Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses

Eugene V. Koonin, Mart Krupovic, Natalya Yutin

Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *DNA Habitats and Their RNA Inhabitants*

Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses

Eugene V. Koonin,[1] Mart Krupovic,[2] and Natalya Yutin[1]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland. [2]Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Paris, France

Address for correspondence: Eugene V. Koonin, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894. koonin@ncbi.nlm.nih.gov

Diverse eukaryotes including animals and protists are hosts to a broad variety of viruses with double-stranded (ds) DNA genomes, from the largest known viruses, such as pandoraviruses and mimiviruses, to tiny polyomaviruses. Recent comparative genomic analyses have revealed many evolutionary connections between dsDNA viruses of eukaryotes, bacteriophages, transposable elements, and linear DNA plasmids. These findings provide an evolutionary scenario that derives several major groups of eukaryotic dsDNA viruses, including the proposed order "*Megavirales*," adenoviruses, and virophages from a group of large virus-like transposons known as Polintons (Mavericks). The Polintons have been recently shown to encode two capsid proteins, suggesting that these elements lead a dual lifestyle with both a transposon and a viral phase and should perhaps more appropriately be named polintoviruses. Here, we describe the recently identified evolutionary relationships between bacteriophages of the family *Tectiviridae*, polintoviruses, adenoviruses, virophages, large and giant DNA viruses of eukaryotes of the proposed order "*Megavirales*," and linear mitochondrial and cytoplasmic plasmids. We outline an evolutionary scenario under which the polintoviruses were the first group of eukaryotic dsDNA viruses that evolved from bacteriophages and became the ancestors of most large DNA viruses of eukaryotes and a variety of other selfish elements. Distinct lines of origin are detectable only for herpesviruses (from a different bacteriophage root) and polyoma/papillomaviruses (from single-stranded DNA viruses and ultimately from plasmids). Phylogenomic analysis of giant viruses provides compelling evidence of their independent origins from smaller members of the putative order "*Megavirales*," refuting the speculations on the evolution of these viruses from an extinct fourth domain of cellular life.

Keywords: Polintons; *Megavirales*; virus evolution; capsid proteins; translation

Introduction

Viruses are the most common and abundant biological entities on earth. Virome studies consistently show that in marine, soil, and animal-associated environments, the number of virus particles typically is 10–100 times greater than the number of cells.[1–3] Viruses and/or other selfish elements, such as transposons and plasmids, parasitize or enter symbiotic relationships with all cellular life forms, with the possible exception of some extremely reduced intracellular parasites.[4] The virus world displays an enormous diversity of genome structures and sizes (viral genomes consist of single-stranded (ss) or double-stranded (ds) RNA or DNA and range in size from less than 2 kb to over 2 Mb), replication and expression mechanisms, and virion structure.[5,6] Furthermore, the overall number of distinct genes present in the genomes of selfish elements appears to substantially exceed the number of genes in all cellular life forms. Viral genes typically evolve much faster than genes of cellular organisms, and as a result, most of the genetic diversity on earth is probably concentrated in the virus world.[2,4,7]

Viruses and related genetic elements do not share a single common ancestor: indeed, there are no genes that would be conserved in all or even in the majority of viral genomes.[8,9] However, viruses and other selfish genetic elements form a

doi: 10.1111/nyas.12728

Ann. N.Y. Acad. Sci. 1341 (2015) 10–24 © 2015 The Authors. *Annals of the New York Academy of Sciences* published by Wiley Periodicals Inc. on behalf of The New York Academy of Sciences.

complex network of evolutionary relationships in which genomes are linked through different sets of shared genes.[6] This evolutionary network apparently emerged owing to extensive gene exchange, often between widely different elements, as well as parallel acquisition of homologous genes from the hosts. Viruses with large genomes contain numerous genes acquired from the hosts at different stages of evolution. However, a small group of virus hallmark genes that encode key proteins involved in genome replication and virion formation, most notably capsid proteins, are represented in a broad variety of elements, comprising a substantial fraction of the edges in the evolutionary network.[6,8,9] Virus hallmark genes do not have obvious ancestors in cellular life forms, suggesting that some types of virus-like elements evolved at a precellular stage of the evolution of life.[10]

The viromes (i.e., the compendia of all viruses and virus-like elements) of prokaryotes (archaea and bacteria) and eukaryotes are dramatically different.[6,11] In prokaryotes, the great majority of viruses have dsDNA genomes, mostly within the range of 10–100 kb. The second most abundant class includes small ssDNA viruses. Retroelements comprise a small minority (no retroviruses are known), whereas RNA viruses are rare.

In stark contrast to bacteria and archaea, eukaryotes are hosts to numerous, enormously diverse RNA viruses, as well as retroelements and retroviruses.[12,13] Compared to RNA viruses and retroelements, ssDNA and dsDNA viruses and mobile elements are less diverse and less abundant in eukaryotes, although both of these classes of selfish elements have been found in most eukaryotes.[6] The dsDNA viruses of eukaryotes have recently enjoyed much attention and even publicity beyond academic circles, thanks to the unexpected discovery of giant viruses, with physical dimensions of the particles and genome sizes exceeding those of many bacteria and archaea and some parasitic unicellular eukaryotes.[14–17] In parallel with the study of giant viruses, and in part stimulated by this discovery, in the last few years, breakthroughs in phylogenomic analysis of eukaryotic dsDNA viruses have been achieved, leading to the emergence of detailed, well-supported evolutionary scenarios for the evolution of the major groups of these viruses. Here, we focus on the key results from these evolutionary genomic studies and discuss the unresolved problems in the evolution of eukaryotic dsDNA viruses.[a]

Diversity of dsDNA viruses in eukaryotes

Altogether, eukaryotes are hosts to 18 recognized families of dsDNA viruses that infect a broad spectrum of unicellular and multicellular organisms, and many unclassified viruses, spanning almost the entire range of viral genome sizes, from approximately 4 kb to almost 2.5 Mb (Table 1). By far the largest and most common group of DNA viruses in eukaryotes consists of seven families of large viruses (including giant mimiviruses and pandoraviruses, with genomes in the Mb range) that share a common viral ancestry, as indicated by the conservation of approximately 50 (inferred) ancestral genes.[15,18,19] This assemblage of virus families is known as nucleocytoplasmic large DNA viruses (NCLDV), or more recently, the proposed order "*Megavirales.*"[20] Notably, it is only the (putative) order "*Megavirales*" that encompasses viruses infecting a broad diversity of unicellular and multicellular eukaryotes; the other recognized families of eukaryotic dsDNA viruses are limited in their spread to individual eukaryotic kingdoms, primarily animals (Table 1).

The giant viruses of the family *Mimiviridae* are themselves "infected" with a distinct class of satellite viruses, known as virophages, that reproduce within the giant virus "factories" inside protist cells and depend on the giant virus for their replication.[21–24] Recently, an evolutionary link has been identified between the virophages and large eukaryotic dsDNA transposons of the Polinton/Maverick family (hereafter referred to as Polintons).[25,26] The Polintons are integrated within the genomes of diverse unicellular protists and animals, suggesting an ancient origin, perhaps coincident with the origin of eukaryotes, as well as substantial evolutionary success. Unexpectedly, we have recently found that the majority of the Polintons encode two proteins homologous to the typical capsid proteins of other viruses, strongly suggesting that, at least under some conditions, these transposons actually produce virions that could infect new hosts.[27] Thus, Polintons, which we believe should be more properly denoted

[a]This review is largely based on two recent articles (Refs. 28 and 78).

