Available online at www.sciencedirect.com
SCIENCE DIRECT
Virus Research 117 (2006) 145–155
Virus Research
www.elsevier.com/locate/virusres

Review

Genomic and evolutionary aspects of *Mimivirus*

M. Suzan-Monti, B. La Scola, D. Raoult *

Unité des Rickettsies et Pathogènes Emergents, Faculté de Médecine, IFR 48, CNRS UMR 6020, Université de la Méditerranée, 27 Bd Jean Moulin, 13385 Marseille Cedex 05, France

Available online 21 September 2005

Abstract

We recently described a giant double stranded DNA virus called *Mimivirus,* isolated from amoebae, which might represent a new pneumonia-associated human pathogen. Its unique morphological and genomic characteristics allowed us to propose *Mimivirus* as a member of a new distinct Nucleocytoplasmic Large DNA viruses family, the *Mimiviridae*. *Mimivirus*-specific features, namely its size and its genomic complexity, ranged it between viruses and cellular organisms. This paper reviews our current knowledge on *Mimivirus* structure, life cycle and genome analysis and discusses its putative evolutionary origin in the tree of species of the three domains of life.

Keywords: *Mimivirus*; Large dsDNA viruses; Structure; Genome; Evolution

Contents

Abbreviations: COG, cluster of ortholog group; dsDNA, double stranded DNA; LLAP, Legionella-like Amoebal pathogen; NCLDV, Nucleocytoplasmic Large DNA Virus; p.i., post-infection; rRNA, ribosomal RNA

* Corresponding author. Tel.: +33 4 91 32 43 75; fax: +33 4 38 77 72.
E-mail addresses: marie.suzan@medecine.univ-mrs.fr (M. Suzan-Monti), didier.raoult@medecine.univ-mrs.fr (D. Raoult).

0168-1702/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2005.07.011

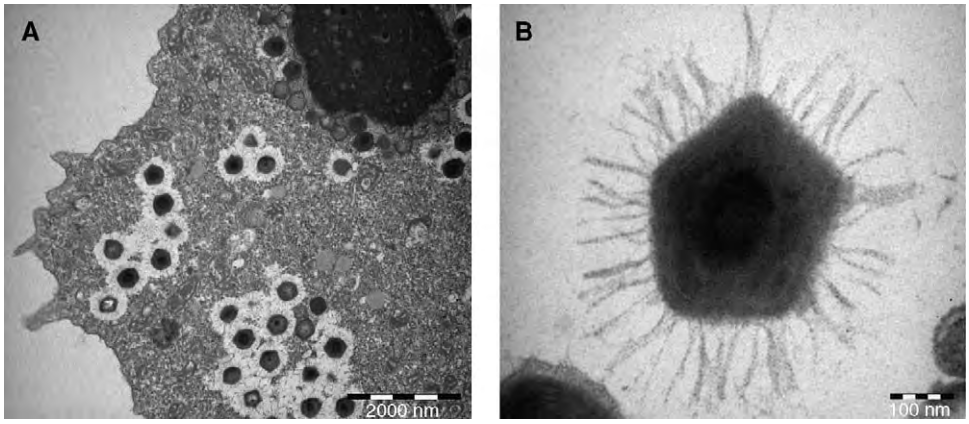

Fig. 1. Appearance of *Mimivirus* in infected cells and in cell-free supernatant. Transmission electron microscopy on *Mimivirus*-infected *A. polyphaga* at 8 h p.i. shows intra-cytoplasmic accumulation of virus particles (A). Bar = 2 μm. *Mimivirus* particles, purified from the supernatant of infected cells at 72 h p.i. as previously described (La Scola et al., 2003), are non-enveloped icosahedral virions surrounded by fibrils (B). Bar = 100 nm.

