCLDV, whereas the major capsid protein is missing in Pandoraviruses and the packaging ATPase in Pithoviruses. The tree consists of three large branches: (1) families *Mimiviridae* and *Phycodnaviridae* along with Pandoraviruses, (2) families *Pithoviridae*, *Marseilleviridae*, *Iridoviridae*, and *Ascoviridae*, and (3) families *Poxviridae* and *Asfarviridae*, together with Faustovirus, Pacmanvirus, and Kaumoebavirus, the recently discovered protist-infecting relatives of asfarviruses. Technically, the tree is unrooted.

Table 1 The Diversity of the NCLDV: 8 Virus Families and Unaffiliated Groups

Virus Family/Group	Host Range	Genome Size Range (kb)	Virion Architecture	Replication Site
"Extended Mimiviridae" Mimiviridae Proposed subfamily "Klosneuvirinae" OLPG group	Acanthamoeba and, probably, other amoebae; algae, heterokonts	280–1570	Icosahedral	Cytoplasm
Phycodnaviridae	Green algae; algal symbionts of paramecia and hydras; heterokonts; Haptophyta	180–400	Icosahedral	Nucleus and cytoplasm
"Pandoraviridae" Mollivirus sibericum Pandoraviruses	Amoebae	650–2470	Spherical; Amphora-shaped	Nucleus and cytoplasm
Pithoviridae	Unknown protists	460–1470	Amphora-shaped	Cytoplasm
Marseilleviridae	Acanthamoeba; probably, also algae	360–380	Icosahedral	Nucleus and cytoplasm
Asco-* and *Iridoviridae	Invertebrates and non-mammalian vertebrates	100–290		
Ascoviridae	Insects, mainly, Noctuids	120–190	Ovoid	Nucleus and cytoplasm
Iridoviridae	Insects, cold-blooded vertebrates	100–290	Icosahedral	Nucleus and cytoplasm
"Extended *Asfarviridae*" Asfarviridae Faustoviruses Pacmanvirus Kaemoebavirus	Amoebae, mammals	170–470	Icosahedral	Cytoplasm
Poxviridae	Animals: vertebrates, insects	130–360	Brick-shaped, icosahedral intermediate	Cytoplasm

OLPG, Organic Lake-Pheocystis globose group of viruses within "extended *Mimiviridae*".

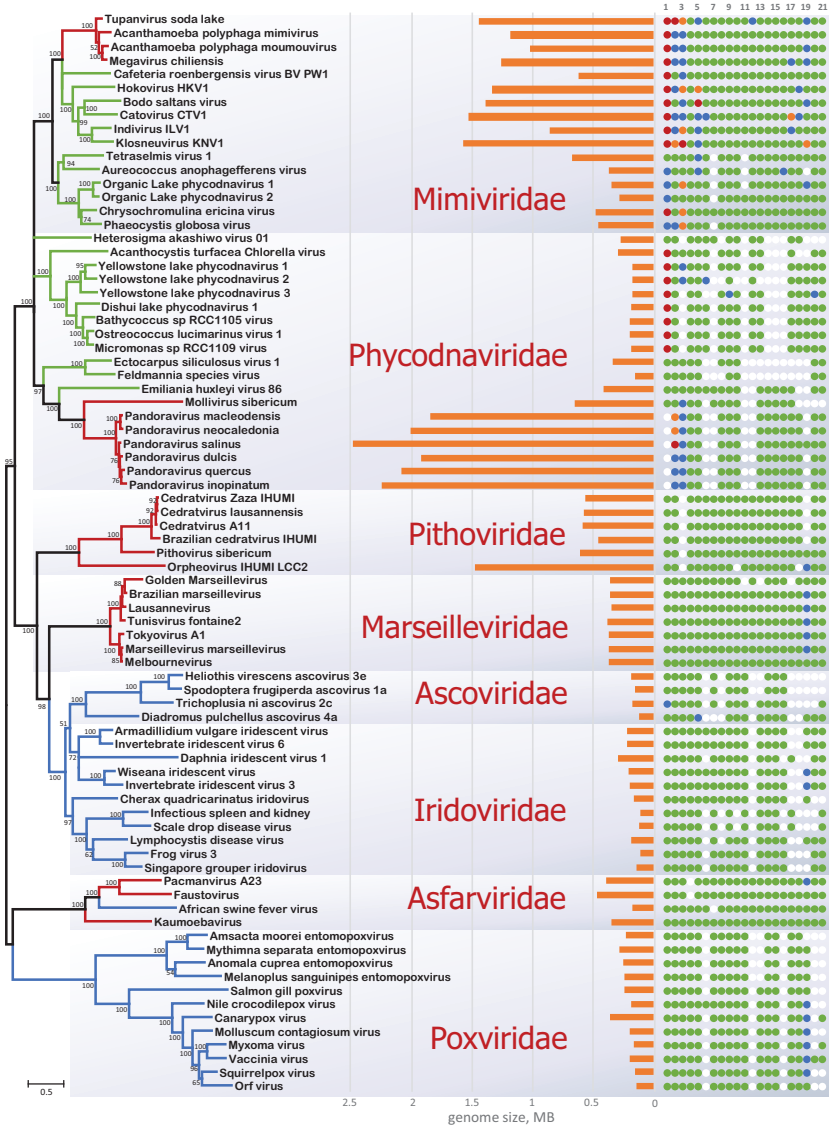

Fig. 1 Phylogenetic tree of nearly universal genes of the NCLDV. The tree was constructed from concatenated multiple alignment of five (nearly) universally conserved NCLDV proteins: DNA polymerase, major capsid protein, packaging ATPase, A18-like helicase, and Poxvirus Late Transcription Factor VLTF3. The branch color indicates confirmed or likely hosts: red, Amoebozoa; green, other protists; blue, Metazoa. The tree was constructed using the FastTree software (Price et al., 2010) with default parameters. The numbers at the internal branches indicate local likelihood-based support (percentage points); the branches with support below 50% were collapsed. Scale bars represent

(Continued)

However, considering additional evidence, namely, that the sequences of the core proteins of viruses in branches 1 and 2 are significantly more similar to each other than any of them are to those in branch 3, as well as the topology of individual phylogenetic that have been previously built for core NCLDV genes with cellular homologs as outgroups (such as DNA and RNA polymerases) (Yutin and Koonin, 2012), the root can be placed at branch 3 with reasonable confidence (Fig. 1).

Branch 1 consists of highly diverse viruses all of which infect various protists (Fig. 1 and Table 1). This branch includes most of the giant viruses that appear in two disjoint clades: (extended) *Mimiviridae* and Pandoraviruses (Fig. 1). Within this branch, the monophyly of two recently identified groups, each combining giant viruses with much smaller ones, is convincingly validated. The first of these groups that have been denoted "extended *Mimiviridae*" (Santini et al., 2013; Yutin et al., 2013) unites the giant viruses in the family *Mimiviridae* with a group of viruses with moderate-sized genomes, such as the Organic Lake phycodnaviruses and *Phaeocystis globosa* virus (OLPG; originally, these viruses have been mislabeled phycodnaviruses, apparently, because these viruses have been isolated form habitats dominated by algae; Yau et al., 2011). Notably, the OLPG clade that, mostly, consists of viruses with genomes in the range of 350–400 kb also includes its own "little giant," the Tetraselmis virus 1, with a 668 kb genome (Schvarcz and Steward, 2018). The second group includes the giant Pandoraviruses and *Mollivirus*

Fig. 1—Cont'd the number of amino acid (aa) substitutions per site. The middle panel shows genome length, on the scale shown in the bottom of the figure. The right panel shows the phyletic distribution of core NCLDV proteins: (1) NCLDV major capsid protein (NCVOG0022); (2) A32-like packaging ATPase (NCVOG0249); (3) Erv1/Alr family disulfide (thiol) oxidoreductase (NCVOG0052); (4) family B DNA polymerase (NCVOG0038); (5) D5-like helicase-primase (NCVOG0023); (6) DNA topoisomerase II (NCVOG0037); (7) FLAP-like endonuclease XPG (NCVOG1060); (8) DNA or RNA helicases of superfamily II (NCVOG0076); (9) Poxvirus late transcription factor VLTF3 (NCVOG0262); (10) Poxvirus late transcription factor VLTF2 (NCVOG1164); (11) Poxvirus early transcription factor VETF (NCVOG0261); (12) Transcription initiation factor IIB (NCVOG1127); (13) Transcription factor S-II (TFIIS) (NCVOG0272); (14) DNA-directed RNA polymerase subunit alpha (NCVOG0274); (15) DNA-directed RNA polymerase subunit beta (NCVOG0271); (16) DNA-directed RNA polymerase subunit 5 (NCVOG0273); (17) mRNA capping enzyme, guanylyltransferase (NCVOG1117); (18) mRNA capping enzyme, methyltransferase (NCVOG1117); (19) Nudix hydrolase (NCVOG0236); (20) ribonucleoside diphosphate reductase, large subunit (NCVOG1353); (21) ribonucleoside diphosphate reductase, beta subunit (NCVOG0276). Green, blue, orange, and red circles indicate, respectively, that one, two, tree, and four or more proteins of the respective family are encoded in the corresponding virus genome.

sibericum which form a clade with Coccolithoviruses as previously reported (Yutin et al., 2014; Yutin and Koonin, 2013). Remarkable features of this group are the unusual virions structures: the asymmetrical, amphora-like virion shape of Pandoraviruses, and the unusual, spherical virions of the Mollivirus. The Coccolithoviruses are generally classified within the family *Phycodnaviridae* (Wilson et al., 2009) but, in our current tree, fail to show affinity with the rest of phycodnaviruses. Finally, Heterosigma akashiwo virus, also considered to be a distinct phycodnavirus (Maruyama and Ueki, 2016), fails to join any of the groups within branch 1 and forms a distinct clade (Fig. 1).

