Chapter 8
A Decade of Giant Virus Genomics: Surprising Discoveries Opening New Questions

Hiroyuki Ogata and Masaharu Takemura

Core Message After the discovery of giant viruses at the beginning of this century, viral research started to exert important influences in ever-broader areas of biology. This chapter presents a review of the discoveries of giant viruses such as mimiviruses and pandoraviruses, their spectacular biology, and revolutionary ideas proposed for their origin and evolution, by particularly addressing the implications that have been brought to reassess our classical perception of a virus.

1   The Nature of Viruses: A Traditional View

Viruses are traditionally regarded as small biological entities, which were once termed *filterable agents*. Since their discovery in the late nineteenth century, and particularly after the observation of crystallized tobacco mosaic virus in 1935, they have rarely been regarded as living organisms. They have no cellular structure, the unit and a common trait of living organisms reproducing by binary division. Viruses do not produce energy (i.e., ATPs) required for their reproduction. In contrast, viruses first replicate their genomic material in large numbers, and then package the genomes into capsids. They cannot replicate autonomously outside their hosts but instead hijack host molecular machinery for their replication.

  Viral particles contain nucleic acids (either DNA or RNA [1]) as well as proteins in most cases, which are enclosed in a capsid. The capsid might in turn be covered by an envelope of lipid bilayer membranes for certain viruses. Viruses are classified

---

H. Ogata, Dr. Sc. (✉)
Bioinformatics Center, Institute for Chemical Research, Kyoto University,
Uji, Kyoto 611-0011, Japan
e-mail: ogata@kuicr.kyoto-u.ac.jp

M. Takemura, Dr. Med. Sc.
Faculty of Science, Tokyo University of Science, Tokyo 162-8601, Japan

into DNA viruses or RNA viruses according to the type of nucleic acids they carry in their particles. Viruses are too small to be visualized easily using optical microscopy. Even the observation of poxviruses (approximately 0.25 μm in length), the largest known viruses until the beginning of the twenty-first century, requires electron microscopy. Most viruses possess a gene (or genes) for genome replication (i.e., DNA polymerase or RNA polymerase), but they often lack genes for transcription and they never encode genes for translation machinery. Viruses depend on their host proteins for these latter steps of the central dogma (i.e., transcription and translation). Therefore, their metabolic capacity is crucially insufficient for autonomous self-reproduction.

A tremendously large body of research has been devoted to understand viruses from medical, agricultural, biochemical, and genetics perspectives. There had always been a clear-cut boundary between viruses and living organisms (i.e., life). However, an extremely large virus now called Acanthamoeba polyphaga mimivirus discovered in 2003 triggered a remarkable change in the perception of viruses, at least among certain microbiologists [2].

2 Discovery of Giant Viruses

Acanthamoeba polyphaga mimivirus (APMV) is an amoeba-infecting large DNA virus, with virus particles reaching 0.75 μm in diameter, including the glycosylated fibrous structure on the surface [3]. In microbiology laboratories, amoebas are used as tools to isolate bacterial pathogens such as *Legionella*. APMV was captured in this type of effort to isolate human pathogens in water samples from a cooling tower of a hospital in England by Timothy Rowbotham [4]. Its particle propagating in the amoeba culture was initially assumed to be an intracellular bacterium because of its large size comparable to small bacterial cells and for its Gram-positive staining property. The particle was therefore given the tentative name of "Bradford coccus," reflecting the name of the city of its isolation. However, efforts to amplify rRNA gene fragments were unsuccessful, leaving the characterization of the bacterium-like particles pending for years.

In 2003, the sample was brought to the group of Didier Raoult in France (Aix-Marseille University), who painstakingly examined the bacterium-like particles using electron microcopy. Unexpectedly, the particles had a regular, icosahedral form that was typical for a virion. Their reproduction cycle had an eclipse period, which was followed by the sudden emergence of hundreds of particles inside their host amoeba cells. The viral characteristics of the large particles were therefore revealed, and the virus was formally designated as "Mimivirus" to emphasize its size (i.e., a bacterium-"mimicking" virus) before its current name, APMV, was assigned.

In 2004, the complete genome sequence of APMV was determined [2]. The linear dsDNA genome, which turned out to be 1.18 Mbp in length, was found to encode more than 1,000 genes, most of which are transcribed during its infection cycle [5, 6]. The viral nature of APMV was also evident from its gene composition (e.g., capsid genes),

and gene phylogenies firmly placed APMV within the nucleocytoplasmic large DNA virus (NCLDV) group [7], which has been proposed, but not yet approved, as a new order, "*Megavirales*" [8]. Certain members of the NCLDV group, such as poxviruses, infecting vertebrates or insects, as well as chloroviruses, infecting unicellular algae such as chlorella, are large dsDNA viruses, already recognized as "giant viruses" even before the discovery of APMV [9] (Table 8.1). However, these classical giants of viruses measure only 0.18–0.25 μm and possess a genome of ca. 300 kbp. Therefore, APMV was truly exceptional in terms of its particle size and genome size among viruses known at that time.