1. Discovery

Cultures of amoebae with water from a cooling tower during an outbreak of pneumonia in Bradford, England, led T.J. Rowbotham to isolate a series of Legionella-like Amoebal Pathogens (LLAP), among which one was Gram-positive and was first called "Bradford coccus" (unpublished results). Different attempts in our laboratory to identify this amoebae-associated bacterium by universal bacterial 16S ribosomal RNA (rRNA) PCR amplification and sequencing reproducibly failed. Initial electron microscopy observations of infected *Acanthamoebae polyphaga* revealed the presence of an intra-cytoplasmic icosahedral structure, surrounded by fibrils of about 650 nm diameter (Fig. 1A). These observations were reproduced on material purified from infected *A. polyphaga* cell-free supernatant (Fig. 1B) by ultracentrifugation on a sucrose cushion (La Scola et al., 2003). This estimated size is larger than that of small mycoplasma *U. urealyticum* (La Scola et al., 2003) and comparable to those of intracellular bacteria such as *R. conorii* or *T. whipplei* (Raoult, 2005). The morphology of what we then called *Mimivirus* (for Mimicking Microbe) was reminding that of *Iridoviruses*, *Asfarviruses* or *Phycodnaviruses* (Salas, 1999; Van Etten et al., 2002; Williams, 1996), three members of the Nucleocytoplasmic Large DNA Virus (NCLDV) family, infecting poikilothermic invertebrate animals, pigs, and eukaryotic algae, respectively. However, the size of *Mimivirus* particle makes it the largest virus ever described until now. *A. polyphaga* is a pathogenic ubiquitous free-living amoebae. Although the natural history of *Mimivirus* is poorly understood, its discovery as an amoebae-associated pathogen makes it a potential causative agent of pneumonia in humans. Currently, we have only indirect evidences supporting this idea since recent results demonstrated the presence of *Mimivirus*-specific antibodies in the sera of patients with community- or hospital-acquired pneumonia (La Scola et al., 2005).

2. *Mimivirus* morphology, life cycle and cellular tropism

2.1. Morphological characteristics of the viral particle

The main characteristics of *Mimivirus* particles observed by transmission electron microscopy are its size of about 650 nm, its icosahedral capsid and the presence of fibrils surrounding the capsid. From outside to inside, the capsid seems to be composed of several membrane layers inside which condensed core material could be observed (Figs. 1B and 2A). In some cases, the core material appeared uncondensed (Fig. 2B). Other morphological aspects of *Mimivirus* particles are shown in Fig. 2C and D. Whether these various forms represent different stages of *Mimivirus* maturation process or defective virions remains to be established. *Mimivirus* particles seem devoid of an outer membrane, suggesting that the way they use to exit from infected cells is not through budding at the cellular cytoplasmic membrane. *Mimivirus* particles appear to share morphological characteristics in common with *Poxviruses* (Moss, 2001), *Iridoviruses* (Williams, 1996), *Phycodnaviruses* (Van Etten et al., 2002)

Table 1
Viral transcripts detected in *Mimivirus* particles

Gene	Definition/putative function	Relationship to other viral genes
R322	DNA polymerase	NCLDV core gene, class I
L425	Capsid protein	NCLDV core gene, class I
R339	TFII-like transcription factor	NCLDV core gene, class II
L124	Tyrosyl-tRNA synthetase	Specific, translation
L164	Cysteinyl-tRNA synthetase	Specific, translation
R663	Arginyl-tRNA synthetase	Specific, translation
L272	Unknown	Specific, orphan
R349	Unknown	Specific, orphan
L520	Unknown	Specific, orphan
L611	Unknown	Specific, orphan

Genes transcripts were detected by RT-PCR analysis on RNAs extracted from purified virions as previously described (Raoult et al., 2004).