Branch 2 unites the families *Pithoviridae*, *Marseilleviridae*, *Iridoviridae*, and *Ascoviridae*. The clustering of Marseillevirus with iridoviruses appeared highly unexpected when observed originally because of the apparent evolutionary affinity of viruses infecting protists with animal viruses (Boyer et al., 2009). Nevertheless, this affinity has been consistently supported by phylogenetic analysis of growing sets of genomes (Yutin et al., 2009, 2014) and buttressed by the discovery of the Pithoviruses which confidently joined the branch (Legendre et al., 2014). Given that Pithoviruses are at the base of the branch (Fig. 1), parsimony implies that the ancestor of branch 2 was a protist virus, whereas the ancestor of the Irido-Asco clade switched hosts to infect animals. The Pithoviruses have the largest virions among all known viruses, with an amphora (pithos) shape resembling that of Pandoraviruses (Andreani et al., 2016, 2017a, 2018; Legendre et al., 2014). The rest of the viruses in this branch has typical icosahedral capsids except for Ascoviruses with ovoid-shaped virions (Asgari et al., 2017). Of the three main branches in the NCLDV phylogeny, branch 2 includes the widest range of genome sizes, from about 100 kb in the smallest iridoviruses to more than 1.5 Mb in Orpheoviruses, a recently discovered member of the *Pithoviridae* (Andreani et al., 2017a).

Branch 3 consists of two clades: Asfarviruses (effectively, numerous strains of ASFV) joined by their larger, protist-infecting relatives, namely, Faustovirus (Reteno et al., 2015), Pacmanvirus (Andreani et al., 2017b), and Kaumoebavirus (Bajrai et al., 2016), and poxviruses. In a pattern that recapitulates those in branches 1 and 2, the viruses in the Asfar-like clade possess typical icosahedral capsids whereas Poxviruses have unique, brick-shaped virions (Moss, 2001). The switch from protists to animal hosts appears to have occurred twice during the evolution of this branch. These switches are likely to have taken place late in asfarviruses (at least, judging from the current knowledge that is limited to a group of closely related viruses that infect a

single mammalian species; Alonso et al., 2018) and early at the base of the poxvirus clade which consists of numerous viruses infecting both arthropods and vertebrates, two animal phyla that radiated from the common ancestor more than half a billion years ago (Moss, 2001). Unlike the other two branches of the NCLDV, branch 3 does not include any giant viruses although the protist viruses in the Asfar-like clade come close to the gigantism threshold (Fig. 1).

The phylogenies of (nearly) universal genes, like the one shown in Fig. 1, have an obvious, severe limitation. Formally, at least, they reflect only the evolution of those genes not that of the respective complete genomes and thus can be easily branded (near) useless "trees of 1%" as it has been done for the phylogenetic trees of universal genes of cellular life forms (Dagan and Martin, 2006). Nevertheless, in the case of cellular microbes (archaea and bacteria), it has been shown that, although much of the genome evolution is defined by gene gain, loss, and exchange and thus defines the tree representation, the phylogenies of universal genes reflect a central vertical trend in evolution and are useful, at least, as templates for evolutionary reconstructions (O'malley and Koonin, 2011; Puigbo et al., 2009). In the next section, we show that the same pertains to the NCLDV although the evolution of these large viruses is even more dynamic than that of cellular life forms.

3. DYNAMIC EVOLUTION OF THE NCLDV, SMALL CORE GENOME AND EXPANSIVE PANGENOME, CAPTURE OF UNIQUE SETS OF GENES, AND MULTIPLE ORIGINS OF VIRUS GIGANTISM

Numerous comparisons of the gene compositions of bacterial and archaeal genomes show that the evolutionary dynamics of these organisms is best described in terms of pangenomes, that is, the entirety of the genes that are represented in the genomes of all members of a given group, typically, a species (notwithstanding the problems with the definition of species in prokaryotes) (Medini et al., 2005; Puigbo et al., 2014; Tettelin et al., 2008). Most of the bacteria and archaea, for which sufficient genomic data are available, have small and stable core genomes but large, dynamic and, apparently, "open" pangenomes which incrementally grow with the addition of each new isolate and do not seem to show signs of saturation. A similar comparative genomic analysis for bacteriophages and archaeal viruses reveals even more fluid genomes, with tiny gene cores, large accessory genomes (genes found only in subgroups of the analyzed group), and

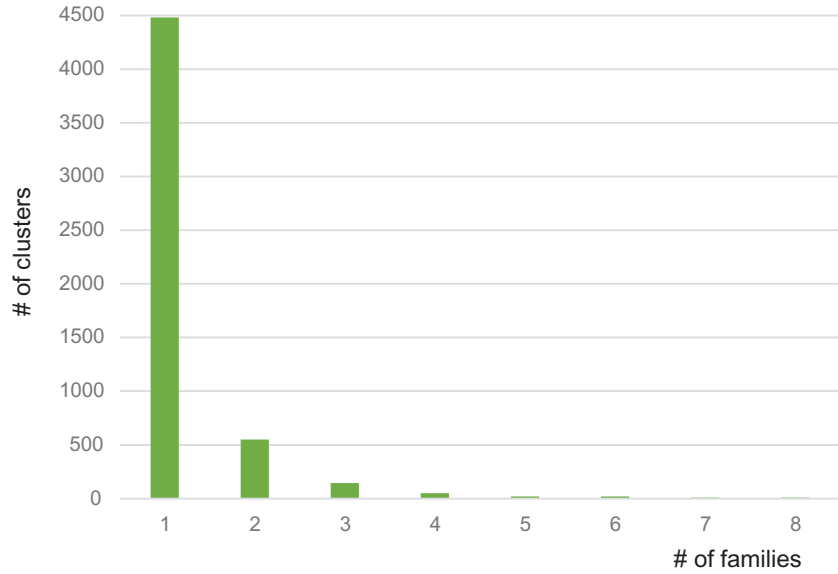

Fig. 2 Representation of NCLDV families in clusters of homologous genes. A comprehensive comparison of the protein sequences encoded by the NCLDV followed by clustering to identify groups of homologous genes yielded 5284 groups most of which are represented only in a single NCLDV family.

pronounced dominance of ORFans, that is, genes found in a single genome or in genomes of a cluster of closely related isolates (Kristensen et al., 2011, 2013). Analysis of the clusters of homologous NCLDV genes shows the same trends, apparently, closer to bacteriophages than to cellular life forms (Fig. 2). The core, even liberally defined by allowing gene losses in multiple lineages, remains at about 40 genes, whereas the pangenome has grown to more than 17,000 genes (the exact number depends on how many isolates are analyzed; in our analysis, we included only a subset of the numerous sequenced genomes of poxviruses and ASFV isolates). The NCLDV pangenome is heavily dominated by ORFans and genes that are conserved in small groups of viruses only (Fig. 2). Thus, the pangenome can be safely expected to grow further, for the foreseeable future.

A complementary look at the genome divergence among the NCLDV using Venn diagrams of shared and unique genes shows that, even among viruses that form tight, well-supported branches in the tree of the (near) universal genes, for example, the family *Mimiviridae*, or ASFV and its relatives, the overlap between the gene sets is surprisingly small (Fig. 3). These observations emphasize the remarkable evolutionary fluidity of the NCLDV

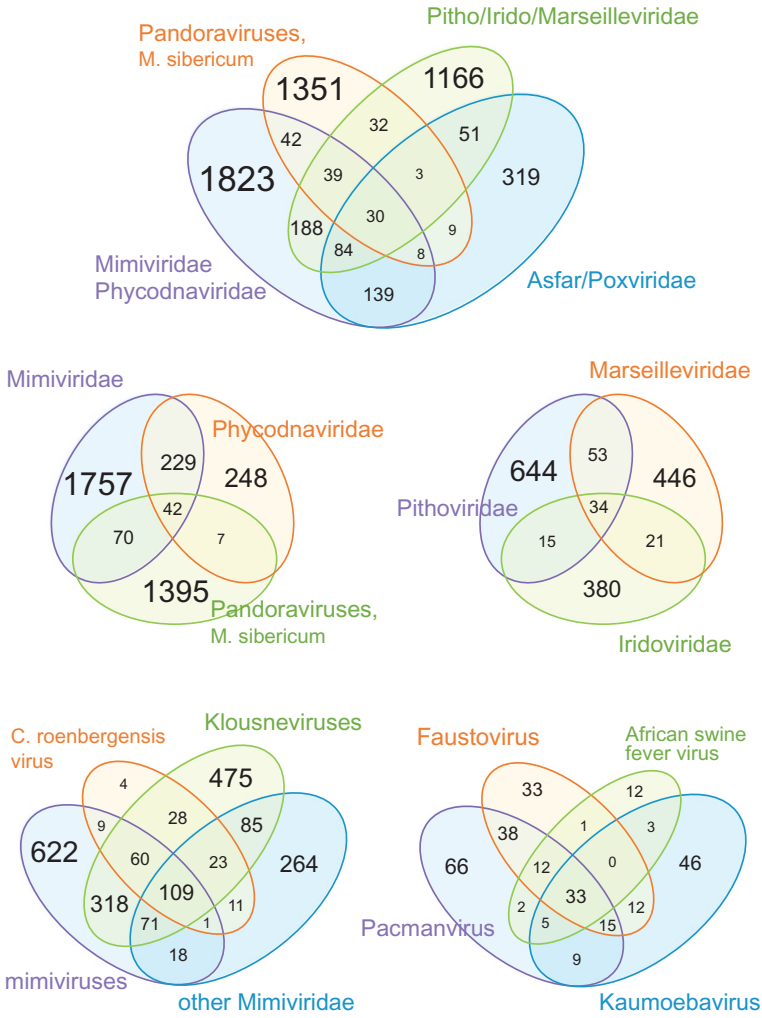

Fig. 3 Overlap between the gene repertoires of different NCLDV families. The Venn diagrams are based on the representation of the 5284 groups of homologous NCLDV genes in the respective virus families. From top to bottom, the diagrams show relationships within increasingly narrow groups of viruses.