Table 1. The major groups of eukaryotic dsDNA viruses[a]

Virus family	Genome size range (kb)	Host range	Comments
Proposed order "*Megavirales*"			
Poxviridae	130–375	Animals	
Asfarviridae	165–190	Mammals, protists	
Iridoviridae	140–303	Animals, protists (?)	
Ascoviridae	150–190	Insects	
Marseilleviridae	346–386	Amoebae	
Phycodnaviridae	154–407	Algae, other protists (?)	
Mimiviridae	370–1.259	Amoebae, algae, other protists	
Pandoraviruses	1.908–2.473	Amoebae	Currently unclassified but likely to become a new family
Pithovirus	600	Amoebae	Currently unclassified but likely to become a new family
Polintoviruses	15–20	Vertebrates, insects, protists	Currently unclassified but likely to become a new family once the existence of virions is validated
Adenoviridae	26–48	Vertebrates	
Virophages	17–26	Satellites/parasites of protist-infecting mimiviruses	Currently unclassified but likely to become a new family
Order *Herpesvirales*			
Alloherpesviridae	134–295	Vertebrates	
Herpesviridae	125–241	Vertebrates	
Malacoherpsviridae	207	Molluscs	
Baculoviridae	80–180	Insects	
Hytrosaviridae	120–190	Insects	
Nimaviridae	300	Crustacea	
Nudiviridae	125–220	Insects	
Polydnaviridae	150–500	Insects	Encapsidated genomes of polydnaviruses consist of multiple dsDNA circles of variable size. The genes encoding the constituents of viral particles are permanently integrated into the insect genome and are not packaged
Papillomaviridae	6.8–8.4	Mammals	
Polyomaviridae	4.7–5.4	Mammals	

[a]When available, the data are from the latest report of the International Committee for Taxonomy of Viruses (ICTV).

polintoviruses, appear to lead a dual lifestyle combining features of viruses and transposons. Although polintoviruses remain to be identified experimentally, in the rest of this paper, we refer to the elements that encode the predicted capsid proteins as polintoviruses, while reserving the designation Polintons for those variants of these elements that appear to have lost the genes for the capsid proteins.

These findings prompted us to investigate in detail the evolutionary connections between the Polintons and other viruses and mobile elements.[28] The results of this analysis, which we discuss in the first part of the present review, point to polintoviruses as the first group of eukaryotic dsDNA viruses that evolved from prokaryotic viral ancestors and played a central role in the evolution of eukaryotic viruses and selfish elements.

Polintoviruses at the root of eukaryotic dsDNA virus, transposon, and plasmid evolution

The genomes of Polintons are large by transposon standards, 15–20 kb, and encode arrays of diverse proteins. However, several proteins are shared by all Polintons, including protein-primed type B DNA polymerase (pPolB), RVE family integrase, FtsK-like DNA-packaging ATPase, and an adenovirus-type cysteine protease implicated in the maturation of capsid proteins.[26,29–31] Polintons are flanked by terminal inverted repeats (TIRs) and, given the universal presence of the pPolB gene, are thought to replicate via the protein-primed mechanism. The terminal protein involved in DNA replication initiation has not been experimentally identified but is predicted to be the N-terminal domain of the polymerase.[32] The presence of genes for two capsid proteins, the putative virion-maturation protease and the genome-packaging ATPase, that are components of the virion morphogenesis apparatus in the members of the "*Megavirales*" implies that Polintons are virus-like transposons. Initially, however, no genes for capsid proteins were identified in the Polintons, rendering the presence of genes for proteins involved in capsid maturation enigmatic. However, recent extensive searches for distant sequence similarity have resulted in the identification of two capsid protein genes, one for the major protein of icosahedral capsids (double jelly-roll protein) and the other one for the minor penton protein.[27] The two proteins are essential and, in principle, sufficient to form icosahedral capsids. This finding explains the presence of other virion-maturation proteins that, with the capsid proteins, constitute a distinct structural–morphogenetic module and suggest that Polintons are actually polintoviruses, a novel group of dsDNA viruses that can exist also in the form of transposon-like provirus elements that are integrated with the host genomes and transmitted vertically across host generations. Some groups of polintoviruses have lost capsid protein genes and apparently became bona fide transposons, for which the name polintons can be retained. Polintoviruses are the second major group of eukaryotic dsDNA viruses, after the "*Megavirales*," that is represented in widely diverse unicellular eukaryotes along with many animals.[28] Polintons/polintoviruses are missing in only one of the five major eukaryotic superkingdoms, namely Archaeplastida, which includes land plants as well as red and green algae.

Polintoviruses share homologous genes and gene blocks with many other viruses, transposons, and plasmids (Fig. 1). In the network of evolutionary connections between all these groups of mobile elements where the edges are homologous genes, Polintoviruses represent the central hub that shares the maximum number of genes with other nodes.[26,28] Of special interest are the multiple connections between bacteriophages of the family *Tectiviridae*, polintoviruses, and the Mavirus virophage, which all share four genes encoding two capsid proteins, DNA-packaging ATPase, and protein-primed DNA polymerase (pPolB) (Fig. 1). Polintoviruses share two additional genes with Mavirus, namely those for the capsid maturation protease and the RVE integrase; the rest of the virophages share the capsid proteins, ATPase, and protease, but lack pPolB and the integrase. The adenoviruses join the network through pPolB, the two capsid proteins, and the protease, and the much larger "*Megavirales*" connect through the capsid proteins, the ATPase, and the protease. Thus, the morphogenetic module is the common denominator that joins all these diverse families of viruses into a polintovirus-centered assemblage. The yeast linear cytoplasmic plasmids bridge the elements with protein-primed replication and the much more complex "*Megavirales*:" these plasmids obviously lack the morphogenetic module but encode pPolB along with four key proteins that are conserved in most of the "*Megavirales*," namely the two largest subunits (the β- and β′-subunits) of multisubunit RNA polymerase, D11-like helicase, and the mRNA-capping enzyme, which show specific evolutionary relationship to the homologues from "*Megavirales*" (Fig. 1). This suite of proteins is believed to be essential for the cytoplasmic replication of DNA genomes.