Is APMV an intriguing, but unique, exception of the virosphere, standing at the extremity in the size spectrum of viruses? Alternatively, have researchers somehow missed opportunities to see and capture such giant viruses (now colloquially called as "giruses" [10])? Soon after the genome sequencing of APMV, comparative sequence studies of genetic data from environmental microbial samples suggested the existence of viruses related to APMV in marine ecosystems [11, 12]. Environmental samplings thus started with the aim of hunting the next giant viruses in different environments including marine ecosystems, and led to the isolation of mimivirus strains from diverse environments [13] and to the discoveries of new giant viruses in the sea, including a large virus infecting bacteriovorus marine nanoflagellate *Cafeteria roenbergensis* (CroV, 750 kbp), which confirmed the predicted presence of mimivirus relatives in the sea [14]. In 2011, again using amoeba cultures, another virus tentatively named "*Megavirus chilensis*" with a genome (1.26 Mbp) slightly larger than that of APMV was isolated from marine sediment sampled at a Chilean coast [15]. In 2013, Philippe et al. reported the discovery of pandoraviruses [16], the largest viruses ever found, with many features that had not been found in the giant viruses reported earlier.

3   Pandoraviruses

Pandoraviruses are atypical among large viruses in their virion morphology. Their virions are not icosahedral, but instead display an irregular ovoid form measuring 1 μm by 0.5 μm with a little apical pore, which makes it reminiscent of Pandora's jar in Greek mythology. They were identified as lytic agents of amoeba cultures, as in the case of mimiviruses and "*Megavirus chilensis*." Pandoravirions are visible by optical microscopy, and were initially given a nickname of "New Life Form (NLF)." Two similar particles were isolated: one from a sediment sample taken at the mouth of the Tunquen River, Chile, and the other from the bottom of a freshwater pond near Melbourne, Australia. Genome sequence analyses revealed that these two parasitic particles (respectively tentatively named "*Pandoravirus salinus*" and "*Pandoravirus dulcis*") represent related but distinct members of a newly proposed genus of giant virus. Except for regions with repetitive sequences at one extremity, their linear dsDNA genomes were sequenced completely. The size of the whole genome was estimated as 2.77 Mb (2,556 predicted genes) for "*P. salinus*" and as 1.91 Mbp (1,502 predicted genes) for "*P. dulcis*."

Table 8.1 Viral genomes larger than 300 kbp recorded in the NCBI/RefSeq database

Virus name	Genome size (bp)	Number of genes	Genome type	NCBI classification	Accession
"Pandoravirus salinus"	2473870[a]	2,541[a]	Linear	Viruses; dsDNA viruses, no RNA stage; unclassified dsDNA viruses[b]	NC_022098
"Pandoravirus dulcis"	1908524[a]	1,487[a]	Linear	Viruses; dsDNA viruses, no RNA stage; unclassified dsDNA viruses[b]	NC_021858
"Megavirus chiliensis"	1259197	1,120	Linear	Viruses; dsDNA viruses, no RNA stage; *Mimiviridae*; unclassified *Mimiviridae*[b]	NC_016072
"Megavirus lba"	1230522	1,176	Linear	Viruses; unclassified viruses[b]	NC_020232
Acanthamoeba polyphaga mimivirus (APMV)	1181549	979[a]	Linear	Viruses; dsDNA viruses, no RNA stage; *Mimiviridae*; *Mimivirus*[b]	NC_014649
"Acanthamoeba polyphaga moumouvirus"	1021348	894	Linear	Viruses; dsDNA viruses, no RNA stage; *Mimiviridae*; unclassified *Mimiviridae*[b]	NC_020104
Cafeteria roenbergensis virus BV-PW1 (CroV)	617453	544	Linear	Viruses; dsDNA viruses, no RNA stage; unclassified dsDNA viruses[b]	NC_014637
Phaeocystis globosa virus	459984	434	Linear	Viruses; dsDNA viruses, no RNA stage; *Phycodnaviridae*; *Prymnesiovirus*; unclassified prymnesiovirus[b]	NC_021312
Emiliania huxleyi virus 86	407339	472	Circular	Viruses; dsDNA viruses, no RNA stage; *Phycodnaviridae*; *Coccolithovirus*[b]	NC_007346
Paramecium bursaria Chlorella virus NY2A	368683	886	Linear	Viruses; dsDNA viruses, no RNA stage; *Phycodnaviridae*; *Chlorovirus*[b]	NC_009898
Acanthamoeba polyphaga marseillevirus (APMaV)	368454	428	Circular	Viruses; dsDNA viruses, no RNA stage; *Marseilleviridae*; *Marseillevirus*[b]	NC_013756
Canarypox virus	359853	328	Linear	Viruses; dsDNA viruses, no RNA stage; *Poxviridae*; *Chordopoxvirinae*; *Avipoxvirus*[b]	NC_005309
Cronobacter phage vB_CsaM_GAP32	358663	545	Linear	Viruses; dsDNA viruses, no RNA stage; *Caudovirales*; *Myoviridae*	NC_019401
"Lausannevirus"	346754	444	Circular	Viruses; dsDNA viruses, no RNA stage; *Marseilleviridae*; *Marseillevirus*[b]	NC_015326