genomes. Dynamic genome evolution necessarily involves both gene gain and gene loss, and evolutionary reconstructions indicate that, in prokaryotes, losses are, on average, more common than gains (Puigbo et al., 2014). In contrast, in the evolution of the NCLDV, the prevailing trend appears to be gene gain as captured by formal, maximum likelihood reconstructions of genome evolution along the tree of the nearly universal genes (Fig. 4),

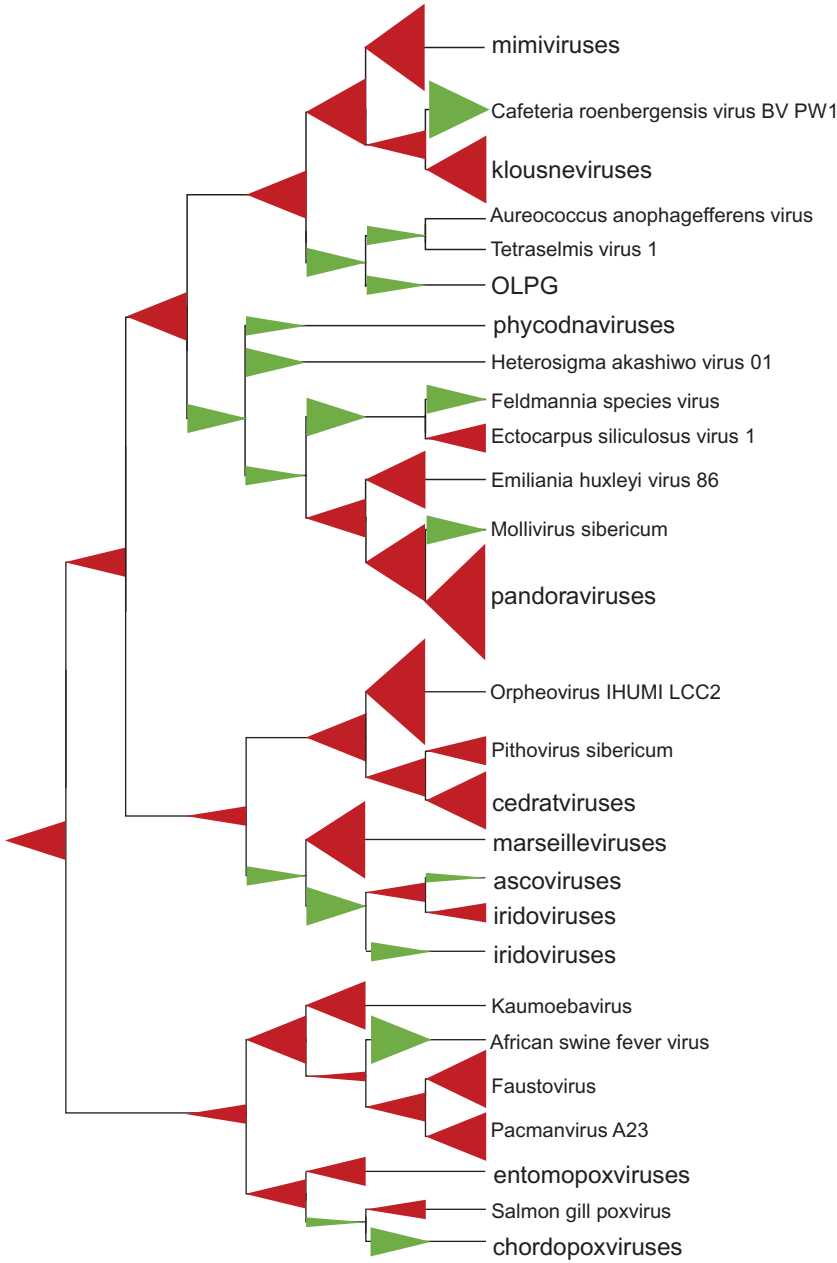

Fig. 4 Reconstruction of gene gain and loss events in the evolution of the NCLDV. The tree topology is from the phylogeny of five nearly universal genes (Fig. 1). The maximum likelihood reconstruction was performed using the COUNT software (Csuros, 2010), from the mapping of the 5284 clusters of homologous genes onto the tree leaves (extant viruses). Red triangles denote gene gains, and green triangles denote gene losses. The size of a triangle is roughly proportional to the maximum likelihood estimate of the number of gains or losses. OLPG stands for "Organic Lake/*Phaeocystis globosa*" virus group.

and also is intuitively clear from the minimal overlap that is observed between the gene sets of viruses within different branches of the NCLDV (Fig. 3). The large fractions of unique and narrowly conserved genes in the genomes of the NCLDV indicate that these genes, mostly, represent lineage-specific gains rather than resulting from losses of ancestral genes. An evolutionary scenario dominated by gene loss would imply unrealistically inflated ancestral "genomes of Eden" containing most of the unique viral genes. Nevertheless, in the evolution of several groups of NCLDV, gene loss was inferred to prevail over gene gain as is the case of most Phycodnaviruses, the OLPG group in the extended family *Mimiviridae*, iridoviruses, and ASFV (Fig. 4). This conclusion largely agrees with the "genome accordion" model which interprets the dynamic evolution of the NCLDV as alternate phases of gene loss and gain (Filee, 2013) although we find that the trend of gene gain is stronger than that of gene loss. Notably, according to our phylogenomic reconstruction (Figs. 1 and 4), giant viruses emerged on at least three independent occasions, during the evolution of the mimiviruses, pandoraviruses, and pithoviruses. Discovery of new groups of giant viruses, for example, among the protist-infecting relatives of asfarviruses or even among poxviruses, appears likely.

Because, by definition, the core set of the NCLDV genes includes those that map to the last common ancestor, among these genes, gain is impossible, whereas gene losses result in the observed patchy distribution of the core genes (Fig. 1). Strikingly, only three genes, those encoding family B DNA polymerase (DNAP), primase-helicase, and a (predicted) transcription factor that is known to be required for late transcription in poxviruses (VLTF3), are conserved in all available NCLDV genomes. All the rest of the core genes, even those encoding the central functions involved in virus morphogenesis and reproduction, such as the major capsid protein (MCP) and RNA polymerase (RNAP) subunits, are lost in certain NCLDV lineages. The loss of some of the core genes is readily interpretable in functional terms. An obvious example is the loss of the MCP gene in pandoraviruses which do not form typical capsids (Philippe et al., 2013). Notably, however, other viruses with derived virion structures, namely, pithoviruses, Mollivirus, poxviruses, and ascoviruses, have retained the MCP genes. In the case of the poxviruses, it has been shown that the MCP ortholog (D13 protein in vaccinia virus), although not incorporated into virions, is involved in the formation of icosahedral intermediates in the virion morphogenesis (Bahar et al., 2011; Szajner et al., 2005). A similar role can be hypothesized for the MCP orthologs encoded by pithoviruses, Mollivirus, and ascoviruses. The interpretation of the loss of the RNAP subunits in phycodnaviruses is also readily

explainable by the apparent existence of a nuclear phase in the reproduction of these viruses whereby transcription of the virus genes is, at least, partially, relegated to the host RNAP (Van Etten and Meints, 1999). Other losses are more difficult to explain. For example, pithoviruses lack the gene encoding the ATPase that is involved in DNA packaging into the virions in other NCLDV (Cassetti et al., 1998; Chelikani et al., 2014; Koonin et al., 1993). A clear explanation seems to be that DNA is incorporated into the amphora-shaped virions of the pithoviruses via a distinct mechanism. However, pandoraviruses that have similarly shaped virions possess multiple paralogs of the ATPase gene (Yutin et al., 2014), rendering this explanation dubious. The erosion of the ancestral core gene set occurred non-uniformly across the NCLDV branches (Fig. 1). In particular, and somewhat paradoxically, pandoraviruses, the genome size record holders in the entire virosphere, have lost more core genes than any other group of the NCLDV, perhaps, in conjunction with the shift in the virion architecture (Yutin and Koonin, 2013). It should be noticed that the core gene set of the NCLDV is, in part, a reflection of the current state of the art in protein sequence and structure comparison: poorly conserved genes can be missed. An example is the minor capsid protein (penton) that appears to be an essential constructive elements of large icosahedral virions but only recently has been identified in many but not all NCLDV with the help of the most sensitive available analytic tools (Krupovic and Koonin, 2015). It cannot be ruled out that several other conserved genes remain undetected in many viruses.

The apparent conservation of some of the core NCLDV genes seems to obscure more complex evolutionary scenarios, such as independent capture of homologous genes from different hosts or displacement of ancestral genes by homologs of different provenance (Yutin and Koonin, 2012). For example, strong indications have been obtained that even a universal NCLDV gene that for the primase-helicase, was replaced with a phage homolog in phycodnaviruses. Several other core genes, such as those for nucleotide metabolism enzymes (ribonucleotide reductase, thymidine, and thymidylate kinases, dUTPase), are clearly polyphyletic, that is, were captured from different hosts and on different occasions.

The origin of the lineage-specific genes in NCLDV remains, to a large extent, an open problem. Although, for some of these genes, acquisition from the hosts or from bacteria, possibly, endosymbionts of protist hosts, is apparent (Filee et al., 2007, 2008; Moreira and Brochier-Armanet, 2008), the majority are of unknown provenance. The only group of the NCLDV, for which the majority of the lineage-specific genes are readily traceable to

functionally characterized host genes, are chordopoxviruses. In these viruses, most of the lineage-specific genes encode homologs of host proteins involved in immune and programmed cell death functions (Moss, 2001; Moss et al., 2000; Senkevich et al., 1997). These viral proteins have been shown or predicted to function, mostly, as dominant-negative inhibitors of the homologous host defense proteins (Howard et al., 1998; Kotwal, 2000; Moss, 2001; Moss et al., 2000; Palumbo et al., 1994; Senkevich et al., 1997). However, these antidefense proteins have been identified primarily in mammalian poxviruses. Apart from the poxvirus core, crocodile poxvirus and, particularly, fish poxviruses encompass primarily genes of unidentifiable origins (Afonso et al., 2006; Gjessing et al., 2015). From a biological perspective, it appears likely that most if not all lineage-specific genes of the NCLDV are involved in virus-host interactions and, more specifically, counteract host defenses via diverse mechanisms. In the next section, we discuss in some detail a remarkable group of such genes in giant viruses, those encoding translation systems components, and trace their evolutionary history. Nevertheless, the origins of most of the lineage-specific virus genes remain enigmatic.