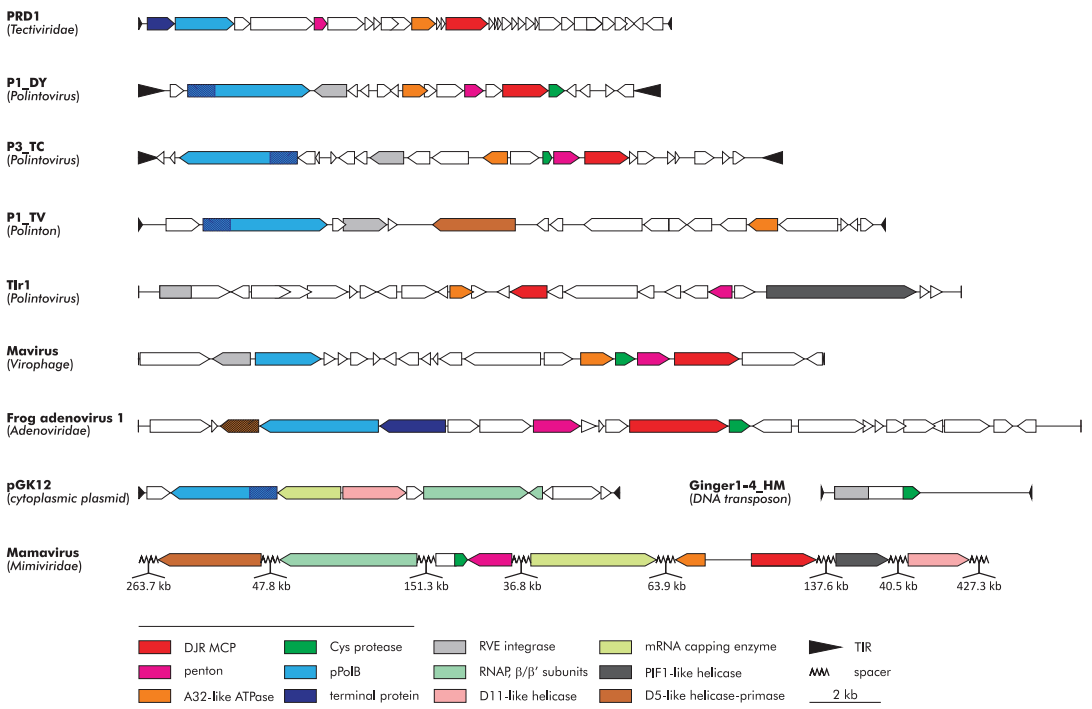

Figure 1. **Gene sharing between polintoviruses/Polintons and other groups of viruses, plasmids, and transposons. Homologous genes are color coded and the color key is provided at the bottom of the figure. Hatched regions in the pPolB genes indicate the position of the (predicted) terminal protein domains. Hatching is also used to indicate the gene encoding the distinct adenoviral genome packaging ATPase IVa2. P1_DY, polintovirus 1 of *Drosophila yakuba*; P3_TC, polintovirus 3 of *Tribolium castaneum*; P1_TV, polinton 1 of *Trichomonas vaginalis*. Modified from Ref. 28.**

From the connections between the polintoviruses and various other groups of viruses and plasmids, we have inferred a unifying scenario under which polintoviruses were the first group of eukaryotic dsDNA viruses and the font of eukaryotic virus, transposon, and plasmid evolution (Fig. 2).[28] Polintoviruses appear to have evolved from bacteriophages of the family *Tectiviridae*, at the onset of eukaryogenesis. This ancestral tectivirus most likely entered the protoeukaryotic cell along with the α-proteobacterial endosymbiont that subsequently gave rise to the mitochondrion (Fig. 2). This route of evolution is compatible with the presence of linear pPolB-encoding plasmids in fungal mitochondria; moreover, in phylogenetic trees of pPolB, the primary split is observed between the pPolBs of mitochondrial plasmids and the rest of the eukaryotic plasmids and viruses.[28]

The key difference between tectiviruses and polintoviruses is that the latter encompass genes for the Ulp1-like cysteine protease and the RVE family integrase. Both of these genes could have been acquired by the tectiviral ancestor of polintoviruses in a single event of recombination with a eukaryotic Ginger 1–like transposon.[33] Indeed, in this group of transposons, the integrase and protease domains are fused in the same polypeptide (Fig. 1). Some tectiviruses persist in bacteria in a linear plasmid form that is conducive for recombination.[34] The Ulp1-like deubiquitinases are characteristic eukaryotic enzymes, whereas bacteria encode only distantly related cysteine proteases, suggesting that the ancestor of polintoviruses had already acquired the protease and integrase genes in the (proto)eukaryotic host. The protease was adopted for capsid maturation and retained in all major virus lineages emerging from polintoviruses, including virophages, adenoviruses, and "*Megavirales,*" although some viruses in the latter group have lost this gene.[28,35]

The acquisition of the RVE integrase gene was a pivotal event in the evolution of the polintoviruses, endowing them with the ability to lead two alternative lifestyles, those typical of transposable elements and bona fide viruses. Such duality is also

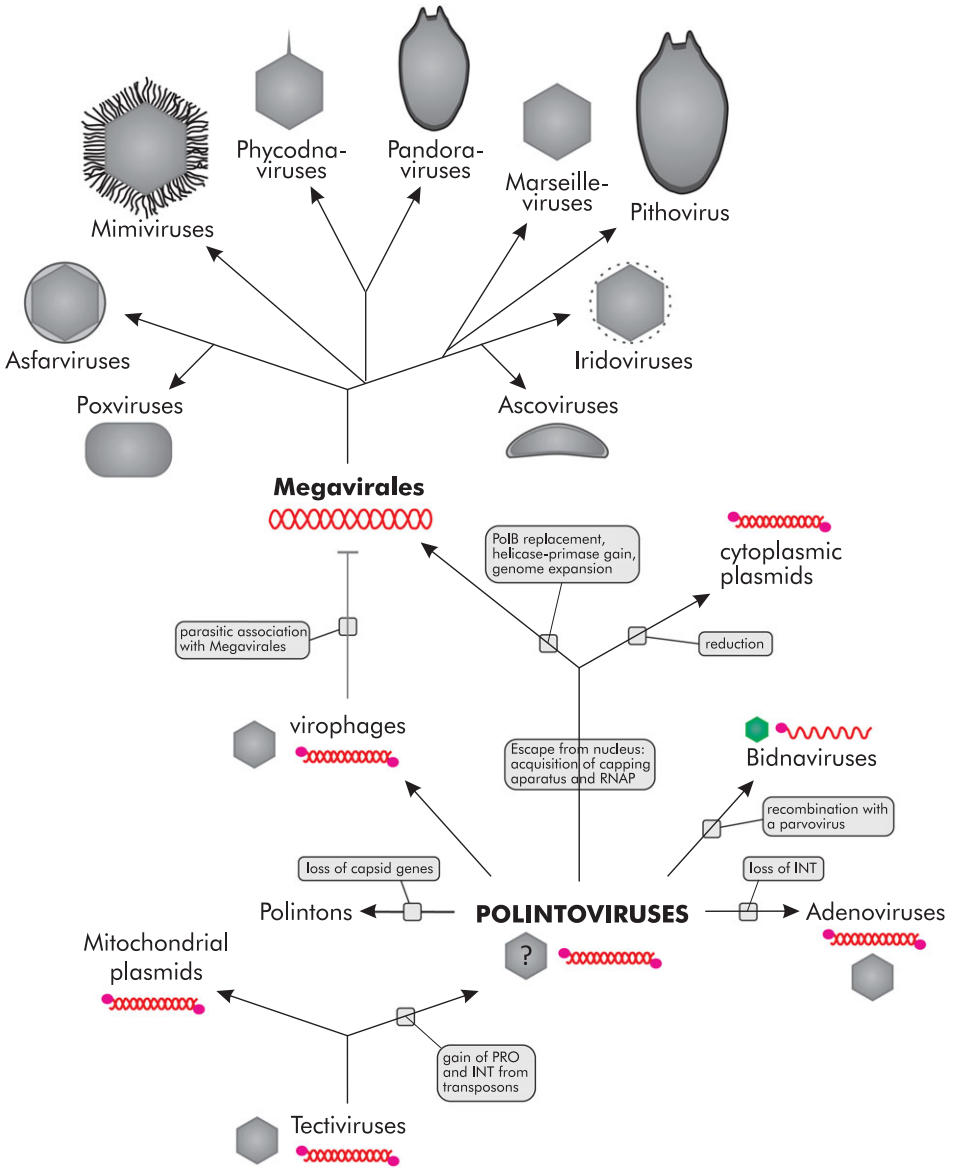

Figure 2. **The polintovirus-centered scenario of evolution for eukaryotic dsDNA viruses and plasmids. INT, RVE family integrase; RNAP, DNA-dependent RNA polymerase; PRO, cysteine protease; PolB, family B DNA polymerase. Green color of the bidnavirus virion indicates that the capsid protein is unrelated to that of polintoviruses. See text for details. Modified from Ref. 28.**