Enterobacteria phage vB_KleM-RaK2	345809	534	Linear	Viruses; dsDNA viruses, no RNA stage; *Caudovirales; Myoviridae*	NC_019526
Paramecium bursaria chlorella virus AR158	344691	814	Linear	Viruses; dsDNA viruses, no RNA stage; *Phycodnaviridae; Chlorovirus*; unclassified chlorovirus[b]	NC_009899
Ectocarpus siliculosus virus 1	335593	240	Circular	Viruses; dsDNA viruses, no RNA stage; *Phycodnaviridae; Phaeovirus*[b]	NC_002687
Aramecium bursaria chlorella virus 1 (PBCV-1)	330611	802	Linear	Viruses; dsDNA viruses, no RNA stage; *Phycodnaviridae; Chlorovirus*[b]	NC_000852
Paramecium bursaria chlorella virus FR483	321240	849	Linear	Viruses; dsDNA viruses, no RNA stage; *Phycodnaviridae; Chlorovirus*[b]	NC_008603
Pseudomonas phage 201phi2-1	316674	461	Linear	Viruses; dsDNA viruses, no RNA stage; *Caudovirales; Myoviridae; Phikzlikevirus*; unclassified phiKZ-like viruses	NC_010821
Choristoneura biennis entomopoxvirus 'L'	307691	334	Linear	Viruses; dsDNA viruses, no RNA stage; *Poxviridae; Entomopoxvirinae; Betaentomopoxvirus*[b]	NC_021248
Shrimp white spot syndrome virus (white spot bacilliform virus)	305108	532	Circular	Viruses; dsDNA viruses, no RNA stage; *Nimaviridae; Whispovirus*	NC_003225

[a]Genome size and the number of predicted genes are based on the records in the NCBI/RefSeq database, and they slightly differ from the numbers from original publications cited in the main text

[b]Viruses classified in the NCLDV group

Pandoraviruses are also unique in terms of their gene contents. Only 401 predicted genes from "*P. salinus*" have significant sequence similarity to other sequences in the current sequence databases, whereas the remaining 2,155 predicted genes (84 %) lack detectable known homologs (i.e., "orphan genes"). A proteomic analysis of "*P. salinus*" virions identified 210 proteins originating from its genome, of which 80 % had no detectable sequence similarity to any other sequence in the public databases. The proteomic identification of gene products from "*P. salinus*" orphan genes suggests the *bona fide* gene status for the many of the predicted orphan genes in its genome.

Electron microscopic analysis revealed the following infection/replication cycle in *Acanthamoeba* cultures. Pandoravirions enter the cytoplasm of amoeba host cells by phagocytosis, as in the case of mimiviruses. The particle empties its contents (genomic DNA and probably associated proteins) into the cytoplasm through its apical pore. The injection process involves the fusion of the internal lipid membrane of the viral particle and the phagocytic vacuolar membrane. This genome delivery step is followed by an eclipse period during which the contents of the particles become invisible, as in other viruses. No binary division was observed. Later, the nuclear membranes disappear gradually and numerous newly assembled virion particles emerge at the periphery of the region formerly occupied by the nucleus. The location of the emergent pandoravirions is therefore different from those produced by mimiviruses and "*Megavirus chilensis*," which create an electron-dense intracellular compartment called a "virion factory," a viral replication and assembly center, in the cytoplasm of the infected cell. Therefore, pandoraviruses are assumed to use the host nucleus for their replication. The replication cycle lasts for 10–15 h.

Another interesting feature of pandoravirus genomes is the presence of spliceosomal type introns in many genes. Although precise delineation of exon/intron structure requires deep sequencing of transcripts, it is estimated that ca. 10 % of "*P. salinus*" genes contain spliceosomal introns. Spliceosomal introns differ from self-splicing introns (found in many viruses) and have been found only rarely in viral genomes. The presence of spliceosomal introns in pandoravirus genes is also indicative of the use of the host nucleus for their transcription.

Pandoraviruses share only a handful of genes with other previously characterized large DNA viruses. Therefore, their evolutionary relation with other viruses might not be readily apparent. Detailed phylogenetic analyses of a few genes common to pandoraviruses and other viruses suggest that they are distantly related to phycodnaviruses in the NCLDV group [17].