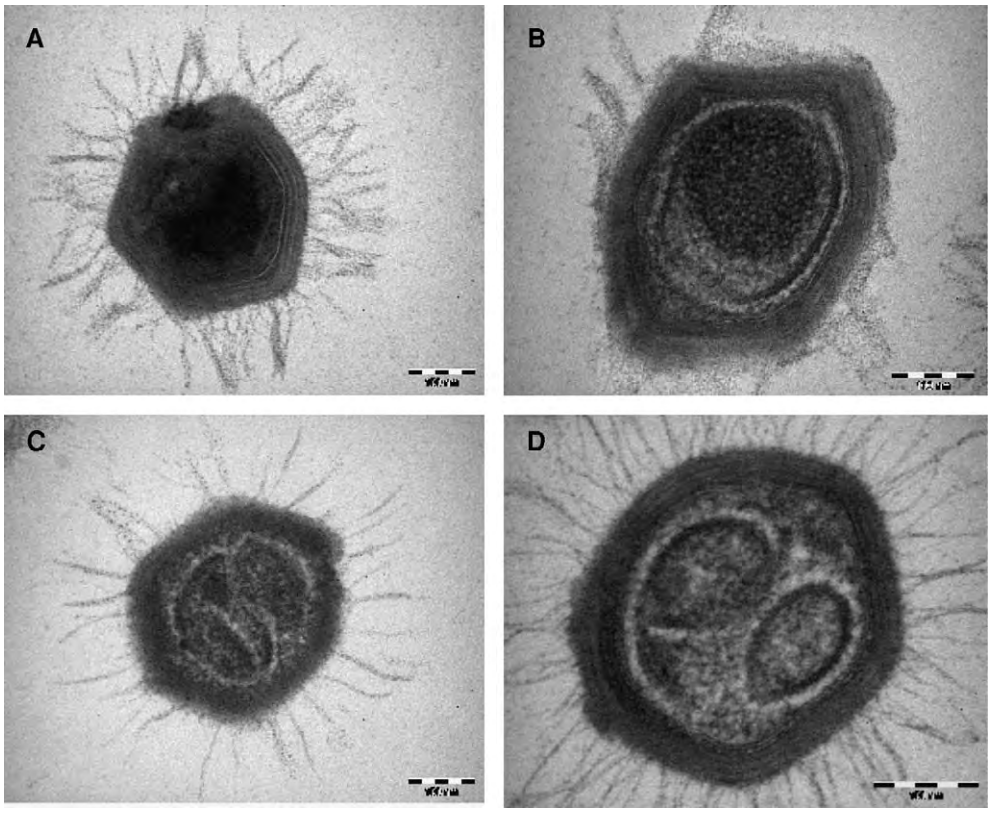

Fig. 2. Different morphological aspects of *Mimivirus* particles. Transmission electron microscopy on cell-free supernatant purified virions shows either mature-like particles with a condensed core (A) or immature-like particles with uncondensed core (B). Diverse ultrastructural pictures of the encapsidated material could be observed (C and D), which might be representative of defective particles. Bar = 100 nm.

and *Asfarviruses* (Brookes et al., 1996), all members of the NCLDV family. An internal lipid membrane surrounding the central core was described in each case, that might also be present in *Mimivirus* particles (Figs. 1B and 2). However, the presence of external fibrils, about 120 nm long, is specific to *Mimivirus*. They seemed to be formed by proteins and glucidic components since they are sensitive to protease and lyzozyme digestion (La Scola et al., submitted manuscript), and one might speculate about their role in the attachment of *Mimivirus* to its target cell. Virion density was determined to be 1.36 g/cm^3 in CsCl (La Scola et al., submitted manuscript).

The condensed core contained the viral genome which is an open 1.2 Mb linear dsDNA (Raoult et al., 2004). We also demonstrated that some viral RNAs were packaged within the virus particle (Raoult et al., 2004; La Scola et al., submitted manuscript). Out of the 27 transcripts tested so far, 10 could be amplified by RT-PCR on RNA extracted from density gradient purified virions (Table 1). We found the transcripts of three genes of the NCLDV class I (R322 DNA polymerase, L425 Capsid protein) and class II (R339 TFII-like transcription factor) core genes as defined by Iyer et al. (2001); three genes coding for amino acyl-tRNA synthetases (L124, L1164 and R663) described for the first time in a

Table 2
Representative large dsDNA viruses

Virus[a]	Virus family	Genome size (bp)	No. of protein coding genes	No. of proteins in virion	Reference
Mimivirus	Mimiviridae	1,181,404	911	?	Raoult et al. (2004)
PBCV-1	Phycodnaviridae	330,744	373	~50	Li et al. (1997)
MsEPV	Poxviridae	230,120	267	~100	Afonso et al. (1999)
MCV	Poxviridae	190,289	182	~100	Senkevich et al. (1996)
ASFV	Asfarviridae	170,101	151	>50	Yanez et al. (1995)
HSV-2	Herpesviridae	154,746	473	~30	Dolan et al. (1998)
AcNPV	Baculoviridae	133,894	154	?	Ayres et al. (1994)
LCDV	Iridoviridae	102,653	110	~40	Tidona and Darai (1997)

[a] PBCV-1: *Paramecium bursaria chlorella* virus 1; MsEPV: *Melanoplus sanguinipes* entomopoxvirus; MCV: *Molluscum contagiosum* virus; ASFV; African swine fever virus; HSV-2: *Herpes simplex virus type 2*; AcNPV: *Autographa californica* nuclear polyhedrosis virus; LCDV: lymphocystis disease virus. Data from Van Etten (2003).

viral genome; and four unique orphan genes. These results suggest that immediate early gene transcripts, that could be translated without viral gene expression, are necessary to *Mimivirus* to perform the first steps of its replicative cycle. Among other dsDNA viruses, viral transcripts were found in *Human Cytomegalovirus* and *Herpes simplex virus type-1* particles (Bresnahan and Shenk, 2000; Greijer et al., 2000; Sciortino et al., 2001), two members of the *Herpesviridae* family.