For organisms that experience significant gene loss and gain, trees constructed from matrices of gene presence-absence can complement phylogenies of conserved genes with a broader view of genome evolution (Snel et al., 2005; Wolf et al., 2002). The gene content tree of the NCLDV (Fig. 5) almost precisely reproduces the topology of the tree for the five conserved genes (Fig. 1). Thus, despite the apparent fast divergence of the virus gene content in the NCLDV evolution, which is demonstrated by the small numbers of shared genes (Fig. 3), gains, and losses of comparatively well-conserved genes that contribute to the gene content tree seem to occur, at least roughly, under a "genomic clock" model, resulting in the congruence of the two independent trees. Such congruence supports the relevance of the tree built for a small number of conserved genes as a reflection of the virus genome evolution.

4. TRANSLATION SYSTEM COMPONENTS IN GIANT VIRUSES: PIECEMEAL CAPTURE OF EUKARYOTIC GENES

Apart from tRNAs encoded by some bacteriophages (Pope et al., 2014; Yoshikawa et al., 2018), viruses, as a rule, lack genes for translation system components. The absence of their own translation system and reliance on that of the host is one of the key features that distinguish viruses from

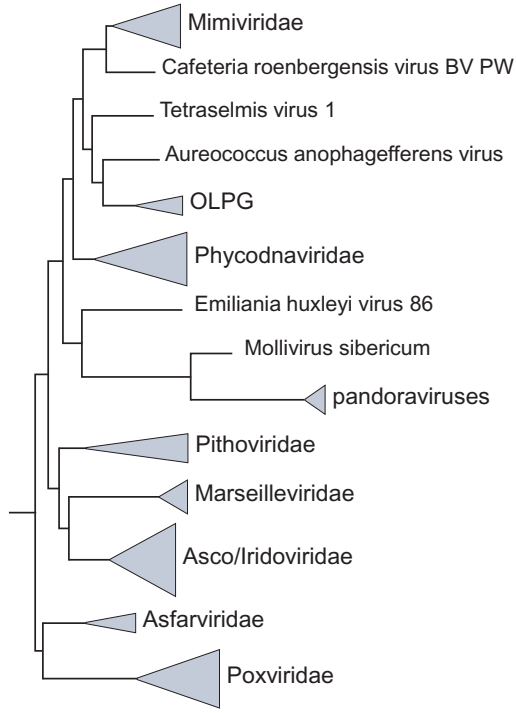

Fig. 5 Gene content tree of the NCLDV. The tree was built using the neighbor-joining method (Saitou and Nei, 1987) from the complete matrix of the presence-absence of the 5284 groups of homologous NCLDV genes in each virus genome.

cellular life forms (Raoult and Forterre, 2008). The giant NCLDV are a major exception to this rule (Table 2). The discovery of genes encoding aaRS and translation factors in the mimivirus came as a big surprise which triggered the fourth domain hypothesis (Colson et al., 2011, 2012; Desnues et al., 2012; Legendre et al., 2012; Nasir et al., 2012, 2017; Raoult et al., 2004). Remarkably, the representation of the translation system components in the mimivirus and its closest relatives is actually quite modest compared to the recently discovered Klosneuvirus and Tupanvirus which encode nearly full complements of the proteins and tRNAs involved in translation, except for the ribosomal RNA and proteins (Table 2) (Abrahao et al., 2018; Schulz et al., 2017). Notably, viruses in the family *Mimiviridae* show dramatic variance in their repertoires of translation-related genes. For example, Tupanvirus, a close relative of the mimiviruses, encodes the complete set of 20 aaRS whereas the mimiviruses have 7 at most. Similarly, Klosneuvirus encodes 19 aaRS (Schulz et al., 2017), whereas the related Bodo sultans virus

Table 2 Translation-Related Genes in NCLDV

	Mimiviruses (4)	CroV (1)	Klosneuviruses (5)	OLPG (6)	Phycodnaviruses (12)	Pandoraviruses, M. Sibericum (7)	Pithoviridae (6)	Marseilleviridae (7)	Asco/Iridoviruses (15)	Extended Asfarviridae (4)	Poxviruses (12)
AlaS	1		1								
ArgS	4		3				1				
AsnS	4		4	1			1				
AspS	1		1				1				
CysS	4		1								
GlnS	1		3								
GlyS	1		2				1				
HisS	1		3				1				
IleS	4	1	5				1				
LeuS	1		3								
LysS	1		3								
MetS	4		4								
ProS	1		3								
PheS	1		2				1				
ThrS	1		3								
SerS	1		2				1				
TrpS	2		3			1					
TyrS	4		2			4	1				

ValS	1		2								
SUI1	4	1	4				1	3		1	
eIF1A	1	1	4	1							
eIF2a	1	1	5						1		
eIF2b	1	1	5					5			
eIF2g	1	1	5								
eIF4a	4	1	4	1							
eIF4e	4	1	5	5		7					1
eIF4g			4								
eIF5A	1		2	1			1				
eIF5B		1	2								
EF1a	4		5	1			3	5			
aEF2	1										
eEF3				1	1						
eRF1	4		4				1	4			

The number of the available complete genomes is indicated in parentheses for each group of viruses. The numbers in the table indicate how many genomes of the given group contain the respective translation-related gene. *OLPG* stands for "Organic Lake/*Phaeocystis globosa* virus group.

has only two (Deeg et al., 2018). These differences imply extensive, repeated gain, and/or loss of translation-related genes in giant viruses.

At face value, the striking abundance of genes for translation system components in Klosneuvirus and Tupanvirus could be construed as being best compatible with the origin of these viruses by reductive evolution from a cellular ancestor. However, phylogenetic analysis of individual translation-related genes tells a different story (Williams et al., 2011; Yutin et al., 2014) as illustrated by four phylogenies of the genes that are most common among the NCLDV in this functional category (Fig. 6). There are three dominant trends in these phylogenies: (i) most of the translation-related genes fall deep within the eukaryotic tree, suggestive of a relatively late acquisition from eukaryotic hosts, (ii) in all cases, the NCLDV genes are polyphyletic which indicates that these genes were acquired repeatedly and independently by different groups of the NCLDV, and (iii) in the phylogenetic trees of the translation-related genes, viruses are mixed in different combinations, suggestive of widespread intervirus gene exchange. Some of the genes appear to have been captured relatively early in the evolutionary history of the NCLDV. For example, eukaryotic translation initiation factor eIF4e apparently was acquired independently by the common ancestors of the mimiviruses and pandora-molliviruses (Fig. 6A). The mimiviruses are monophyletic also in the phylogenies of the tyrosyl-tRNA synthetase (TyrS) (Fig. 6B) and release factor eRF1 (Fig. 6C). Otherwise, however, the tree topologies are complicated, implying convoluted scenarios of gene gain, loss, and exchange. For example, the TyrS tree includes five distinct branches of NCLDV which are affiliated with different groups of eukaryotes (Fig. 6B). One of these branches includes Tupanvirus and two Klosneuviruses, whereas mimiviruses, the relatives of Tupanvirus, and Catovirus, a member of the Klosenuvirus group, belong in a different subtree (Fig. 6B). Conceivably, TyrS was already present in the common ancestor of the Mimiviridae, including Tupanvirus, and Klosneuviruses but was independently displaced by a homolog from a distinct eukaryotic host in the Mimiviruses and Catovirus. The TyrS genes of Pandoraviruses and Orpheovirus appear to have been acquired independently from yet other eukaryotic hosts. In the case of Pandoraviruses, a relatively recent capture of this gene from Acanthamoeba appears likely. The tree of the release factor eRF1 also includes two NCLDV clades (Fig. 6C). One of these consists of Marseilleviruses, with a clear eukaryotic affinity, whereas the second one is an odd assortment of viruses, combining Mimiviruses, Pacmanvirus (Asfarvirus relative), and Orpheovirus (giant virus in the family Pithoviridae), and unexpectedly

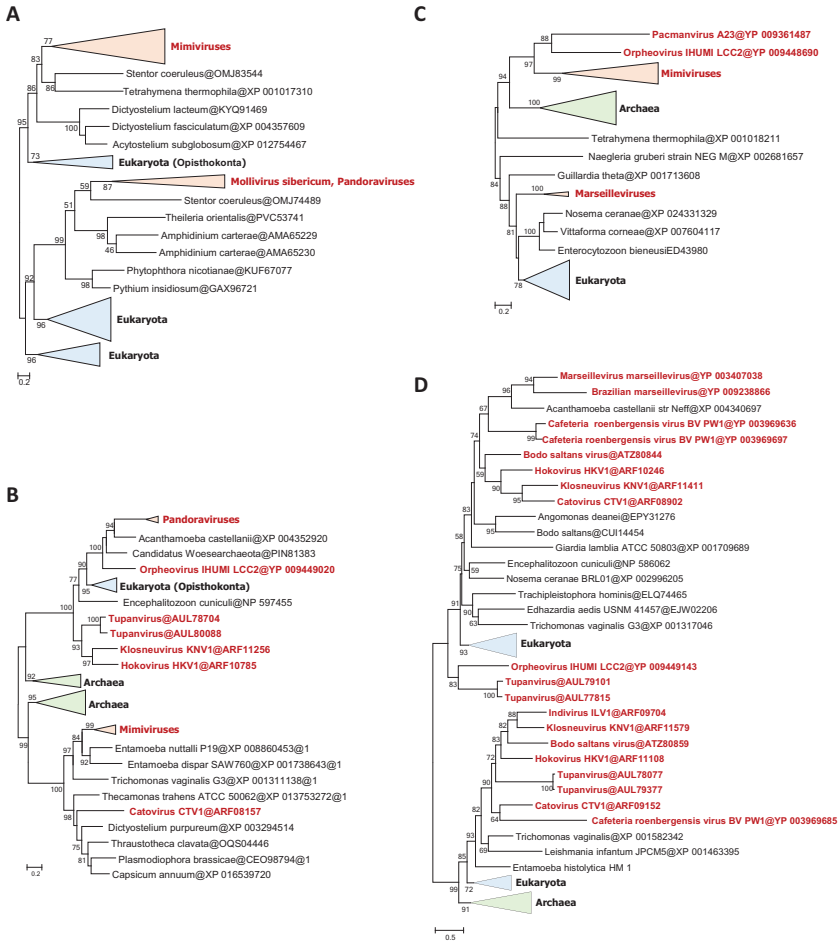

Fig. 6 Phylogenies of selected translation system components encoded by the NCLDV. (A) translation initiation factor eIF4e; (B) tyrosyl-tRNA synthetase, TyrS; (C) peptide chain release factor 1, eRF1; and (D) translation initiation factor eIF2b. The trees were constructed using the FastTree software (Price et al., 2010) with default parameters. The numbers at the internal branches indicate local likelihood-based support (percentage points).