4   Are Viruses Alive?

Are viruses really non-organisms? The discoveries of these giant viruses strongly shook some of the beliefs of microbiologists and evolutionary biologists, and either reactivated or initiated old and new issues related to the concept of viruses. Succinctly put, a clear boundary that had been perceived between organisms (cellular life forms)
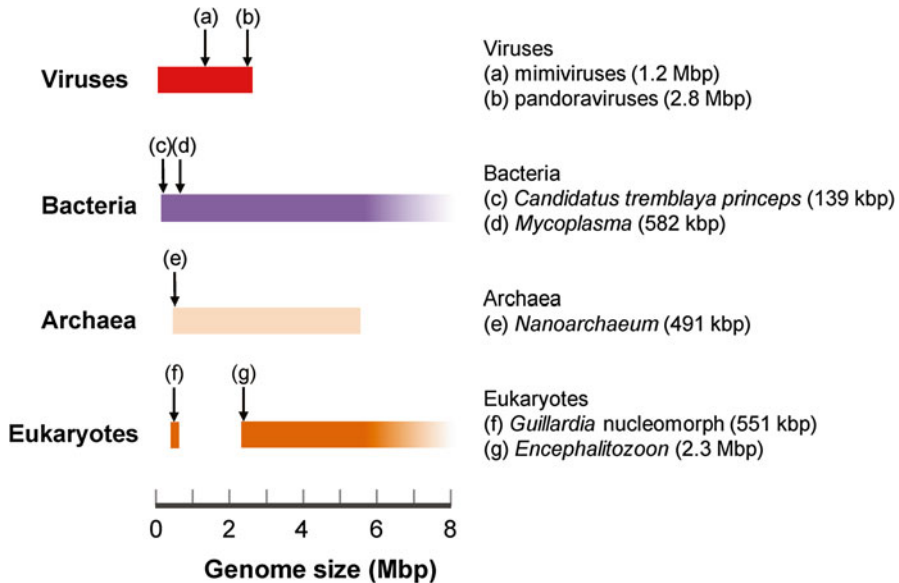
Fig. 8.1 Genome size of viruses and cellular organisms

and viruses became blurred considerably for the first time in the history of virology. Viruses had been thought to represent an ultimate form of parasite that carries a minimal set of genes that are necessary for nucleic acid replication and packaging. However, the size ranges of viral genomes and cellular genomes now mutually overlap. Mimivirus genome sizes exceed those of parasitic bacterial and archaeal genomes, whereas the size of the "*P. salinus*" genome falls in the size range of standard bacterial genomes and exceeds the size of parasitic eukaryotic genomes (Fig. 8.1).

Given the diversity of genes encoded in giant virus genomes, their reproduction strategies would not be expected to be simpler than those of cellular organisms. Furthermore, the APMV genome was found to encode genes for part of the translation system, which is regarded as a hallmark molecular apparatus distinguishing cellular organisms from viruses. APMV has four aminoacyl-tRNA synthetase genes and three genes for translation initiation, elongation, and termination, in addition to six tRNA genes. To date, no virus has been found that encodes genes for the ribosome, but the presence of these translation-related genes in APMV suggests that giant viruses actively participate in the translation process. They are not completely dependent on their host at every phase of the central dogma during their reproduction. CroV and "*Megavirus chilensis*" have several translational genes, although no such gene was identified in the genomes of pandoraviruses. In addition, APMV particles were found to contain both RNA (mRNAs) and DNA (genomic DNA), which further blurred the conceptual barrier between organisms and viruses.

The mimivirus DNA delivery system also illustrates how the molecular machineries of giant viruses are sophisticated [18]. Upon infection of particle contents into

the host cell cytoplasm, APMV opens five triangular faces around a vertex of its icosahedral capsid. The machinery, called "stargate," shows no structural similarity to the DNA delivery systems in other viruses such as the tails of bacteriophages.

The discovery of virophages supports the self-contained characteristics of the reproduction machinery of giant viruses, at least to a certain degree. Virophages are small viruses with genomes ca. 20 kbp dsDNA in length [19–22]. They are incapable of infecting cellular organisms independently, but they start reproduction when they are co-infected with a giant virus such as APMV. In fact, virophages infect the virion factory that giant viruses build inside the cytoplasm of the host. Infection of virophages can lead to abortive forms and abnormal capsid assembly of giant viruses. Therefore, virophages are small viruses (with their own DNA replication genes) that infect other larger viruses. The existence of virophages now appears to be a common phenomenon associated with giant viruses of the family *Mimiviridae* [23]. These observations indicate a high level of integrity and flexibility of the virion factory, and revived the old contention, that is, "are viruses not alive?" [24].

These discoveries during the last decade provided opportunities to reexamine the concept of viruses and their placement in the evolutionary history of life.

5   Fourth Domain Hypothesis

Mimiviruses possess several genes that are widely conserved in cellular organisms, such as RNA polymerase genes and aminoacyl-tRNA synthetase genes. Molecular-phylogenetic analyses suggest deep evolutionary origins for those genes, which might predate the radiation of the eukaryotic kingdom. Based on this observation, Raoult et al. reported that mimiviruses and related giant viruses might constitute a fourth domain of life [2], in addition to the other three established domains of life composed of eukaryotes, bacteria, and archaea (a domain is the highest taxonomic rank of organisms) (Fig. 8.2). This initial proposal was followed by others that supported the same idea or that extended the hypothesis by providing different evidence and arguments [25–29]. However, as expected, several lines of counter-argument have been raised by others [30–32].