The protein composition of *Mimivirus* particles is currently under investigation by proteomic analysis. When compared to other large DNA viruses (Table 2), it should be expected a large number of proteins associated to the virion. Characterization of either the structural proteins, among which is the L425 Capsid protein, or the "accessory" proteins would provide helpful informations in our understanding of the viral replication cycle, especially in the interactions between different viral proteins and cellular host factors.

2.2. Replication cycle

Observation of *Mimivirus*-infected *A. polyphaga* revealed a typical viral development cycle (La Scola et al., 2003 and submitted manuscript). There is first an eclipse phase until 4 h post-infection (p.i.) (Fig. 3A), followed by the clustered cytoplasmic accumulation of newly synthesised virions at 8 h p.i. (Fig. 3B). At later time p.i., almost all the cytoplasmic space is filled up with viral particles (Figs. 3C and 4A). *Mimivirus* is a lytic virus since most of the infected amoebae are lysed at 24 h post-infection (not shown). Although not clearly established, release of progeny virions might most probably occur through infected cell lysis. Actually, little is known about the different stages of this replication cycle namely attachment to the cell surface and entry, viral core release, DNA replication, transcription, translation, assembly and release of progeny virions. However, electron microscopy analyses of infected cells revealed interesting pictures notably capsid assembly and budding from the nucleus (Fig. 4D and E), as well as the presence of both "empty" (Fig. 4B and C) and "full" particles in the cytoplasm. These features are reminiscent of the "virus factories" that have been described for many different viruses but especially for all the NCLDVs members *Poxviridae, Iridoviridae, Phycodnaviridae* and *Asfarviridae*, but also for *Herpesviridae* (Novoa et al., 2005; Van Etten et al., 2002). Viral factories are functional dynamic structures that modify large areas of infected cells to allow efficient viral morphogenesis. The early DNA replication of *Herpes viruses* begins in nuclear factories where several events take place: DNA synthesis, capsid assembly, DNA encapsidation, assembly of empty shells in which viral DNA is inserted and capsid acquisition of a primary envelope by budding through the nuclear membrane. The late replication steps occur in cytoplasmic viral factories. The existence of both nuclear and cytoplasmic viral factories in *Mimivirus* infected cells is a likely hypothesis. Indeed our results showed that *Mimivirus* capsids were assembled in the nucleus of infected cells and acquired their inner membrane by budding through the nuclear membrane (Fig. 4D and E). Moreover, the electronlucent appearance of the cytoplasmic areas where *Mimivirus* particles accumulated looked like the one described for other complex dsDNA viruses. Although many points are still unresolved, on the