embraced by certain bacteriophages and eukaryotic Ty1-copia retrotransposons (also known as pseudoviruses) and Ty3-gypsy retrotransposons (metaviruses).[6,11,36] Conceivably, the ability to persist in the integrated form in the host genome and the ensuing flexibility of parasite–host interaction strategies played a key role in the further diversification and spread of polintoviruses (Polintons) and their derivatives in eukaryotes.

Some polintoviruses have given up the virus lifestyle altogether by losing the genes involved in virion formation and becoming pure transposons (Polintons).[27] A striking example is the extraordinary expansion of capsid-less Polintons in *Trichomonas vaginalis*, where they constitute up to 30% of the host genome.[30] Adenoviruses followed the opposite evolutionary route, losing the integrase gene together

with the integration/transposition ability (Fig. 2) and thereby committing to a strictly viral reproduction strategy. Polintoviruses also contributed the pPolB gene to the evolution of a remarkable family of ssDNA viruses, the *Bidnaviridae*, which emerged as a result of extensive gene shuffling between four groups of selfish elements.[32]

Most importantly, Polintoviruses seem to have played a key role in the emergence of the "*Megavirales*," the most abundant and diverse group of eukaryotic dsDNA viruses. The "*Megavirales*" apparently inherited from the Polintoviruses the virion morphogenetic module, consisting of the DJR major capsid protein (MCP), the genome-packaging ATPase, the maturation protease, and likely also the minor capsid (penton) protein (Fig. 1). Notably, among all the numerous DJR proteins, the major capsid protein of the polintoviruses is most closely related to those of phycodnaviruses,[27] suggesting a direct evolutionary connection between polintoviruses and the "*Megavirales*." The packaging ATPases and the maturation proteases show high levels of sequence divergence, complicating definitive phylogenetic analysis. However, the topologies of the respective phylogenetic trees are at least compatible with the existence of such a link.[26]

Polintoviruses reside in the nucleus of the host cell and, accordingly, rely on host enzymes for transcription. A key event for the emergence of the "*Megavirales*" was the escape from the nucleus, concomitant with the acquisition of the RNAP and the capping apparatus from the host. The escaped element that would replicate in the cytoplasm using the ancestral Polinton pPolB spawned two groups of mobile elements (Fig. 2), namely cytoplasmic plasmids (so far found only in fungi) and the "*Megavirales*," which share the unique trifunctional capping enzyme, RNAP, and D11-like helicase (Fig. 1). The cytoplasmic plasmids retain pPolB but have lost the morphogenetic module, thus succumbing to the exclusive intracellular lifestyle.

By contrast, the "*Megavirales*" evolved via the route of increasing complexity and autonomy. The essential events in the evolution of "*Megavirales*" from the cytoplasmic polintovirus-like ancestor include the replacement of pPolB with the RNA/DNA-primed PolB and acquisition of the D5-like helicase–primase (Fig. 2). It appears likely that pPolB, which, by definition, starts DNA replication at the end of the genome, cannot ensure efficient replication of genomes above a certain threshold (probably about 45 kb, as in adenoviruses), owing to the lack of a dedicated primase that would make multiple internal primers along the genome to ensure the completion of lagging strand synthesis. Thus, to sustain the reproduction of larger dsDNA genomes, the ability to prime synthesis on multiple sites in the genome appears to be required. Some polintoviruses encode divergent D5-like primases–helicases (Fig. 1) which often cluster with "*Megavirales*" in phylogenetic trees,[26] suggesting that "*Megavirales*" inherited this key enzyme from Polintons. Several other genes that are common and probably ancestral in the "*Megavirales*" are also shared with various Polintons.[26] The PolB of "*Megavirales*" was probably acquired from the eukaryotic host,[37] replacing the ancestral pPolB and, together with the primase–helicase, providing an opportunity for almost unlimited genome expansion. This expansion, which involved massive acquisition of genes not only from eukaryotes but also from bacteria,[38] was so dramatic that the genes apparently inherited from the polintoviruses constitute only a tiny fraction of the gene complement of the giant viruses. Nevertheless, a few losses notwithstanding, these proteins of polintovirus descent make up the morphogenetic module of the "*Megavirales*" as well as important parts of the replication apparatus.

The virophages retain many features of polintoviruses but followed a different strategy to adapt to reproduction in the cytoplasm of the host cells. Instead of encoding their own machinery for cytoplasmic propagation, these viruses evolved to parasitize their giant relatives by exploiting the transcription apparatus and other functions of the giant viruses.[22–25]

Evolution of eukaryotic dsDNA viruses beyond the polintovirus-related assemblage

There are 10 families of eukaryotic dsDNA viruses that do not show clear evolutionary relationship to the polintoviruses or the "*Megavirales*" (Table 1). These viruses are characterized by rather narrow host ranges, most of them being found only in animals. The evolutionary-genomic analysis of these viruses has not yet been performed in a comprehensive manner as it was done in the case

Ann. N.Y. Acad. Sci. 1341 (2015) 10–24 © 2015 The Authors. *Annals of the New York Academy of Sciences* published by Wiley Periodicals Inc. on behalf of The New York Academy of Sciences.

of "*Megavirales*." Nevertheless, a brief overview of what has become apparent is instructive.

Five families of large eukaryotic dsDNA viruses (*Baculoviridae*, *Hytrosaviridae*, *Nimaviridae*, *Nudiviridae*, and *Polydnaviridae*) have been isolated exclusively from arthropods, mainly insects. These viruses, especially the latter three families, encode highly diverged protein sequences and at this time are represented by only a handful of virus species (a single species in the case of *Nimaviridae*), hampering phylogenomic analysis. All these viruses seem to compose a monophyletic group that shares several signatures genes not found in other viruses.[39,40] Polydnaviruses are an extremely unusual group of viruses that propagate vertically, with the virus genomes permanently integrated in the genomes of the insect hosts. Nevertheless, phylogenetic analysis of the retained viral genes indicates that polydnaviruses are highly derived descendants of nudiviruses.[41] Preliminary phylogenetic study of highly conserved genes, such as the DNA polymerase, RNAP subunits, helicase–primase, and thiol oxidoreductase suggests that this entire assemblage of viruses could be a highly derived offshoot of the "*Megavirales*."[42] However, a comprehensive phylogenomic analysis of these very interesting viruses remains to be performed.