In fact, some mimivirus gene phylogenies that were used to support the viruses' deep evolutionary origins in early studies later supported more recent origins after additional data for eukaryotic genomes were added [31]. Nevertheless, the phylogenetic analyses of different genes (including RNA polymerase and DNA polymerase gene) still support early branching positions for these genes near the roots of trees of cellular homologs.

In support of recent origins of giant virus genes, Moreira and Brochier-Armanet used phylogenetic analyses to prove that many mimivirus genes were acquired from cellular organisms by horizontal gene transfer in the course of evolution, and suggested "giant chimeric" characteristics of mimivirus genomes [33]. This type of result, implying that large virus genomes were derived from smaller virus genomes through the accretion of genes, enjoys certain popularity. However, the evidence of
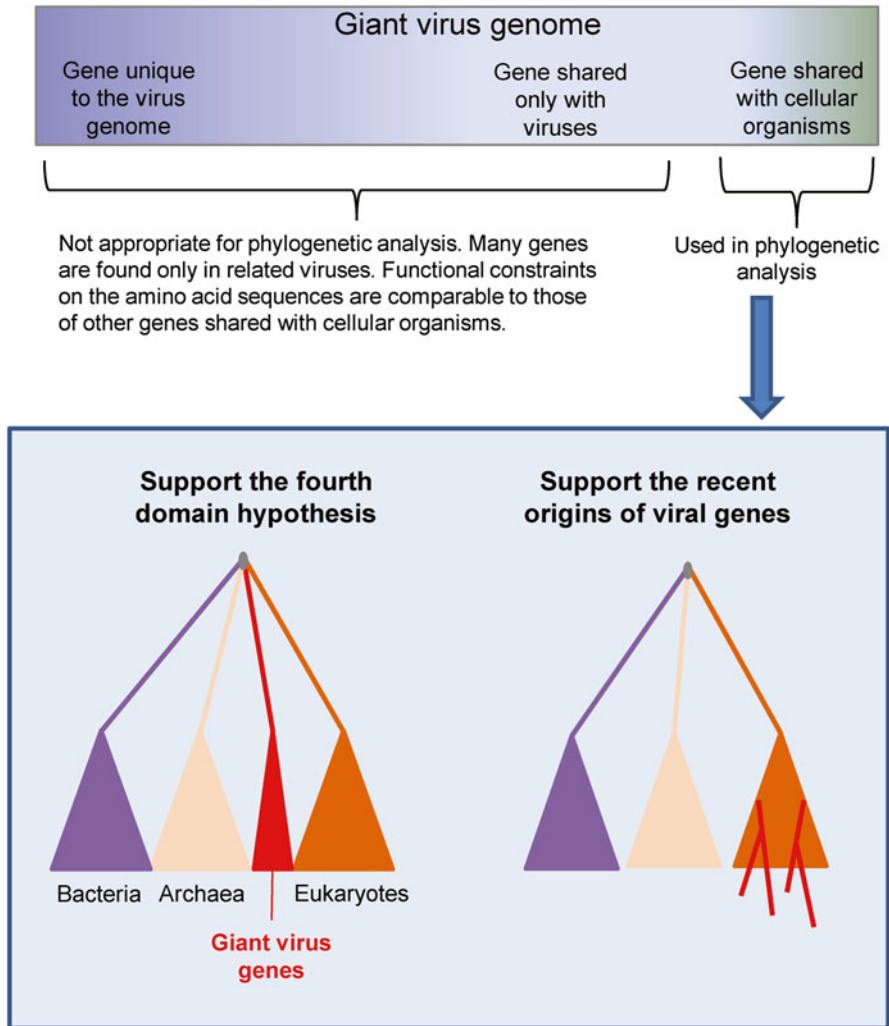
Fig. 8.2 Schematic drawing for the gene composition of typical giant virus genome and evolutionary analysis results

lateral gene acquisitions is in fact limited to a rather small subset (i.e., <10 %) of the entire gene set encoded in giant virus genomes [24, 34]. Such a level of detection of gene transfer is at the same level as that of bacterial genomes. For instance, "*P. salinus*" has 92 genes that might be of host amoeba origin, but this corresponds to only 3.6 % of the 2,556 genes encoded in its genome [16].

Giant viruses possess numerous genes with no detectable homologs in any cellular organism. Several authors have inferred that this fact might result from deep evolutionary origins of giant viruses [35]. More specifically, the existence of those

genes with obscure origins is not compatible with the classical idea that viral genomes are (mainly) derived from cellular genomes. A typical counter-argument to this invokes the high evolutionary rate of viral genomes that can erase the trace of homology between viral and cellular gene homologs. Ogata and Claverie refuted this counter-argument by demonstrating that no significant difference exists in the relative rates of evolution (more specifically, in the levels of functional constraints on sequences) between genes found only in closely related large DNA viruses and those with cellular homologs [36]. RNA viruses and ssDNA viruses (albeit at a lesser extent than RNA viruses) are known to evolve rapidly, but currently no reliable estimate exists for the evolutionary rate of giant virus genomes that can be compared directly with those of cellular genomes [37, 38].