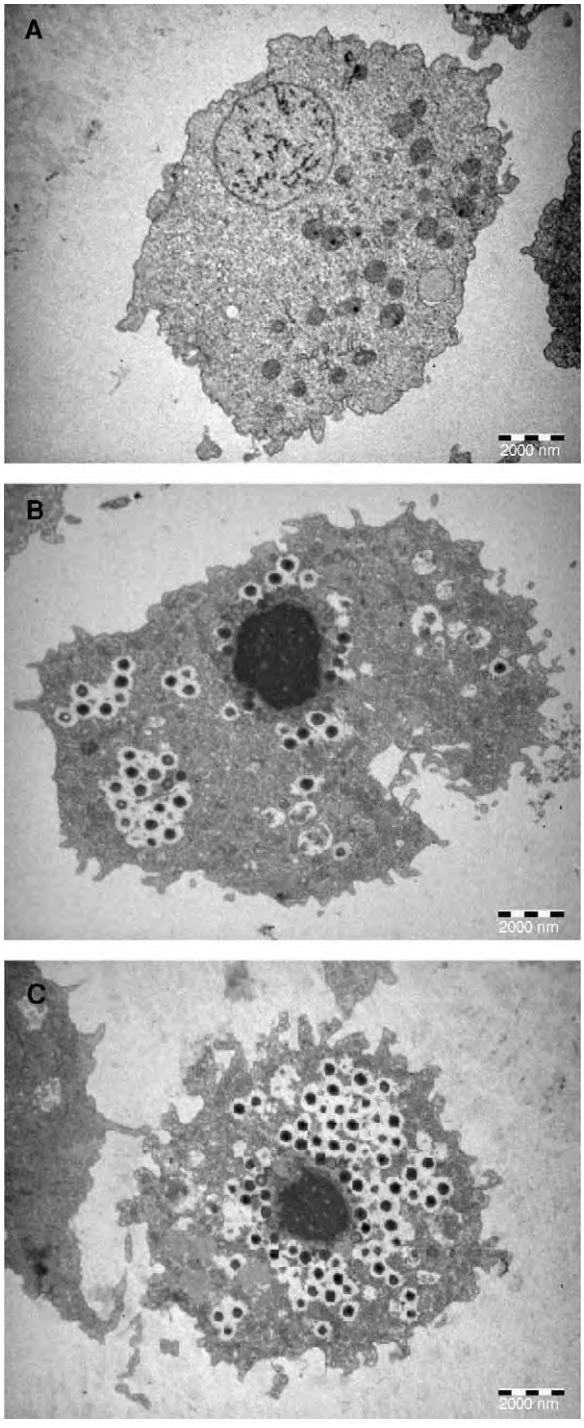

Fig. 3. *Mimivirus* life cycle in *A. polyphaga*. Transmission electron microscopy on *Mimivirus*-infected *A. polyphaga* reveals an eclipse phase at 4 h p.i. (A), followed by perinuclear production and cytoplasmic accumulation of newly formed viruses at 8 h p.i. (B), increasing at 16 h p.i. (C). Bar = 2 μm.