The highly diverse order *Herpesvirales* is of special interest from an evolutionary standpoint because, in this case, a distinct bacteriophage connection to the morphogenetic module is traceable. The bacteriophages involved, namely members of the order *Caudovirales*, are unrelated to the tectiviruses that apparently gave rise to the polintovirus-related majority of eukaryotic dsDNA viruses (Figs. 1 and 2). Herpesviruses share with the *Caudovirales* homologous major capsid proteins of the HK97 fold that are unrelated to the jelly-roll fold present in the capsid proteins of most icosahedral viruses (including the polintovirus-centered assemblage), terminases (packaging ATPase–nucleases), and capsid maturation proteases.[43–46] The inheritance of the morphogenesis module consisting of a capsid protein, ATPases, and protease from a bacteriophage closely parallels the similar evolutionary route of the polintovirus ancestor, but the actual proteins involved are unrelated (or in the case of the ATPase, distantly related). It appears somewhat paradoxical that the most common bacteriophages of the order *Caudovirales* gave rise to a single (even if diverse) family of eukaryotic dsDNA viruses, whereas the narrowly spread tectiviruses seeded the bulk of eukaryotic dsDNA virus evolution. Conceivably, the key event for the success of the polintoviruses that defined the wide spread of their derivatives was the acquisition of the integrase (see above). Furthermore, the fact that herpesviruses (so far) are limited to animal hosts might indicate that this virus family appeared relatively late in the course of eukaryotic evolution, with the ancestor bacteriophage coming not from the protomitochondrion but from a distinct (perhaps transient) bacterial symbiont of early animals.

Finally, the two families of dsDNA viruses with tiny, circular genomes, *Papillomaviridae* and *Polyomaviridae*, appear to share an origin that is completely distinct from the origins of all larger dsDNA viruses of eukaryotes. The large replicative protein of these viruses, known as the large T antigen in polyomaviruses and the E1 protein in papillomaviruses, is homologous to the replication proteins of ssDNA viruses, such as circoviruses, parvoviruses, and geminiviruses. This large protein has a typical domain architecture consisting of a superfamily 3 helicase and a rolling-circle replication-initiation endonuclease that, however, is inactivated in papillomaviruses and polyomaviruses, concomitant with the switch from rolling circle to the "θ-like" replication mode.[47,48] Thus, the small dsDNA viruses of eukaryotes are apparently derivatives of ssDNA viruses that themselves evolved via recombination of bacterial rolling circle–replicating plasmids and ssRNA viruses.[49]

Origin of giant viruses: dramatic inflation of viral genomes, not degeneration of cells from a fourth domain of life

The discovery of giant viruses (sometimes called "giruses") infecting protists, pioneered by the isolation of *Acanthamoeba polyphaga mimivirus*, is one of the most unexpected and certainly the most widely publicized breakthrough in virology in decades.[14,22,50–54] The giant viruses defy the textbook definition of viruses as "filterable" infectious agents because their enormous virions do not pass bacterial filters, and once and for all invalidate the separation between viruses and cellular life forms based on size. Not only are the particles of giant viruses bigger than the cells of numerous bacteria and archaea, but also the genomes of pandoraviruses, the current virus world record

Table 2. Origin of universal cellular genes in giant viruses

Gene/protein	Acanthamoeba polyphaga mimivirus	Acanthamoeba castellanii mamavirus	Acanthamoeba polyphaga lentillevirus	Acanthamoeba polyphaga moumouvirus	Moumouvirus	Moumouvirus Monve	Moumouvirus goulette	Courdo11 virus	Terra1 virus	Megavirus chiliensis	Megavirus courdo7	Megavirus courdo11	Cafeteria roenbergensis virus BV-PW1	Pandoravirus salinus	Pandoravirus dulcis	Pithovirus sibericum	Outcome of phylogenetic analysis
RNA polymerase, subunit α	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Inside eukaryotes, no 4th domain
RNA polymerase, subunit β	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Inside eukaryotes, no 4th domain
Arginyl-tRNA synthetase	Y	Y	Y	Y		Y	Y	Y	Y	Y		Y					Inside eukaryotes, no 4th domain
Aspartyl/asparaginyl-tRNA synthetase				Y		Y	Y			Y		Y					Inside bacterial/eukaryotic group
Cysteinyl-tRNA synthetase	Y	Y	Y		Y	Y	Y	Y	Y	Y		Y					Inside eukaryotes, no 4th domain
Isoleucyl-tRNA synthetase				Y			Y			Y	Y	Y	Y				Sister group to eukaryotes; formally compatible with 4th domain hypothesis
Methionyl-tRNA synthetase	Y	Y	Y			Y	Y	Y	Y	Y		Y					Inside eukaryotes, no 4th domain
Tryptophanyl-tRNA synthetase										Y		Y		Y			Inside eukaryotes; no 4th domain; different eukaryotic origins in mimiviruses and pandoraviruses
Tyrosyl-tRNA synthetase	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y		Inside eukaryotes; no 4th domain; different eukaryotic origins in mimiviruses and pandoraviruses
Pseudouridine synthase													Y				Inside bacteria, no 4th domain
Peptide chain release factor 1 (eRF1)						Y	Y			Y	Y	Y					Inside eukaryotes; no 4th domain; different eukaryotic origins in mimiviruses and pandoraviruses
Putative translation initiation inhibitor, yjgF family														Y	Y		Inside bacteria, no 4th domain
Putative translation factor (SUA5)														Y			Inside eukaryotes, no 4th domain
Translation elongation factor EF-1α (GTPase)	Y	Y	Y		Y	Y	Y	Y	Y	Y		Y					Inside eukaryotes, no 4th domain; polyphyletic origin among mimiviruses
Translation initiation factor 1 (eIF-1/SUI1)	Y	Y		Y		Y	Y			Y	Y	Y	Y				Inside eukaryotes, no 4th domain; polyphyletic origin within the family *Mimiviridae*

Continued

Table 2. *Continued*

Gene/protein	Acanthamoeba polyphaga mimivirus	Acanthamoeba castellanii mamavirus	Acanthamoeba polyphaga lentillevirus	Acanthamoeba polyphaga moumouvirus	Moumouvirus	Moumouvirus Monve	Moumouvirus goulette	Courdo11 virus	Terra1 virus	Megavirus chiliensis	Megavirus courdo7	Megavirus courdo11	Cafeteria roenbergensis virus BV-PW1	Pandoravirus salinus	Pandoravirus dulcis	Pithovirus sibericum	Outcome of phylogenetic analysis
Translation initiation factor 2 (IF-2; GTPase)													Y				Inside eukaryotes, no 4th domain
Translation initiation factor 2, α subunit (eIF-2α)													Y				Inside eukaryotes, no 4th domain
Translation initiation factor 2, β subunit (eIF-2β)/eIF-5 N-terminal domain													Y				Inside eukaryotes, no 4th domain
Translation initiation factor 2, γ subunit (eIF-2γ; GTPase)													Y				Inside eukaryotes, no 4th domain
Translation initiation factor 3, subunit g (eIF-3g)													Y				Inside eukaryotes, no 4th domain
Translation initiation factor 4F, cap-binding subunit (eIF-4E) and related cap-binding proteins	Y	Y	Y	Y		Y	Y			Y	Y	Y	Y	Y	Y		Inside eukaryotes, no 4th domain; polyphyletic origin within the "*Megavirales*"
Translation initiation factor 4F, helicase subunit (eIF-4A) and related helicases	Y	Y	Y								Y	Y	Y				Inside eukaryotes, no 4th domain; polyphyletic origin within the family *Mimiviridae*

The data are from Ref. 78.

holders at approximately 2.5 Mb, are substantially larger and richer in gene content than many bacterial and archaeal genomes, from both parasites and free-living microbes, and even the genomes of some parasitic eukaryotes, such as Microsporidia.[16] The recent identification of pandoraviruses and the pithovirus,[17] which are not only huge, by viral standards, but also possess a novel, asymmetrical virion structure, shows that the actual diversity of giant viruses remains to be uncovered.