The fourth domain issue would be revolutionary if the hypothesis is true, but the issue might be more complicated than the third (Archaea) domain proposition by Carl Woese in 1977 [39], which has now become widely established after a long debate (but see [40] for a recent discussion). Several important but different points might be readily apparent in the debates on the fourth domain hypothesis. They are discussed at different levels: some scientific and others epistemological. Crucial questions include the following: Are viruses organisms? Are viruses as old as cellular organisms? Even if we accept their deep ancestry, does the evidence from gene sequences support their old origins? How are they connected with the early history of the evolution of cells? Can we regard all viruses, from small RNA viruses to large DNA viruses, as a single biological group? Further characterization of giant viruses is expected to contribute to the resolution of these entangled issues.

6   Viral Origin of the Nucleus

Presumably, an important issue in the biology of giant viruses is the elucidation of the virion factory, an intracellular compartment for viral replication, and assembly that large viruses create inside their host cells. Nearly nothing is known about the virion factory, which can be as large as the nucleus and which would involve hundreds of viral proteins and other host factors. Investigation of the composition, structure, and function of the virion factory will definitely engender a better understanding of the nature of giant viruses. Here we briefly revisit a hypothesis that links the ancestor of large DNA viruses (and virion factory) with the origin of the nucleus.

Before the discovery of APMV, it had been proposed that eukaryotic DNA polymerase genes originated from ancient large DNA viruses based on the deep phylogenetic positions suggested for DNA polymerase genes of large DNA viruses [41, 42]. Takemura [42] and Bell [43] independently proposed further that an ancient large dsDNA virus infecting an archaeal ancestor of eukaryotes might be the origin of the eukaryotic nucleus (i.e., viral eukaryogenesis hypothesis). These authors identified several intriguing functional similarities between the nucleus and large DNA viruses such as poxviruses. Both poxviruses and the nucleus replicate only inside the
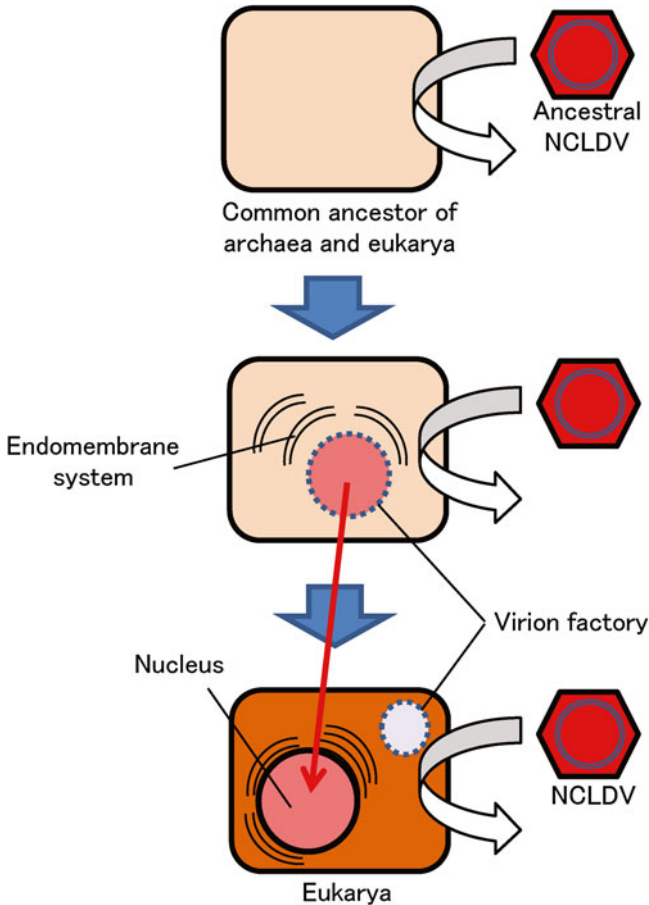
Fig. 8.3 Virion factory and the origin of the eukaryotic nucleus

cytoplasm of a eukaryotic cell. The translation process requires the translation system (ribosomes) located in the cytoplasmic region of the cell in both cases, although the spatial arrangement of cytoplasmic ribosomes and the virion factory are not known. Both poxviruses and the nucleus possess mechanisms to export mRNAs. The presence of repetitive sequences at the extremities of their linear dsDNA genomes is common between large DNA viruses and the nucleus. The virion factory of poxviruses arranges endoplasmic reticulum (ER) membranes at its periphery [44], reminiscent of the membrane surrounding the nucleus.