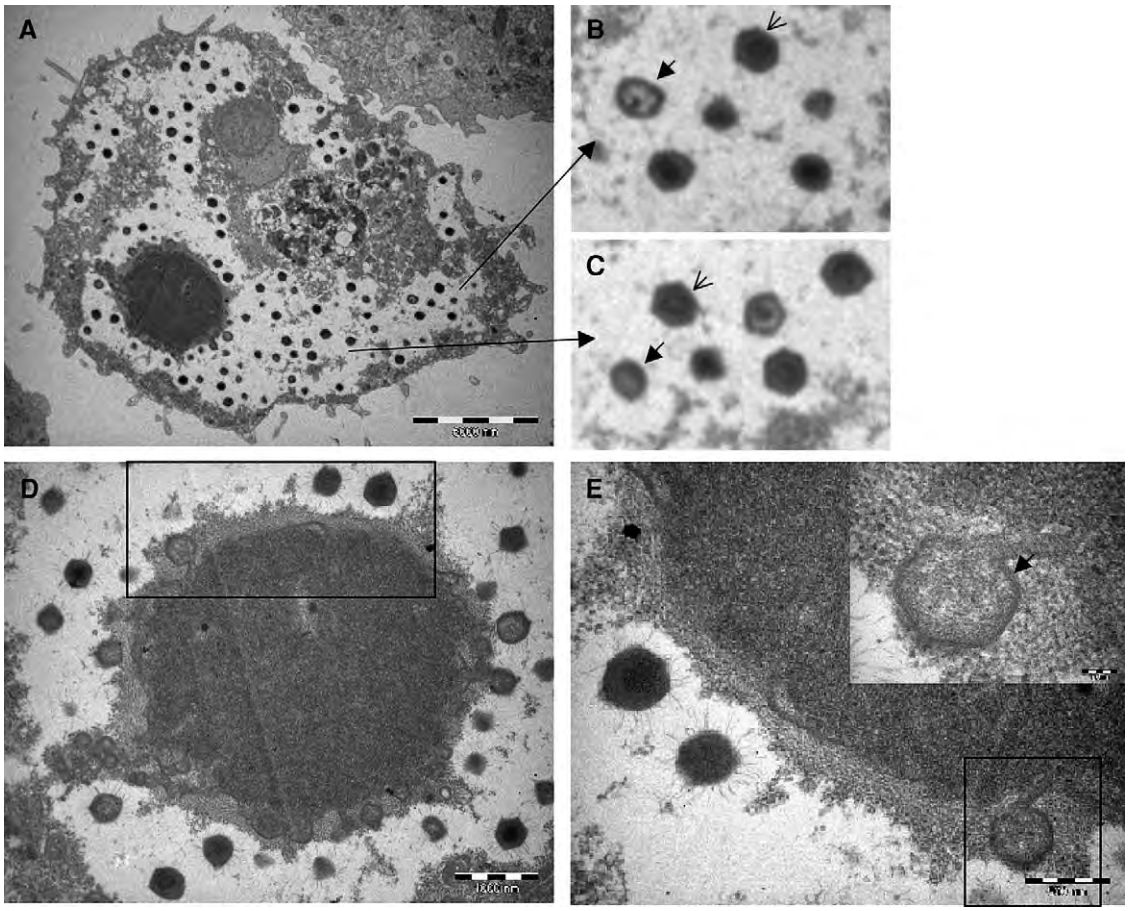

Fig. 4. Assembly of *Mimivirus* particles. At 16 h p.i. the cytoplasm of *Mimivirus* infected *A. polyphaga* is filled up with virions (A). Bar = 5 μm. Two cytoplasmic areas are enlarged in (B and C) where "empty" (closed arrow) and "full" (opened arrow) particles could be seen. Details of the nucleus of this infected cell are shown in (D), showing the release of newly formed capsids at the nuclear periphery. Bar = 1 μm. The boxed area is enlarged in (E). *Mimivirus* assembly starts in the nucleus of infected cells and a capsid budding from the nucleus could be seen. Bar = 500 nm. The boxed area is enlarged in the upper right. Bar = 100 nm.

basis of these initial observations, and according to what was proposed by Van Etten and Meints (1999) for the *PBCV-1 Phycodnavirus* life cycle, we would like to tentatively propose a scheme for the replication cycle of *Mimivirus* (Fig. 5).

2.3. Host cells

As any other virus, *Mimivirus* is an obligate intracellular parasite which requires cellular metabolism and functions to multiply. Initially isolated and then propagated on cultures of amoebae, we tried to determine the cellular tropism of *Mimivirus*. A large number of primary or established cell lines from vertebrates or invertebrates were checked for their ability to support *Mimivirus* infection and replication. It appears that only A. *polyphaga, A. castellanii* and *A. mauritaniensis* could be productively infected by a cell-free viral suspension (La Scola et al., submitted manuscript). This result suggested either a very narrow range of target cell specificity, restricted to protozoans, or the need for a cell to cell transmission of infection from amoebae infected cells to higher eukaryotic cells.

3. Genome analysis

3.1. General features

The genome of *Mimivirus* is a linear dsDNA of 1.2 Mb long which was entirely sequenced (GenBank accession number AY653733). The viral genome might adopt a circular topology upon annealing of the 900 nt inverted repeats found at each end (Raoult et al., 2004). This genome is larger than the genome of several bacteria and archeon parasites (Koonin, 2005) and is the largest of the dsDNA viral genomes described (Table 2). Apart from this exceptional large size, the *Mimivirus* genome is mainly characterized by its high coding capacity of 90.5%, since 1262 putative Open Reading Frames (ORF) were described. Of the 911 coded proteins, 298 had a functional attribute. This represents about 24% of the predicted genes, which is a rather low number when compared to the 70% of the predicted genes of small bacteria and archeal genomes which are of known function (Galperin and Koonin, 2004), but which states the question about the origin and the significance of the very high proportion of