The "cell-like" features of giant viruses prompted several researchers to propose the striking idea that giant viruses represent a "fourth domain of life" that is distinct from but comparable to the three cellular domains, bacteria, archaea, and eukaryotes.[14,23,52,55–57] The fourth-domain concept is sometimes presented as a general idea and on other occasions as a specific hypothesis on the evolution of viruses.[58] In general terms, the claim that giant viruses represent a fourth domain of life simply refers to the cell-like character of these viruses in terms of size of the virions and genomes and, in addition, to the observation that many genes of these viruses have no detectable homologs and so could have come from some unknown source. In this general form, the fourth-domain concept does not make any falsifiable predictions and, accordingly, is not a scientific hypothesis *sensu* Popper.

In contrast, the specific fourth-domain hypothesis drives directly from the original definition of the

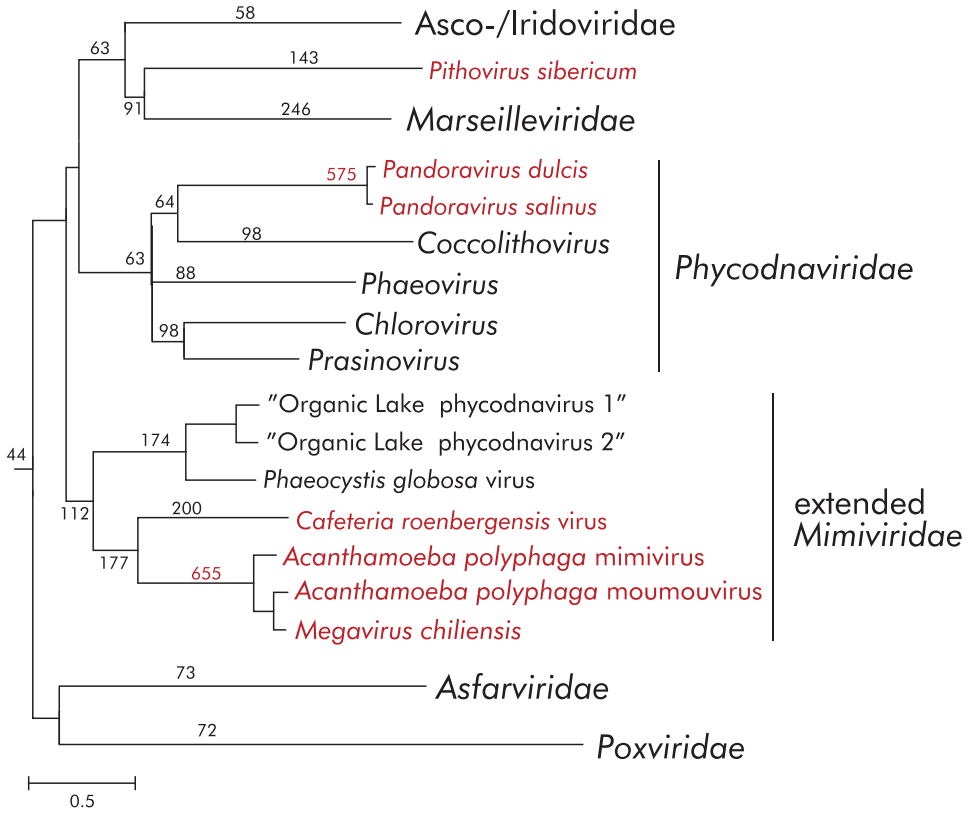

Figure 3. **Origin of giant viruses. The schematic tree shows the phylogeny of the genes that are conserved across the "*Megavirales*," with the three independently evolved groups of giant viruses highlighted in red. The numbers at internal branches indicated the estimated number of genes inferred to have been gained at the respective stage of evolution. The data are from Ref. 78.**

three domains of cellular life. These three domains, Bacteria, Archaea, and Eukaryota, are the three major trunks in the unrooted phylogenetic tree of 16S ribosomal RNA,[59–64] which is topologically consistent with the phylogenies of most of the other (nearly) universal genes that encode primarily components of the translation and the core transcription machineries.[65–67] Unlike other viruses, the giant viruses encode several proteins that are universal among cellular life forms, in particular translation system components, such as aminoacyl-tRNA synthetases and translation factors (Table 2). The presence of these universal genes allows one to formally include the giant viruses in the "tree of life" that constitutes the basis of the three-domain system.[14]

The outcome of the phylogenetic analysis of the universal genes of giant viruses is (at least, in principle) readily interpretable: the placement of the viral genes outside the three traditional domains of cellular life is compatible with the fourth-domain hypothesis, whereas their reliable placement inside any of the three domains would falsify this hypothesis. Several analyses, starting with the original study of the mimivirus genome, have reported phylogenetic trees that appeared compatible with the fourth-domain hypothesis.[14,68–70] However, there is an inherent problem with such observations. The point is that rapid evolution of viral genes, a common phenomenon that is likely to have occurred in the history of the giant viruses, especially immediately following the acquisition of the respective genes from the host, can obscure their affinity with homologs from cellular organisms within one of the recognized domains. This effect of nonuniform evolutionary rates is a pervasive problem in the analysis of deep phylogenies.[71] Indeed, a subsequent reanalysis of the phylogenies of several universal genes present in giant viruses has failed to find support for the fourth-domain hypothesis.[72]

Ann. N.Y. Acad. Sci. 1341 (2015) 10–24 © 2015 The Authors. *Annals of the New York Academy of Sciences* published by Wiley Periodicals Inc. on behalf of The New York Academy of Sciences.

Notwithstanding their unusual size, high genetic complexity, and the presence of some universal cellular genes, all giant viruses contain the set of core genes that define the NCLDV[15,18,73] or the proposed order "*Megavirales*."[20] Evolutionary reconstructions have mapped about 50 genes encoding essential viral functions to the putative common ancestor of the "*Megavirales*," although some of these putative ancestral genes have been lost in certain groups of viruses.[15,37,74] Importantly, this ancestral gene set does not include genes for components of the translation system or any other genes that might be considered suggestive of a cellular nature of the common ancestor of the "*Megavirales*" implied by the fourth-domain hypothesis.

Phylogenetic analysis of the genes that are conserved across the "Megavirales" reveals apparent evolutionary relationships between giant and smaller viruses. Specifically, the mimiviruses cluster with the so-called organic lake phycodnaviruses and *Phaeocystis globosa* viruses,[75,76] pandoraviruses cluster with phycodnaviruses, in particular coccolithoviruses,[77] and pithovirus clusters with marseilleviruses and iridoviruses[17] (Fig. 3). Together with reconstructions of gene-repertoire evolution from the phyletic patterns (matrices of gene presence and absence) of "*Megavirales*" genes, these relationships suggest that different groups of giant viruses independently evolved from smaller ancestral viruses.