The viral eukaryogenesis hypothesis is an endosymbiotic hypothesis for the nucleus. It has been revisited and extended since the discovery of APMV and its large virion factory [45, 46] (Fig. 8.3). The endosymbiotic origin of the mitochondrion and the chloroplast is now widely accepted among biologists. In contrast, various theories have been proposed for the origin of the nucleus by researchers. These are

divisible into two categories, symbiotic or non-symbiotic theories, but none has yet been widely accepted. Symbiotic hypotheses are based on the symbiosis of organisms belonging to two species, such as archaeal and bacterial cells. The syntrophic eukaryogenetic theory, proposed by Moreira and López-García, invokes a syntrophic association of a sulfate-reducing δ-proteobacterium and a methanogenic archaeon [47]. However, the non-syntrophic eukaryogenetic theory, proposed by Cavalier-Smith, emphasizes the co-evolution of organelles including the nucleus, and postulates the fusion of ER membranes as the origin of the nuclear membranes [48]. Martin and Koonin hypothesized that nucleus–cytosol compartmentalization occurred to separate an mRNA splicing reaction, which proceeds more slowly, from a translation reaction, which proceeds more rapidly [49].

In spite of the proposal of these hypotheses corroborated by updated biological knowledge, an enigma remains. When and how did these events start? Was there a critical event that started everything, or did they occur gradually? The viral eukaryogenesis hypothesis has acquired more attention because of the discoveries of giant viruses and their properties consistent with the hypothesis, as described above. In a recent work, Takemura suggests that the infection of an ancestral NCLDV to the common ancestor of archaea and eukaryotes was a critical evolutionary event that spurred the emergence of the cell nucleus (Takemura, submitted).

Another effort to establish evolutionary and conceptual links between viruses with cellular organisms is the examination of the definition of viruses. For instance, Raoult and Forterre suggested the definition of viruses as "capsid-encoding organisms" and cellular organisms as "ribosome-encoding organisms" [50]. Traditionally, the term virus refers to a viral particle (i.e., virion) [1]. Claverie and Forterre proposed that the crucially important characteristics of "metabolically active state" of a virus reside in the virion factory [46] or in the whole infected cell (i.e., the "virocell" concept) [35, 51]. It is noteworthy that viral research is changing after the discoveries of giant viruses as well as other previously unrecognized viruses such as archaea viruses and symbiotic viruses, and now has marked influence in ever-broader areas of biology [52]. It is likely that further studies of giant viruses will continue to reveal their fascinating biology and will engender a unified evolutionary picture of the viral and cellular worlds.

References

1. Lwoff A. The concept of virus: the third Marjory Stephenson memorial lecture. J Gen Microbiol. 1957;17:239–53.
2. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM. The 1.2-megabase genome sequence of Mimivirus. Science. 2004;306:1344–50.
3. Rossmann MG. Structure of viruses: a short history. Q Rev Biophys. 2013;46:133–80.
4. Claverie JM, Abergel C, Ogata H. Mimivirus. Curr Top Microbiol Immunol. 2009;328: 89–121.
5. Legendre M, Santini S, Rico A, Abergel C, Claverie JM. Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. Virol J. 2011;8:99.

6. Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D, Lartigue A, Lescot M, Bernadac A, Poulain J, Abergel C, Claverie JM. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. Genome Res. 2010;20:664–74.

7. Yutin N, Wolf YI, Raoult D, Koonin EV. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol J. 2009;6:223.

8. Colson P, De Lamballerie X, Yutin N, Asgari S, Bigot Y, Bideshi DK, Cheng XW, Federici BA, Van Etten JL, Koonin EV, La Scola B, Raoult D. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch Virol. 2013;158:2517–21.

9. Van Etten JL, Meints RH. Giant viruses infecting algae. Annu Rev Microbiol. 1999;53:447–94.

10. Claverie JM, Ogata H, Audic S, Abergel C, Suhre K, Fournier PE. Mimivirus and the emerging concept of "giant" virus. Virus Res. 2006;117:133–44.

11. Ghedin E, Claverie JM. Mimivirus relatives in the Sargasso Sea. Virol J. 2005;2:62.

12. Monier A, Claverie JM, Ogata H. Taxonomic distribution of large DNA viruses in the sea. Genome Biol. 2008;9:R106.

13. Pagnier I, Reteno DG, Saadi H, Boughalmi M, Gaia M, Slimani M, Ngounga T, Bekliz M, Colson P, Raoult D, La Scola B. A decade of improvements in Mimiviridae and Marseilleviridae isolation from amoeba. Intervirology. 2013;56:354–63.

14. Fischer MG, Allen MJ, Wilson WH, Suttle CA. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci U S A. 2010;107:19508–13.

15. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci U S A. 2011;108:17486–91.

16. Philippe N, Legendre M, Doutre G, Coute Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie JM, Abergel C. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science. 2013;341:281–6.