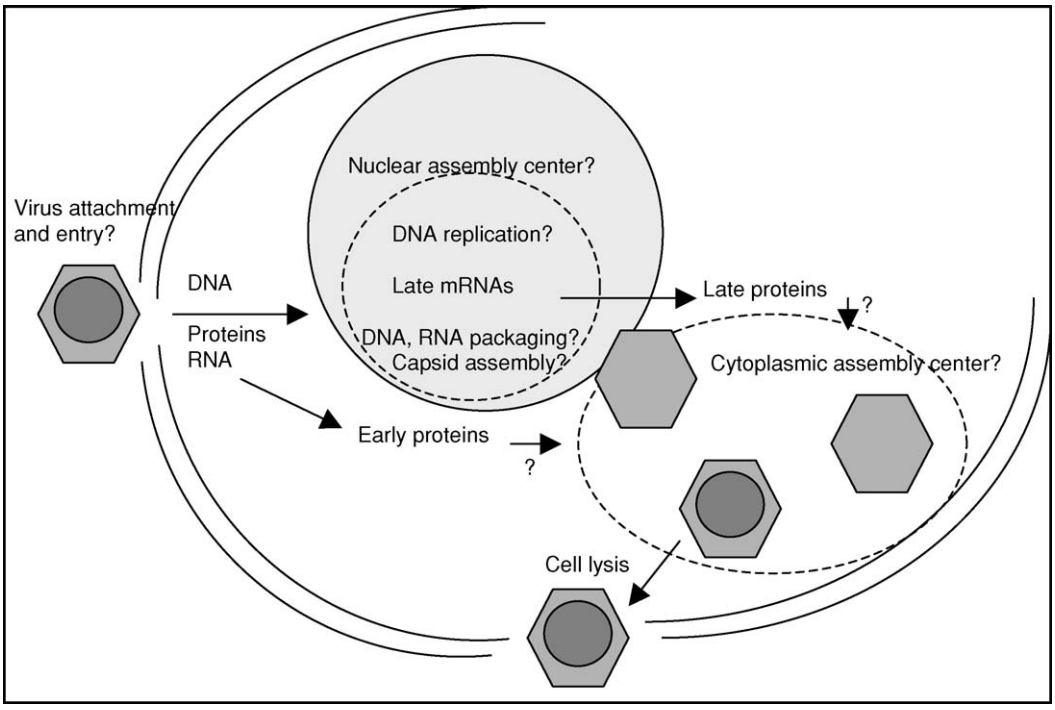

Fig. 5. Proposed life cycle for *Mimivirus*. Following entry into the amoebae, *Mimivirus* DNA might be decapsidated and transported to the nucleus, and immediate early RNAs delivered to the cytoplasm to be translated. DNA replication might begin within the nucleus, possibly in a nuclear assembly center. Early and late genes are transcribed and most probably transported to the cytoplasm to be translated. Capsid assembly might take place within the nucleus as well as DNA encapsidation. Coating and fibrils acquisition of the viral particles might occur in the cytoplasm, possibly in a cytoplasmic assembly center. Newly synthesised *Mimivirus* particles might be released by lysis of the infected cells.

genes with unassigned functions in *Mimivirus* genome. Of the predicted *Mimivirus* ORFs, 194 matched significantly with 108 clusters of orthologs groups (COG) representing 17 different functional classes (Table 3). Functional categories of amino acid transport and metabolism (E), translation (J) and post-translational modification (O) were overrepresented compared to other NCLDVs genomes (Raoult et al., 2004). Detailed analysis of the *Mimivirus* genome revealed features shared with other viral, bacterial or cellular genomes as well as features unique to this virus, that after phylogenetic studies allowed us to assign it a unique position in the tree of species from the three domains of life.

Table 3
Clusters of orthologs groups identified in *Mimivirus*

COG[a]	Category	Function	*Mimivirus* genes
2	A	RNA processing and modification	2
1	D	Cell cycle control and mitosis	1
3	E	Amino acid transport and metabolism	8
5	F	Nucleotide transport and metabolism	5
1	G	Carbohydrate metabolism and transport	1
3	I	Lipid metabolism	3
7	J	Translation	7
11	K	Transcription	11
22	L	Replication, recombination and repair	31
7	M	Cell wall/membrane biogenesis	8
1	N	Cell motility	4
12	O	Post-translational modification, protein turnover and chaperones	18
2	Q	Secondary metabolites biosynthesis, transport and catabolism	4
15	R	General function prediction only	75
12	S	Function unknown	12
3	T	Signal transduction	3
1	U	Intracellular trafficking and secretion	1

[a] Number of clusters of orthologs groups (COG) in each category with one or more homolog in *Mimivirus*.