joining the archaeal branch of the tree. Conceivably, the evolutionary history of this gene involved acquisition of the release factor gene from archaea, perhaps a protist endosymbiont, by one of the giant viruses, possibly the ancestor of the Mimiviruses, followed by multiple intervirus gene transfers. The tree of the translation initiation factor eIF2b appears to reflect an even more complicated history (Fig. 6D). There are four distinct

NCLDV branches in this tree, all of which are affined with different groups of eukaryotes. Two of the branches include identical sets of viruses, namely, members of the *Mimiviridae* except for the mimiviruses sensu stricto. Thus, the ancestors of the *Mimiviridae* appear to have acquired independently two homologous genes from different eukaryotic hosts which was apparently followed by the loss of these genes in the mimiviruses.

In summary, phylogenetic analysis of the translation-related genes of the NCLDV reveals a highly complicated history that is hardly compatible with their origin from cellular ancestors via reductive evolution but rather implies piecemeal capture of these genes from different eukaryotes (and in a few cases, bacteria, or archaea), multiple gene losses and displacements, and extensive intervirus exchange of genes. Convergent, independent acquisition of homologous translation-related genes by NCLDV from widely different branches occurred many times in evolution. This trend suggests a strong selective pressure for the capture and retention of these genes which culminated in the accretion of a near complete translation system (save for the ribosome) in Tupanvirus and Klosneuviruses. The causes of such an evolutionary pressure are unknown. A plausible explanation seems to be that the protist hosts react to the giant virus infection by shutting down their translation system, so that viruses strongly benefit from the capability to, at least partially, replenish the translation system components and restore translation to produce virus proteins. Unfortunately, virus-host interactions in protists barely have been studied, and the prospects of extensive experimentation in this field appear rather dim cue to the lack of well-established model systems.

5. THE BIGGER PICTURE: DNA VIROSPHERE AS A NETWORK AND EVOLUTIONARY RELATIONSHIPS BETWEEN NCLDV AND SMALLER dsDNA VIRUSES

Even the simplest of the NCLDV, such as iridoviruses with genomes of about 100 kb, are large, complex viruses, so the question of their potential origin from smaller, simpler viruses is pertinent. Comparison of the morphogenetic modules of diverse viruses gives some clues. Homologous genes for the major and minor jelly-roll capsid proteins, DNA packaging ATPase and protease involved in capsid maturation are shared by a broad range of eukaryotic dsDNA viruses (Iranzo et al., 2016; Krupovic and Koonin, 2015). The viruses sharing these four major morphogenetic genes include adenoviruses, virus-like transposons known as polintons (after polymerase and integrase) and polinton-like viruses (PLV), virophages, and NCLDV

(Iranzo et al., 2016; Krupovic and Koonin, 2015; Yutin et al., 2015). The polintons encode the entire morphogenetic module and hence have been dubbed "polintoviruses" although virions remain to be identified for any of them (Krupovic and Koonin, 2015). However, some virions of polinton-like viruses (PLV), an expansive group of viruses identified by metagenomic sequence analysis and distantly related to polintons, have been detected (Yutin et al., 2015). Structural modeling shows that the predicted MCPs encoded by the polintons are closely similar to the MCPs of both NCLDV and tailless bacteriophages of the family *Tectiviridae* and their numerous relatives (Krupovic et al., 2014; Yutin et al., 2018). Polintons, thus, seem to be a likely intermediate between tectivirus-like prokaryotic ancestors and "hotbed" of eukaryotic dsDNA virus evolution, giving rise to a variety of eukaryotic dsDNA viruses of widely different sizes, including the virophages and the NCLDV (Fig. 7) (Krupovic and Koonin, 2015).

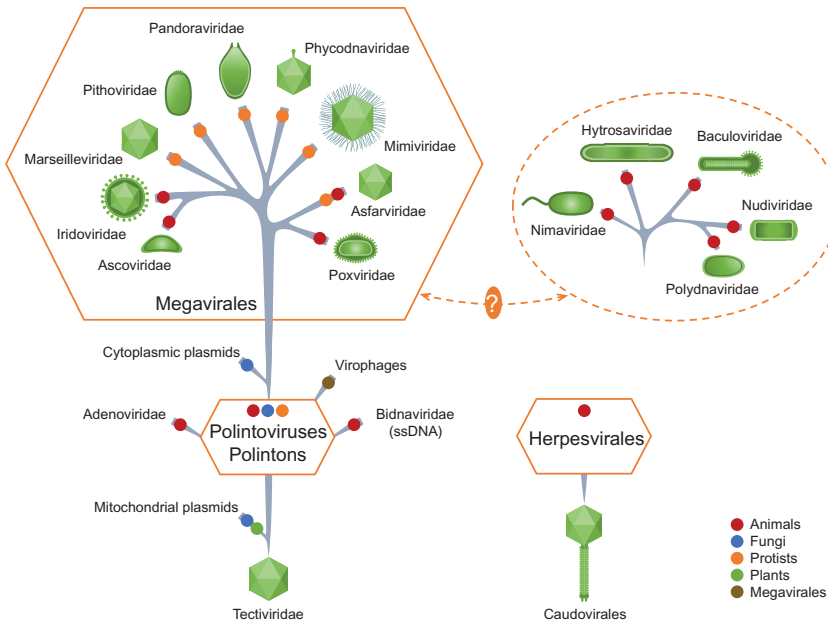

Fig. 7 The big picture of the NCLDV origins and the evolutionary links with other viruses. Several large dsDNA viruses infecting arthropods are tentatively linked to the NCLDV through several core genes and might represent highly derived forms of the NCLDV. The origin of herpesviruses that are not related to the NCLDV from tailed bacteriophages is included to present a complete picture of the evolution of the dsDNA virosphere. *The figure is adopted from Koonin, E.V., Dolja, V.V., Krupovic, M., 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. Virology, 479–480, 2–25, under Creative Common Attribution License (CC BY).*

In addition to the morphogenetic module, tectiviruses, polintons, adenoviruses, and some virophages encode a DNA polymerase (DNAP) that, however, is belongs to the protein-primed subfamily of the family B DNAPs and is only distantly related to the DNAPs of the NCLDV. Some polintons also encode a primase-helicase distantly related to those that are universally present in the NCLDV. Given that these replication proteins do not appear to be monophyletic with the corresponding proteins of the NCLDV, the evolution of the latter apparently involved recruitment of the replication and transcription machineries from the eukaryotic hosts, combining these with the polinton-derived morphogenetic module and providing for the dramatic genome expansion (Krupovic and Koonin, 2015). It is notable that the giant mimiviruses and their parasites, the virophages, ultimately share the same ancestry (Fig. 7). Finally, it should be noted that several apparently related, even if diverse, families of large dsDNA viruses infecting arthropods, namely, *Baculoviridae, Nudiviridae*, *Polydnaviridae*, *Hytrosoviridae*, and *Nimoviridae*, share subsets of core genes with the NCLDV and might represent highly derived NCLDV forms (Koonin et al., 2015).

The evolutionary scenario illustrated in Fig. 7 is compatible with the results of the analysis of a bipartite network of dsDNA viruses in which viral genomes are connected through shared genes (Iranzo et al., 2016). This network shows pronounced modularity, and the two supermodules that encompass the vast majority of viruses consist of: (1) prokaryotic virus families *Tectiviridae*, *Corticoviridae*, *Sphaerolipoviridae*, and *Turriviridae* together with the eukaryotic adenoviruses, polintons, PLV, virophages, mitochondrial dsDNA plasmids, and the NCLDV and (2) the prokaryotic order *Caudovirales* (tailed bacteriophages) and the order *Herpesvirales* that includes only animal viruses. The principal distinguishing features of these two supermodules of dsDNA viruses are their unrelated morphogenetic gene blocks. All the uncertainties inherent in deep evolutionary reconstructions notwithstanding, these analyses show that the NCLDV, including the giant members of this group, are descendants of much smaller and simpler, typical viruses.

6. THE TAXONOMIC PERSPECTIVE: ORDER, CLASS, PHYLUM?

The evolutionary coherence and common ancestry of the NCLDV are beyond reasonable doubt. However, the most distant NCLDV, for example, Pandoraviruses vs Poxviruses, are dramatically different from each

other, with the genome sizes varying by more than an order of magnitude and only about 20 core genes shared. It appears highly desirable that evolutionarily coherent groups of viruses received a formal taxonomic status but so far, the proposal to formalize the relationships among the NCLDV by creating the order "Megavirales" (Colson et al., 2013) has not been successful with the ICTV. And, perhaps, for a good reason. Now, with the dramatically increased diversity of sequenced virus genomes, we are gaining a deeper understanding of the evolutionary relationships among all types of viruses, and it is becoming clear that the full hierarchical taxonomic system modeled after that employed for cellular life forms must be adopted for viruses. The initial steps in this direction are being made, with the first virus phylum, Negarnaviricota (for minus-strand RNA viruses), with two subphyla, already approved by the ICTV. Gradual incorporation of the entire virosphere into the hierarchical taxonomic structure appears to be imminent.