17. Yutin N, Koonin EV. Pandoraviruses are highly derived phycodnaviruses. Biol Direct. 2013;8:25.

18. Zauberman N, Mutsafi Y, Halevy DB, Shimoni E, Klein E, Xiao C, Sun S, Minsky A. Distinct DNA exit and packaging portals in the virus *Acanthamoeba polyphaga mimivirus*. PLoS Biol. 2008;6:e114.

19. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, Merchat M, Suzan-Monti M, Forterre P, Koonin E, Raoult D. The virophage as a unique parasite of the giant mimivirus. Nature. 2008;455:100–4.

20. Yutin N, Raoult D, Koonin EV. Virophages, polintons, and transpovirons: a complex evolutionary network of diverse selfish genetic elements with different reproduction strategies. Virol J. 2013;10:158.

21. Fischer MG, Suttle CA. A virophage at the origin of large DNA transposons. Science. 2011;332:231–4.

22. Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Gibson JA, Cavicchioli R. Virophage control of Antarctic algal host-virus dynamics. Proc Natl Acad Sci U S A. 2011;108:6163–8.

23. Santini S, Jeudy S, Bartoli J, Poirot O, Lescot M, Abergel C, Barbe V, Wommack KE, Noordeloos AA, Brussaard CP, Claverie JM. Genome of *Phaeocystis globosa* virus PgV-16 T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. Proc Natl Acad Sci U S A. 2013;110:10800–5.

24. Ogata H, Claverie JM. Microbiology. How to infect a mimivirus. Science. 2008;321:1305–6.

25. Colson P, Gimenez G, Boyer M, Fournous G, Raoult D. The giant *Cafeteria roenbergensis* virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. PLoS One. 2011;6:e18935.

26. Legendre M, Arslan D, Abergel C, Claverie JM. Genomics of Megavirus and the elusive fourth domain of Life. Commun Integr Biol. 2012;5:102–6.

27. Nasir A, Kim KM, Caetano-Anolles G. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. BMC Evol Biol. 2012;12:156.

28. Wu D, Wu M, Halpern A, Rusch DB, Yooseph S, Frazier M, Venter JC, Eisen JA. Stalking the fourth domain in metagenomic data: searching for, discovering, and interpreting novel, deep branches in marker gene phylogenetic trees. PLoS One. 2011;6:e18011.

29. Claverie JM, Ogata H. Ten good reasons not to exclude giruses from the evolutionary picture. Nat Rev Microbiol. 2009;7:615. author reply 615.

30. Moreira D, Lopez-Garcia P. Comment on "The 1.2-megabase genome sequence of Mimivirus". Science. 2005;308:1114. author reply 1114.

31. Moreira D, Lopez-Garcia P. Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol. 2009;7:306–11.

32. Williams TA, Embley TM, Heinz E. Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. PLoS One. 2011;6:e21080.

33. Moreira D, Brochier-Armanet C. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. BMC Evol Biol. 2008;8:12.

34. Monier A, Claverie JM, Ogata H. Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. BMC Genomics. 2007;8:456.

35. Forterre P. Giant viruses: conflicts in revisiting the virus concept. Intervirology. 2010; 53:362–78.

36. Ogata H, Claverie JM. Unique genes in giant viruses: regular substitution pattern and anomalously short size. Genome Res. 2007;17:1353–61.

37. Moreau H, Piganeau G, Desdevises Y, Cooke R, Derelle E, Grimsley N. Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. J Virol. 2010;84:12555–63.

38. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol. 2010;84:9733–48.

39. Goldenfeld N, Pace NR. Retrospective. Carl R. Woese (1928-2012). Science. 2013;339:661.

40. Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only two primary domains of life. Nature. 2013;504:231–6.

41. Villarreal LP, DeFilippis VR. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. J Virol. 2000;74:7079–84.

42. Takemura M. Poxviruses and the origin of the eukaryotic nucleus. J Mol Evol. 2001;52:419–25.

43. Bell PJ. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? J Mol Evol. 2001;53:251–6.

44. Tolonen N, Doglio L, Schleich S, Krijnse Locker J. Vaccinia virus DNA replication occurs in endoplasmic reticulum-enclosed cytoplasmic mini-nuclei. Mol Biol Cell. 2001;12:2031–46.

45. Bell PJ. Sex and the eukaryotic cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. J Theor Biol. 2006;243:54–63.

46. Claverie JM. Viruses take center stage in cellular evolution. Genome Biol. 2006;7:110.

47. Moreira D, Lopez-Garcia P. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J Mol Evol. 1998;47:517–30.

48. Cavalier-Smith T. Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. Biol Direct. 2010;5:7.

49. Martin W, Koonin EV. Introns and the origin of nucleus-cytosol compartmentalization. Nature. 2006;440:41–5.

50. Raoult D, Forterre P. Redefining viruses: lessons from Mimivirus. Nat Rev Microbiol. 2008;6:315–9.

51. Forterre P. The virocell concept and environmental microbiology. ISME J. 2013;7:233–6.

52. Dolja VV, Krupovic M. Accelerating expansion of the viral universe. Curr Opin Virol. 2013;3:542–5.