The observation that giant viruses encompass many ancestral genes shared with the smaller members of "*Megavirales*" constrains the fourth-domain hypothesis to a specific version under which a viral ancestor of the giant viruses reproduced in a host that belonged to a fourth domain of cellular life and acquired numerous genes, including some that are universal in cellular life forms. In this scenario, after the fourth cellular domain went extinct, the giant viruses remained the only surviving "living fossils" of their original hosts.

In a recent study, we sought to formally and comprehensively test this specific fourth-domain hypothesis through phylogenetic analysis of all universal genes identified in giant viruses.[38] The results of this phylogenomic analysis effectively falsify the fourth-domain hypothesis by showing that the great majority of universal genes in giant viruses confidently cluster within the eukaryotic domain of the tree of life (Table 2). Furthermore, in cases where the same universal gene was detected in distantly related giant viruses, the respective genes typically were polyphyletic (e.g., tyrosyl-tRNA synthetase in mimiviruses and pandoraviruses; Table 2). Thus, these genes have been acquired by the giant viruses from the eukaryotic hosts at different stages of evolution, not inherited from a fourth domain of cellular life.

A complementary phylogenetic analysis of the genes that are conserved across the "*Megavirales*" clearly reveals the roots of the giant viruses (Fig. 3). Combined with probabilistic reconstruction of ancestral gene sets,[78] the phylogeny of the core genes shows that evolution of the "*Megavirales*" was generally characterized by genome expansion, which was pushed to the extreme in at least three independently evolved groups of giant viruses: (1) *Mimiviridae* that are a distant sister group of the *Phycodnaviridae*, (2) pandoraviruses that evolved from a common ancestor with coccolithoviruses within the *Phycodnaviridae*, and (3) pithoviruses that evolved from a common ancestor with *Iridoviridae* and *Marseilleviridae*.[17,38] Discovery of additional groups of giant viruses appears quite likely. Remarkably, the ancestral icosahedral capsid constructed from DJR MCPs was replaced with less regular, ovoid, or brick-shaped virions in several groups within the "*Megavirales*," namely in ascoviruses, poxviruses, pandoraviruses, and pithoviruses, indicating that all aspects of the viral life cycle are prone to dramatic change.

Conclusions

The world of viruses and selfish elements is enormously diverse, even when one considers only viruses with a certain type of genomic nucleic acid (dsDNA) that infect hosts from only one domain of cellular life (eukaryotes). The recent discovery of giant viruses infecting protists shows that we are still far from saturation in the study of viral diversity. Nevertheless, paraphrasing Einstein, the most astonishing finding about the virus world is that its history appears to be tractable. As discussed in this paper, we are now close to having coherent scenarios for the evolution of all major groups of dsDNA viruses of eukaryotes and perhaps deciphering certain general principles of this evolution. One such key principle is the unity of the evolutionary processes among viruses and capsid-less selfish elements.[6] However, the equally fundamental,

complementary principle is the evolutionary persistence of viral morphogenetic modules.[46,78] Certainly, the present reconstructions of the history of the virus world are schematic and only cover a small number of "hallmark" genes.[8] However, these genes encode most of the key functions in viral morphogenesis, replication, and expression, whereas the rest of the genes, of which giant viruses possess thousands, are involved in various aspects of virus–host interaction and widely differ between viruses. For all dsDNA viruses of eukaryotes, we can detect clear ancestry among viruses (and in some cases, plasmids) of bacteria, an observation that is compatible with the key role of the (proto)mitochondrial endosymbiont (and possibly other bacterial symbionts) in the evolution of eukaryotes.[79,80] Some of the most exciting ideas on viral evolution, such as the fourth-domain hypothesis, do not survive phylogenomic scrutiny. However, the emerging picture of virus evolution is hardly less remarkable.

Acknowledgments

E.V.K. and N.Y. are supported by intramural funds of the U.S. Department of Health and Human Services (to the National Library of Medicine).

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Edwards, R.A. & F. Rohwer. 2005. Viral metagenomics. *Nat. Rev. Microbiol.* **3:** 504–510.
2. Rohwer, F. 2003. Global phage diversity. *Cell.* **113:** 141.
3. Suttle, C.A. 2007. Marine viruses: major players in the global ecosystem. *Nat. Rev. Microbiol.* **5:** 801–812.
4. Koonin, E.V. & V.V. Dolja. 2013. A virocentric perspective on the evolution of life. *Curr. Opin. Virol.* **3:** 546–557.
5. Krupovic, M. & D.H. Bamford. 2010. Order to the viral universe. *J. Virol.* **84:** 12476–12479.
6. Koonin, E.V. & V.V. Dolja. 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol. Mol. Biol. Rev.* **78:** 278–303.
7. Kristensen, D.M. *et al.* 2013. Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. *J. Bacteriol.* **195:** 941–950.
8. Koonin, E.V., T.G. Senkevich & V.V. Dolja. 2006. The ancient Virus World and evolution of cells. *Biol. Direct.* **1:** 29.
9. Holmes, E.C. 2011. What does virus evolution tell us about virus origins? *J. Virol.* **85:** 5247–5251.
10. Koonin, E.V. 2009. On the origin of cells and viruses: primordial virus world scenario. *Ann. N. Y. Acad. Sci.* **1178:** 47–64.
11. Krupovic, M. *et al.* 2011. Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol. Mol. Biol. Rev.* **75:** 610–635.
12. Kazazian, H.H., Jr. 2004. Mobile elements: drivers of genome evolution. *Science* **303:** 1626–1632.
13. Goodier, J.L. & H.H. Kazazian, Jr. 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* **135:** 23–35.
14. Raoult, D. *et al.* 2004. The 1.2-megabase genome sequence of Mimivirus. *Science* **306:** 1344–1350.
15. Koonin, E.V. & N. Yutin. 2010. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. *Intervirology* **53:** 284–292.
16. Philippe, N. *et al.* 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* **341:** 281–286.
17. Legendre, M. *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.
18. Iyer, L.M., L. Aravind & E.V. Koonin. 2001. Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* **75:** 11720–11734.
19. Iyer, L.M. *et al.* 2006. Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res.* **117:** 156–184.
20. Colson, P. *et al.* 2013. "Megavirales," a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch. Virol.* **158:** 2517–2521.
21. La Scola, B. *et al.* 2008. The virophage as a unique parasite of the giant mimivirus. *Nature* **455:** 100–104.
22. Claverie, J.M. & C. Abergel. 2009. Mimivirus and its virophage. *Annu. Rev. Genet.* **43:** 49–66.
23. Desnues, C., M. Boyer & D. Raoult. 2012. Sputnik, a virophage infecting the viral domain of life. *Adv. Virus Res.* **82:** 63–89.
24. Krupovic, M. & V. Cvirkaite-Krupovic. 2011. Virophages or satellite viruses? *Nat. Rev. Microbiol.* **9:** 762–763.
25. Fischer, M.G. & C.A. Suttle. 2011. A virophage at the origin of large DNA transposons. *Science* **332:** 231–234.
26. Yutin, N., D. Raoult & E.V. Koonin. 2013. Virophages, polintons, and transpovirons: a complex evolutionary network of diverse selfish genetic elements with different reproduction strategies. *Virol. J.* **10:** 158.
27. Krupovic, M., D.H. Bamford & E.V. Koonin. 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.
28. Krupovic, M. & E.V. Koonin. 2015. Polintons: the hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat. Rev. Microbiol.* **13:** 105–115.
29. Kapitonov, V.V. & J. Jurka. 2006. Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **103:** 4540–4545.
30. Pritham, E.J., T. Putliwala & C. Feschotte. 2007. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **390:** 3–17.
31. Jurka, J. *et al.* 2007. Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* **8:** 241–259.