In light of these developments, what should be the taxonomic status of the NCLDV? The complex structure of the phylogenetic tree of the core genes (Fig. 1) that is largely congruent with the gene content tree (Fig. 4) implies a commensurate taxonomic rank that should be higher than an order. A distinct possibility seems to be that the three major branches of the NLCDV revealed by phylogenetic analysis, namely, (1) mimi-phycodnaviruses (including pandoraviruses), (2) pitho-irido-marseilleviruses, and (3) expanded asfarviruses-poxviruses become orders within the NCLDV class (that will, obviously, receive a new name). Potentially, one could think even about the inclusion of the NCLDV into a phylum that would encompass all dsDNA viruses with the double jelly-roll MCP. Whatever the ultimate solution, the NCLDV can be expected to become a high-rank taxon with a rich internal structure.

7. CONCLUDING REMARKS

The NCLDV comprise a highly diverse group of large viruses that include the size record holders of the virosphere and show an enormous diversity of shapes, genetic layouts, and host ranges. All this variation notwithstanding, the identification of about 40 core genes that are shared by the majority of the NCLDV and, in evolutionary reconstructions, map to the last common ancestor of the NCLDV leaves no reasonable doubt of the monophyly of this expansive group of viruses. There is also a biological counterpart to that genetic coherence of the NCLDV: most of these viruses reproduce in the cytoplasm, inside virus factories (Katsafanas and Moss, 2007;

Suzan-Monti et al., 2007), although some have secondarily moved to the nucleus. Hence the notable autonomy of the NCLDV that encode most if not all components of their replication and expression systems as well as some additional pathways, such as the one for disulfide bond formation in the cytosol (Senkevich et al., 2002). The giant viruses take this functional autonomy to an unprecedented level, with some encoding a nearly complete translation system, except for the ribosome.

Phylogenomic analysis yields well-supported scenarios for the evolution of the NCLDV that is traceable through the phylogeny of core genes complemented by the reconstructions of gene gain and loss. These processes were extensive during the evolution of the NCLDV, so that the gene repertoires of viruses of different families show little overlap. In most branches of the NCLDV, gene gain prevailed over gene loss, and evolutionary reconstructions indicate that giant viruses independently evolved from smaller ones on at least three, and probably, more occasions. In particular, these viruses captured the genes for translation system components in a piecemeal manner, apparently, from different eukaryotic hosts. Moreover, on numerous occasions, homologous translation-related genes were acquired convergently by different emerging giant viruses.

Phylogenomic reconstructions yield a fairly detailed picture of the evolution of the NCLDV, revealing the key evolutionary processes and transitions. Especially notable are the changes in virion morphology, from the icosahedral capsids that, evidently, are ancestral for NCLDV and beyond, to unique particle shapes. Such a transition occurred at least five times during the NCLDV evolution, yielding the spherical virions of the Mollivirus, the amphora-shaped giant virions of pandoraviruses and pithoviruses, the brick-shaped poxvirus particles, and the ovoid virions of ascoviruses. It does not seem unlikely that more odd virion shapes are discovered or, at least, that additional groups of giant viruses with amphora-shaped virions pop up. A particularly intriguing question seems to be whether there exists some limit on genome size for incorporation into icosahedral virions and whether klosneuviruses and Orpheovirus, with the genomes in the range of 1.4–1.5 Mb, already push that limit, whereas pandoraviruses cross the line. Perhaps, true super-giants that would dwarf pandoraviruses appear among viruses with amphora-shaped virions? Whether or not the evolution of the unique virion shapes represents specific adaptations for virus penetration into specific host cells or for other aspects of virus-host interaction is at present completely unknown.

Of no lesser interest are the evolutionary transitions in the host range of the NCLDV. The majority and the greatest diversity of the NCLDV are associated with various protist hosts, and it appears obvious that the NCLDV originated in a unicellular eukaryotic host, most likely, at an early stage in the evolution of eukaryotes. The leap to animal hosts occurred on at least three independent occasions, namely, in the ancestor of irido-ascoviruses, poxviruses, and asfarviruses. The first two events are likely to have been ancient because the respective groups of viruses infect both vertebrates and arthropods. In contrast, asfarvirus appear to be a case of recent host switch because in this case, a single virus species (albeit highly successful) is known to infect a single animal host. It is a question of considerable interest whether any common adaptations can be identified in the viruses that have switched to animal hosts but so far, no such adaptations have been detected and, in particular, no genes uniquely shared by animal-infecting NCLDV have been detected. Interestingly, although NCLDV are common in algae but, apparently, absent in land plants, integrated genes originated from a distinct NCLDV of unclear provenance have been identified integrated in moss genomes (Maumus et al., 2014), suggesting that NCLDV were lost from the plant lineage at a relatively late stage of evolution.

The phylogenomic reconstructions also provide the basis for new taxonomy. It can be confidently expected that the NCLDV become a high-rank taxon, such as a class, with several constituent orders that will contain already recognized and new families. All the progress in phylogenomics notwithstanding, major enigmas remain with regard to the forces that drive the evolution of the NCLDV and, especially, the biology of these complex viruses. Apart from the poxviruses and the iridoviruses that have the smallest genomes among the NCLDV, a substantial majority of the genes in most other members of the group, particularly, the giant viruses, remain ORFans, with unknown provenance and functions in virus reproduction. It seems to be a safe bet that these genes are involved in virus-host interactions and, in particular, counteract host defenses, but the specifics remain obscure. Moreover, even when the biochemical activity of NCLDV genes is quite clear, as in the case of the translation system components, the biological underpinning of the apparent drive for the capture of these genes in NCLDV remains unclear, even if plausible hypotheses can be proposed.

The causes of the growth of some (but not other) virus genomes that repeatedly led to virus gigantism are not well understood. Generally, the evolutionary factors that promote the genomic expansion in multiple

lineages of the NCLDV are likely to be related to distinct aspects of virus-host interaction in protists. One plausible possibility is that, to be able to enter the cells of protists, such as amoeba, via phagocytosis, viruses should exceed a minimal particle size, about 1 μm (Rodrigues et al., 2016), and this limitation could drive the evolution of giant virions that could accommodate large genomes. In accord with this scenario, marseillevirus virions that are far below the phagocytosis size threshold have been shown to form "giant" multi-particle aggregates that are phagocytized by amoeba (Arantes et al., 2016).

Arguably, to attain such understanding, complementing comparative genomics with actual biology of the viruses in their natural hosts is a must. Given the "exotic" nature of most of these hosts, this is a major challenge.

From a general perspective, we believe that the study of the NCLDV evolution does not support radical ideas on giant viruses breaking through the virus-cell divide which remains the fundamental demarcation line between autonomous and parasitic biological entities (Koonin, 2010; Koonin and Wolf, 2012). However, this does not make these highly complex viruses any less interesting as an evolutionary oddity and the object of fascinating research.

ACKNOWLEDGMENTS

The authors thank Yuri Wolf for expert help with computational analyses. The authors' research is supported by intramural funds of the National Institutes of Health of the United States.

REFERENCES

Abeles, S.R., Pride, D.T., 2014. Molecular bases and role of viruses in the human microbiome. J. Mol. Biol. 426, 3892–3906.

Abergel, C., Legendre, M., Claverie, J.M., 2015. The rapidly expanding universe of giant viruses: mimivirus, pandoravirus, pithovirus and mollivirus. FEMS Microbiol. Rev. 39, 779–796.

Abrahao, J., Silva, L., Silva, L.S., Khalil, J.Y.B., Rodrigues, R., et al., 2018. Tailed giant tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat. Commun. 9, 749.

Afonso, C.L., Tulman, E.R., Delhon, G., Lu, Z., Viljoen, G.J., et al., 2006. Genome of crocodilepox virus. J. Virol. 80, 4978–4991.

Aherfi, S., Colson, P., La Scola, B., Raoult, D., 2016. Giant viruses of amoebas: an update. Front. Microbiol. 7, 349.

Allen, M.J., Forster, T., Schroeder, D.C., Hall, M., Roy, D., et al., 2006. Locus-specific gene expression pattern suggests a unique propagation strategy for a giant algal virus. J. Virol. 80, 7699–7705.

Alonso, C., Borca, M., Dixon, L., Revilla, Y., Rodriguez, F., et al., 2018. ICTV virus taxonomy profile: asfarviridae. J. Gen. Virol. 99, 613–614.

Andreani, J., Aherfi, S., Bou Khalil, J.Y., Di Pinto, F., Bitam, I., et al., 2016. Cedratvirus, a double-cork structured giant virus, is a distant relative of pithoviruses. Viruses 8, pii: E300.
Andreani, J., Khalil, J.Y.B., Baptiste, E., Hasni, I., Michelle, C., et al., 2017a. Orpheovirus IHUMI-LCC2: a new virus among the giant viruses. Front. Microbiol. 8, 2643.
Andreani, J., Khalil, J.Y.B., Sevvana, M., Benamar, S., Di Pinto, F., et al., 2017b. Pacmanvirus, a new giant icosahedral virus at the crossroads between asfarviridae and faustoviruses. J. Virol. 91, pii: e00212-17.
Andreani, J., Verneau, J., Raoult, D., Levasseur, A., La Scola, B., 2018. Deciphering viral presences: two novel partial giant viruses detected in marine metagenome and in a mine drainage metagenome. Virol. J. 15, 66.
Arantes, T.S., Rodrigues, R.A., Dos Santos Silva, L.K., Oliveira, G.P., De Souza, H.L., et al., 2016. The large marseillevirus explores different entry pathways by forming giant infectious vesicles. J. Virol. 90, 5246–5255.
Asgari, S., Bideshi, D.K., Bigot, Y., Federici, B.A., Cheng, X.W., et al., 2017. ICTV virus taxonomy profile: ascoviridae. J. Gen. Virol. 98, 4–5.
Bahar, M.W., Graham, S.C., Stuart, D.I., Grimes, J.M., 2011. Insights into the evolution of a complex virus from the crystal structure of vaccinia virus D13. Structure 19, 1011–1020.
Bajrai, L.H., Benamar, S., Azhar, E.I., Robert, C., Levasseur, A., et al., 2016. Kaumoebavirus, a new virus that clusters with faustoviruses and asfarviridae. Viruses 8, 278.
Baltimore, D., 1971a. Expression of animal virus genomes. Bacteriol. Rev. 35, 235–241.
Baltimore, D., 1971b. Viral genetic systems. Trans. N. Y. Acad. Sci. 33, 327–332.
Barr, J.J., 2017. A bacteriophages journey through the human body. Immunol. Rev. 279, 106–122.
Beijerinck, M.W., 1898. Über ein contagium vivum fluidum als ursache der fleckenkrankheit der tabaksblätter. In: Verhandelingen der Koninklijke akademie van Wetenschappen te, vol. 65. Muller, Amsterdam, pp. 1–22.
Bergh, O., Borsheim, K.Y., Bratbak, G., Heldal, M., 1989. High abundance of viruses found in aquatic environments. Nature 340, 467–468.
Birtles, R.J., Rowbotham, T.J., Storey, C., Marrie, T.J., Raoult, D., 1997. Chlamydia-like obligate parasite of free-living amoebae. Lancet 349, 925–926.
Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., et al., 2009. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc. Natl. Acad. Sci. U. S. A. 106, 21848–21853.
Cassetti, M.C., Merchlinsky, M., Wolffe, E.J., Weisberg, A.S., Moss, B., 1998. DNA packaging mutant: repression of the vaccinia virus A32 gene results in noninfectious, DNA-deficient, spherical, enveloped particles. J. Virol. 72, 5769–5780.
Chelikani, V., Ranjan, T., Kondabagil, K., 2014. Revisiting the genome packaging in viruses with lessons from the "giants". Virology 466–467, 15–26.
Chinchar, V.G., Waltzek, T.B., Subramaniam, K., 2017. Ranaviruses and other members of the family iridoviridae: their place in the virosphere. Virology 511, 259–271.
Claverie, J.M., Abergel, C., 2009. Mimivirus and its virophage. Annu. Rev. Genet. 43, 49–66.
Claverie, J.M., Ogata, H., Audic, S., Abergel, C., Suhre, K., et al., 2006. Mimivirus and the emerging concept of "giant" virus. Virus Res. 117, 133–144.
Claverie, J.M., Abergel, C., Ogata, H., 2009. Mimivirus. Curr. Top. Microbiol. Immunol. 328, 89–121.
Colson, P., Gimenez, G., Boyer, M., Fournous, G., Raoult, D., 2011. The giant cafeteria roenbergensis virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of life. PLoS One 6, e18935.

Colson, P., De Lamballerie, X., Fournous, G., Raoult, D., 2012. Reclassification of giant viruses composing a fourth domain of life in the new order megavirales. Intervirology 55, 321–332.
Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., et al., 2013. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch. Virol. 158, 2517–2521.
Colson, P., La Scola, B., Levasseur, A., Caetano-Anolles, G., Raoult, D., 2017a. Mimivirus: leading the way in the discovery of giant viruses of amoebae. Nat. Rev. Microbiol. 15, 243–254.
Colson, P., La Scola, B., Raoult, D., 2017b. Giant viruses of amoebae: a journey through innovative research and paradigm changes. Annu. Rev. Virol. 4, 61–85.
Csuros, M., 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26, 1910–1912.
Dagan, T., Martin, W., 2006. The tree of one percent. Genome Biol. 7, 118.
Deeg, C.M., Chow, C.T., Suttle, C.A., 2018. The kinetoplastid-infecting bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. eLife 7, pii: e33014.
Desnues, C., Boyer, M., Raoult, D., 2012. Sputnik, a virophage infecting the viral domain of life. Adv. Virus Res. 82, 63–89.
Filee, J., 2013. Route of NCLDV evolution: the genomic accordion. Curr. Opin. Virol. 3, 595–599.
Filee, J., Siguier, P., Chandler, M., 2007. I am what i eat and i eat what i am: acquisition of bacterial genes by giant viruses. Trends Genet. 23, 10–15.
Filee, J., Pouget, N., Chandler, M., 2008. Phylogenetic evidence for extensive lateral acquisition of cellular genes by nucleocytoplasmic large DNA viruses. BMC Evol. Biol. 8, 320.
Fischer, M.G., 2016. Giant viruses come of age. Curr. Opin. Microbiol. 31, 50–57.
Fischer, M.G., Hackl, T., 2016. Host genome integration and giant virus-induced reactivation 1 of the virophage mavirus. Nature 540, 288–291.
Fischer, M.G., Suttle, C.A., 2011. A virophage at the origin of large DNA transposons. Science 332, 231–234.
Forterre, P., Krupovic, M., Prangishvili, D., 2014. Cellular domains and viral lineages. Trends Microbiol. 22, 554–558.
Gjessing, M.C., Yutin, N., Tengs, T., Senkevich, T., Koonin, E., et al., 2015. Salmon gill poxvirus, the deepest representative of the chordopoxvirinae. J. Virol. 89, 9348–9367.
Howard, J., Justus, D.E., Totmenin, A.V., Shchelkunov, S., Kotwal, G.J., 1998. Molecular mimicry of the inflammation modulatory proteins (IMPs) of poxviruses: evasion of the inflammatory response to preserve viral habitat. J. Leukoc. Biol. 64, 68–71.
Iranzo, J., Krupovic, M., Koonin, E.V., 2016. The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. mBio 7, pii: e00978-16.
Iwanowski, D.I., 1892. Über die mosaikkrankheit der tabakspflanze. Bull. scient. Acad. imp. Sci. St. Petersb. 35, 7–70. Nouvelle Serie III.
Iyer, L.M., Aravind, L., Koonin, E.V., 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. 75, 11720–11734.
Iyer, L.M., Balaji, S., Koonin, E.V., Aravind, L., 2006. Evolutionary genomics of nucleocytoplasmic large DNA viruses. Virus Res. 117, 156–184.
Katsafanas, G.C., Moss, B., 2007. Colocalization of transcription and translation within cytoplasmic poxvirus factories coordinates viral expression and subjugates host functions. Cell Host Microbe 2, 221–228.
Khalil, J.Y.B., Andreani, J., La Scola, B., 2016. Updating strategies for isolating and discovering giant viruses. Curr. Opin. Microbiol. 31, 80–87.

Koonin, E.V., 1991. Genome replication/expression strategies of positive-strand RNA viruses: a simple version of a combinatorial classification and prediction of new strategies. Virus Genes 5, 273–281.
Koonin, E.V., 2009. Evolution of genome architecture. Int. J. Biochem. Cell Biol. 41, 298–306.
Koonin, E.V., 2010. The two empires and three domains of life in the postgenomic age. Nat. Edu. 3, 27.
Koonin, E.V., Starokadomskyy, P., 2016. Are viruses alive? The replicator paradigm sheds a decisive light on an old but misguided question. Stud. Hist. Philos. Biol. Biomed. Sci. 59, 125–134.
Koonin, E.V., Wolf, Y.I., 2012. Evolution of microbes and viruses: a paradigm shift in evolutionary biology? Front. Cell. Infect. Microbiol. 2, 119.
Koonin, E.V., Yutin, N., 2010. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology 53, 284–292.
Koonin, E.V., Senkevich, T.G., Chernos, V.I., 1993. Gene A32 product of vaccinia virus may be an ATPase involved in viral DNA packaging as indicated by sequence comparisons with other putative viral ATPases. Virus Genes 7, 89–94.
Koonin, E.V., Senkevich, T.G., Dolja, V.V., 2006. The ancient virus world and evolution of cells. Biol. Direct 1, 29.
Koonin, E.V., Dolja, V.V., Krupovic, M., 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. Virology 479-480, 2–25.
Kotwal, G.J., 2000. Poxviral mimicry of complement and chemokine system components: what's the end game? Immunol. Today 21, 242–248.
Kristensen, D.M., Cai, X., Mushegian, A., 2011. Evolutionarily conserved orthologous families in phages are relatively rare in their prokaryotic hosts. J. Bacteriol. 193, 1806–1814.
Kristensen, D.M., Waller, A.S., Yamada, T., Bork, P., Mushegian, A.R., et al., 2013. Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. J. Bacteriol. 195, 941–950.
Krupovic, M., Cvirkaite-Krupovic, V., 2011. Virophages or satellite viruses? Nat. Rev. Microbiol. 9, 762–763.
Krupovic, M., Koonin, E.V., 2015. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. Nat. Rev. Microbiol. 13, 105–115.
Krupovic, M., Bamford, D.H., Koonin, E.V., 2014. Conservation of major and minor jelly-roll capsid proteins in polinton (maverick) transposons suggests that they are bona fide viruses. Biol. Direct 9, 6.
Krupovic, M., Kuhn, J.H., Fischer, M.G., 2016. A classification system for virophages and satellite viruses. Arch. Virol. 161, 233–247.
La Scola, B., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., et al., 2003. A giant virus in amoebae. Science 299, 2033.
La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., et al., 2008. The virophage as a unique parasite of the giant mimivirus. Nature 455, 100–104.
Legendre, M., Arslan, D., Abergel, C., Claverie, J.M., 2012. Genomics of megavirus and the elusive fourth domain of life. Commun. Integr. Biol. 5, 102–106.
Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., et al., 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. Proc. Natl. Acad. Sci. U. S. A. 111, 4274–4279.
Legendre, M., Lartigue, A., Bertaux, L., Jeudy, S., Bartoli, J., et al., 2015. In-depth study of mollivirus sibericum, a new 30,000-y-old giant virus infecting acanthamoeba. Proc. Natl. Acad. Sci. U. S. A. 112, E5327–E5335.
Legendre, M., Fabre, E., Poirot, O., Jeudy, S., Lartigue, A., et al., 2018. Diversity and evolution of the emerging pandoraviridae family. Nat. Commun. 9, 2285.