32. Krupovic, M. & E.V. Koonin. 2014. Evolution of eukaryotic single-stranded DNA viruses of the *Bidnaviridae* family from genes of four other groups of widely different viruses. *Sci. Rep.* **4:** 5347.
33. Bao, W., V.V. Kapitonov & J. Jurka. 2010. Ginger DNA transposons in eukaryotes and their evolutionary relationships with long terminal repeat retrotransposons. *Mob. DNA.* **1:** 3.
34. Gillis, A. & J. Mahillon. 2014. Phages preying on *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*: past, present and future. *Viruses* **6:** 2623–2672.
35. Yutin, N. & E.V. Koonin. 2012. Hidden evolutionary complexity of nucleo-cytoplasmic large DNA viruses of eukaryotes. *Virol. J.* **9:** 161.
36. Sandmeyer, S.B. & T.M. Menees. 1996. Morphogenesis at the retrotransposon-retrovirus interface: gypsy and copia families in yeast and Drosophila. *Curr. Top Microbiol. Immunol.* **214:** 261–296.
37. Filée, J., P. Forterre, T. Sen-Lin & J. Laurent. 2002. Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *J. Mol. Evol.* **54:** 763–773.
38. Yutin, N., Y.I. Wolf & E.V. Koonin. 2014. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* **466–467:** 38–52.
39. Wang, Y. & J.A. Jehle. 2009. Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: new insights on an old topic. *J. Invertebr. Pathol.* **101:** 187–193.
40. Jehle, J.A., A.M. Abd-Alla & Y. Wang. 2013. Phylogeny and evolution of Hytrosaviridae. *J. Invertebr. Pathol.* **112**(Suppl.): S62–S67.
41. Theze, J. *et al.* 2011. Paleozoic origin of insect large dsDNA viruses. *Proc. Natl. Acad. Sci. U.S.A.* **108:** 15931–15935.
42. Wang, Y., O.R. P. Bininda-Emonds & J.A. Jehle. 2012. "Nudivirus genomics and phylogeny." In *Molecular Structure, Diversity, Gene Expression Mechanisms and Host-Virus Interactions.* M. Garcia, Ed. Rijeka: InTech pp. 33–52.
43. Selvarajan Sigamani, S. *et al.* 2013. The structure of the herpes simplex virus DNA-packaging terminase pUL15 nuclease domain suggests an evolutionary lineage among eukaryotic and prokaryotic viruses. *J. Virol.* **87:** 7140–7148.
44. Pietilä, M.K. *et al.* 2013. Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. *Proc. Natl. Acad. Sci. U.S.A.* **110:** 10604–10609.
45. Rixon, F.J. & M.F. Schmid. 2014. Structural similarities in DNA packaging and delivery apparatuses in Herpesvirus and dsDNA bacteriophages. *Curr. Opin. Virol.* **5:** 105–110.
46. Krupovic, M. & D.H. Bamford. 2011. Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly. *Curr. Opin. Virol.* **1:** 118–124.
47. Ilyina, T.V. & E.V. Koonin. 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res.* **20:** 3279–3285.
48. Iyer, L.M. *et al.* 2005. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res.* **33:** 3875–3896.
49. Krupovic, M. 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Curr. Opin. Virol.* **3:** 578–586.
50. La Scola, B. *et al.* 2003. A giant virus in amoebae. *Science* **299:** 2033.
51. Koonin, E.V. 2005. Virology: Gulliver among the Lilliputians. *Curr. Biol.* **15:** R167–R169.
52. Claverie, J.M. *et al.* 2006. Mimivirus and the emerging concept of "giant" virus. *Virus Res.* **117:** 133–144.
53. Van Etten, J.L. 2011. Another really, really big virus. *Viruses* **3:** 32–46.
54. Van Etten, J.L., L.C. Lane & D.D. Dunigan. 2010. DNA viruses: the really big ones (giruses). *Annu. Rev. Microbiol.* **64:** 83–99.
55. Colson, P. *et al.* 2011. Viruses with more than 1000 genes: Mamavirus, a new Acanthamoeba polyphaga mimivirus strain, and reannotation of Mimivirus genes. *Genome Biol. Evol.* **3:** 737–742.
56. Yoosuf, N. *et al.* 2012. Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. *Genome Biol. Evol.* **4:** 1324–1330.
57. Legendre, M. *et al.* 2012. Genomics of Megavirus and the elusive fourth domain of Life. *Commun. Integr. Biol.* **5:** 102–106.
58. Forterre, P., M. Krupovic & D. Prangishvili. 2014. Cellular domains and viral lineages. *Trends Microbiol.* **22:** 554–558.
59. Woese, C.R. 1987. Bacterial evolution. *Microbiol. Rev.* **51:** 221–271.
60. Woese, C.R. & G.E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* **74:** 5088–5090.
61. Woese, C.R., O. Kandler & M.L. Wheelis. 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.
62. Woese, C.R., L.J. Magrum & G.E. Fox. 1978. Archaebacteria. *J. Mol. Evol.* **11:** 245–251.
63. Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. *Science* **276:** 734–740.
64. Pace, N.R. 2006. Time for a change. *Nature* **441:** 289.
65. Brown, J.R. & W.F. Doolittle. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiolm Molm Biolm Rev.* **61:** 456–502.
66. Brown, J.R. 2001. Genomic and phylogenetic perspectives on the evolution of prokaryotes. *Syst. Biol.* **50:** 497–512.
67. Puigbo, P., Y.I. Wolf & E.V. Koonin. 2009. Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J. Biol.* **8:** 59.
68. Colson, P. *et al.* 2012. Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. *Intervirology* **55:** 321–332.
69. Colson, P. *et al.* 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.
70. Nasir, A., K.M. Kim & G. Caetano-Anolles. 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.
71. Felsenstein, J. 2004. *Inferring Phylogenies.* Sunderland, MA: Sinauer Associates.
72. Williams, T.A., T.M. Embley & E. Heinz. 2011. Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. *PLoS One* **6:** e21080.
73. Iyer, L.M. *et al.* 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* **117:** 156–184.
74. Yutin, N. *et al.* 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* **6:** 223.
75. Yutin, N. *et al.* 2013. Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. *Virol. J.* **10:** 106.
76. Santini, S. *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.
77. Yutin, N. & E.V. Koonin. 2013. Pandoraviruses are highly derived phycodnaviruses. *Biol. Direct.* **8:** 25.
78. Yutin, N., Y.I. Wolf & E.V. Koonin. 2014. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* **466–467:** 38–52.
79. Krupovic, M. & D.H. Bamford. 2008. Virus evolution: how far does the double beta-barrel viral lineage extend? *Nat. Rev. Microbiol.* **6:** 941–948.
80. Embley, T.M. & W. Martin. 2006. Eukaryotic evolution, changes and challenges. *Nature* **440:** 623–630.
81. Martin, W. & E.V. Koonin. 2006. Introns and the origin of nucleus-cytosol compartmentation. *Nature* **440:** 41–45.