Loeffler, F., Frosch, P., 1897. Summarischer Bericht uber die Ergebnisse der Untersuchungen der Kommission zur Erforschung der Maul- und Klauenseuche bei dem Institut fu$ r Infektionskrankheiten in Berlin. In: Centralblatt fuXr Bakteriologie, Parasitenkunde und Infektionskrankheiten, Abt. I. vol. 22, pp. 257–259.
Lopez-Garcia, P., 2012. The place of viruses in biology in light of the metabolism- versus-replication-first debate. Hist. Philos. Life Sci. 34, 391–406.
Maruyama, F., Ueki, S., 2016. Evolution and phylogeny of large DNA viruses, mimiviridae and phycodnaviridae including newly characterized heterosigma akashiwo virus. Front. Microbiol. 7, 1942.
Maumus, F., Epert, A., Nogue, F., Blanc, G., 2014. Plant genomes enclose footprints of past infections by giant virus relatives. Nat. Commun. 5, 4268.
Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R., 2005. The microbial pan-genome. Curr. Opin. Genet. Dev. 15, 589–594.
Moreira, D., Brochier-Armanet, C., 2008. Giant viruses, giant chimeras: the multiple evolutionary histories of mimivirus genes. BMC Evol. Biol. 8, 12.
Moreira, D., Lopez-Garcia, P., 2005. Comment on "the 1.2-megabase genome sequence of mimivirus". Science 308, 1114. author reply 1114.
Moreira, D., Lopez-Garcia, P., 2009. Ten reasons to exclude viruses from the tree of life. Nat. Rev. Microbiol. 7, 306–311.
Moreira, D., Lopez-Garcia, P., 2015. Evolution of viruses and cells: do we need a fourth domain of life to explain the origin of eukaryotes? Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 370, 20140327.
Moss, B., 2001. Poxviridae: the viruses and their replication. In: Fields, B.N., Knipe, D.M., Howley, P.M., Griffin, D.E. (Eds.), Fields Virology. Lippincott, Williams & Wilkins, Philadelphia, pp. 2849–2884.
Moss, B., Shisler, J.L., Xiang, Y., Senkevich, T.G., 2000. Immune-defense molecules of molluscum contagiosum virus, a human poxvirus. Trends Microbiol. 8, 473–477.
Nasir, A., Kim, K.M., Caetano-Anolles, G., 2012. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms archaea, bacteria and eukarya. BMC Evol. Biol. 12, 156.
Nasir, A., Kim, K.M., Caetano-Anolles, G., 2017. Phylogenetic tracings of proteome size support the gradual accretion of protein structural domains and the early origin of viruses from primordial cells. Front. Microbiol. 8, 1178.
Ogilvie, L.A., Jones, B.V., 2015. The human gut virome: a multifaceted majority. Front. Microbiol. 6, 918.
O'malley, M.A., Koonin, E.V., 2011. How stands the tree of life a century and a half after the origin? Biol. Direct 6, 32.
Palumbo, G.J., Buller, R.M., Glasgow, W.C., 1994. Multigenic evasion of inflammation by poxviruses. J. Virol. 68, 1737–1749.
Pearson, H., 2008. 'Virophage' suggests viruses are alive. Nature 454, 677.
Philippe, N., Legendre, M., Doutre, G., Coute, Y., Poirot, O., et al., 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341, 281–286.
Pope, W.H., Anders, K.R., Baird, M., Bowman, C.A., Boyle, M.M., et al., 2014. Cluster M mycobacteriophages bongo, pegleg, and rey with unusually large repertoires of tRNA isotypes. J. Virol. 88, 2461–2480.
Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490.
Puigbo, P., Wolf, Y.I., Koonin, E.V., 2009. Search for a 'Tree of Life' in the thicket of the phylogenetic forest. J. Biol. 8, 59.
Puigbo, P., Lobkovsky, A.E., Kristensen, D.M., Wolf, Y.I., Koonin, E.V., 2014. Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. BMC Biol. 12, 66.

Raoult, D., Forterre, P., 2008. Redefining viruses: lessons from mimivirus. Nat. Rev. Microbiol. 6, 315–319.
Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., et al., 2004. The 1.2-megabase genome sequence of mimivirus. Science 306, 1344–1350.
Reteno, D.G., Benamar, S., Khalil, J.B., Andreani, J., Armstrong, N., et al., 2015. Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. J. Virol. 89, 6585–6594.
Rodrigues, R.a.L., Abrahao, J.S., Drumond, B.P., Kroon, E.G., 2016. Giants among larges: how gigantism impacts giant virus entry into amoebae. Curr. Opin. Microbiol. 31, 88–93.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Santini, S., Jeudy, S., Bartoli, J., Poirot, O., Lescot, M., 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. U. S. A. 110, 10800–10805.
Schulz, F., Yutin, N., Ivanova, N.N., Ortega, D.R., Lee, T.K., et al., 2017. Giant viruses with an expanded complement of translation system components. Science 356, 82–85.
Schvarcz, C.R., Steward, G.F., 2018. A giant virus infecting green algae encodes key fermentation genes. Virology 518, 423–433.
Senkevich, T.G., Koonin, E.V., Bugert, J.J., Darai, G., Moss, B., 1997. The genome of molluscum contagiosum virus: analysis and comparison with other poxviruses. Virology 233, 19–42.
Senkevich, T.G., White, C.L., Koonin, E.V., Moss, B., 2002. Complete pathway for protein disulfide bond formation encoded by poxviruses. Proc. Natl. Acad. Sci. U. S. A. 99, 6667–6672.
Snel, B., Huynen, M.A., Dutilh, B.E., 2005. Genome trees and the nature of genome evolution. Annu. Rev. Microbiol. 59, 191–209.
Suttle, C.A., 2005. Viruses in the sea. Nature 437, 356–361.
Suttle, C.A., 2007. Marine viruses—major players in the global ecosystem. Nat. Rev. Microbiol. 5, 801–812.
Suzan-Monti, M., La Scola, B., Barrassi, L., Espinosa, L., Raoult, D., 2007. Ultrastructural characterization of the giant volcano-like virus factory of acanthamoeba polyphaga mimivirus. PLoS One 2, e328.
Szajner, P., Weisberg, A.S., Lebowitz, J., Heuser, J., Moss, B., 2005. External scaffold of spherical immature poxvirus particles is made of protein trimers, forming a honeycomb lattice. J. Cell Biol. 170, 971–981.
Tettelin, H., Riley, D., Cattuto, C., Medini, D., 2008. Comparative genomics: the bacterial pan-genome. Curr. Opin. Microbiol. 11, 472–477.
Tulman, E.R., Delhon, G.A., Ku, B.K., Rock, D.L., 2009. African swine fever virus. Curr. Top. Microbiol. Immunol. 328, 43–87.
Van Etten, J.L., Meints, R.H., 1999. Giant viruses infecting algae. Annu. Rev. Microbiol. 53, 447–494.
Weinbauer, M.G., 2004. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. 28, 127–181.
Wigington, C.H., Sonderegger, D., Brussaard, C.P., Buchan, A., Finke, J.F., et al., 2016. Re-examination of the relationship between marine virus and microbial cell abundances. Nat. Microbiol. 1, 15024.
Williams, T.A., Embley, T.M., Heinz, E., 2011. Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. PLoS One 6, e21080.
Wilson, W.H., Van Etten, J.L., Allen, M.J., 2009. The phycodnaviridae: the story of how tiny giants rule the world. Curr. Top. Microbiol. Immunol. 328, 1–42.

Woese, C.R., 1987. Bacterial evolution. Microbiol. Rev. 51, 221–271.
Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. U. S. A. 74, 5088–5090.
Woese, C.R., Kandler, O., Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. Proc. Natl. Acad. Sci. U. S. A. 87, 4576–4579.
Wolf, Y.I., Rogozin, I.B., Grishin, N.V., Koonin, E.V., 2002. Genome trees and the tree of life. Trends Genet. 18, 472–479.
Wommack, K.E., Colwell, R.R., 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64, 69–114.
Yau, S., Lauro, F.M., Demaere, M.Z., Brown, M.V., Thomas, T., et al., 2011. Virophage control of antarctic algal host-virus dynamics. Proc. Natl. Acad. Sci. U. S. A. 108, 6163–6168.
Yoshikawa, G., Askora, A., Blanc-Mathieu, R., Kawasaki, T., Li, Y., et al., 2018. Xanthomonas citri jumbo phage XacN1 exhibits a wide host range and high complement of tRNA genes. Sci. Rep. 8, 4486.
Yutin, N., Koonin, E.V., 2012. Hidden evolutionary complexity of nucleo-cytoplasmic large DNA viruses of eukaryotes. Virol. J. 9, 161.
Yutin, N., Koonin, E.V., 2013. Pandoraviruses are highly derived phycodnaviruses. Biol. Direct 8, 25.
Yutin, N., Wolf, Y.I., Raoult, D., Koonin, E.V., 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol. J. 6, 223.
Yutin, N., Colson, P., Raoult, D., Koonin, E.V., 2013. Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. Virol. J. 10, 106.
Yutin, N., Wolf, Y.I., Koonin, E.V., 2014. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology 466–467, 38–52.
Yutin, N., Shevchenko, S., Kapitonov, V., Krupovic, M., Koonin, E.V., 2015. A novel group of diverse polinton-like viruses discovered by metagenome analysis. BMC Biol. 13, 95.
Yutin, N., Backstrom, D., Ettema, T.J.G., Krupovic, M., Koonin, E.V., 2018. Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. Virol. J. 15, 67.