3.2. Genome analysis

3.2.1. Genes shared with all or some NCLDVs

Comparative analysis of the genomes of the 4 families of NCLDVs allowed Iyer et al. (2001) to propose an hypothetical common ancestral origin to these viruses based on the presence of a set of 31 core genes among which 9 class I genes conserved in all families, 8 class II genes conserved in all families but missing in at least one lineage within families and 14 class III genes conserved in 3 out of 4 families. The genome of *Mimivirus* comprises the 9 class I genes, 6 out of 8 class II genes and 11 out of 14 class III genes with closest homologs scattered among the 4 different families of NCLDVs: 11 close to *Phycodnaviruses*, 7 close to *Poxviruses*, 6 close to *Iridoviruses* and 1 close to *Asfarviruses* (Table 4). Interestingly, it was observed that the transcripts of 2 class I genes (DNA polymerase and capsid protein) and of one class II genes (TFII-like transcription factor) were found packaged within the viral particle (Table 1). Of the five core genes missing

Table 4
NCLDVs core genes identified in *Mimivirus*

NCLDV core gene	Putative function	*Mimivirus* gene	Best homolog in NCLDV
I	Helicase III/ATPase	L206	*Iridoviruses*
I	DNA polymerase (B family)	R322	*Phycodnaviruses*
I	ATPase	L437	*Iridoviruses*
I	Helicase II	L396	*Phycodnaviruses*
I	Capsid	L425	*Phycodnaviruses*
I	Thiol oxidoreductase	R596	*Phycodnaviruses*
I	Helicase II	R350	*Poxviruses*
I	Ser/Thr protein kinase	R400	*Phycodnaviruses*
I	Transcription factor	R450	*Phycodnaviruses*
II	TFII-like transcription factor	R339	*Phycodnaviruses*
II	MuT-like NTP pyrophosphatase	L524	*Iridoviruses*
II	Myristoylated virion protein A	L323	*Iridoviruses*
II	PCNA	R493	*Phycodnaviruses*
II	Ribonucleotide reductase large sub.	R313	*Poxviruses*
II	Ribonucleotide reductase small sub.	L312	*Poxviruses*
III	Unknown	R429	*Phycodnaviruses*
III	BroA, KilA-N terminal domain	L37	*Poxviruses*
III	mRNA capping enzyme	R382	*Asfarviruses*
III	RNA polymerase subunit 2	L244	*Iridoviruses*
III	RNA polymerase subunit 1	R501	*Iridoviruses*
III	Thioredoxin/Glutaredoxin	R195	*Phycodnaviruses*
III	Ser/Tyr phosphatase	R622	*Poxviruses*
III	BIR domain	R311	
III	Virion-associated membrane protein	L65	*Poxviruses*
III	Topoisomerase II	R480	*Phycodnaviruses*
III	SW1/SNF2 helicase	L364	*Poxviruses*

in *Mimivirus*, it should be noticed that their protein products are enzymes implied in DNA precursor synthesis (the class II thymidylate kinase and the class II dUTPase), DNA replication (the class III ATP-dependent DNA ligase and RuvC-like Holliday junction resolvase), or transcription (the class III RNA polymerase subunit 10). These results suggested that, although undoubtedly related to NCLDVs, *Mimivirus* had a specific pattern of NCLDVs core genes. Phylogenetic analysis based on the concatenated sequences of the class I genes confirmed this hypothesis and defined *Mimivirus* as a new and independent lineage of NCLDVs, the *Mimiviridae* (Raoult et al., 2004).

Besides these 26 common core genes, the *Mimivirus* genome contains a large amount of ORFs with characteristics specific to this virus. The main group represented about 80% of the ORFs with no known homologs (Raoult et al., 2004). The origin of these orphan genes as well as the function of the protein products are still unknown. However, since the transcripts of some of these orphan genes were detected within the viral particles (Table 1), it might be thought that at least part of these genes might code for proteins useful to the *Mimivirus* replication cycle. Whether some of these orphan proteins are also packaged within the virions is currently under investigation. In the *Mimivirus* genome, another group of genes was identified that were never described in a viral genome until then, and that brought us some information on the origin and evolution of this virus.

3.2.2. Genes homolog to prokaryotes and/or eukaryotes

The detailed analysis of these genes revealed different functional categories that are shown in Table 5, with their inferred origin (Raoult et al., 2004; Koonin, 2005). Strikingly, genes dealing with tRNA modification, translation, molecular chaperones and protein folding, amino acid metabolism, protein modification and lipid metabolism were more closely related to eukaryotic homologs whereas gene products implied in nucleotide synthesis and polysaccharide metabolism were closer to bacterial homologs. Genes coding DNA repair enzymes showed best homologies with either eukaryotic or bacterial genes. Together with the presence of genes of the translation functional category, the presence of DNA repair genes is the most remarkable feature of *Mimivirus* genome. Although tRNA-like and elongation factor genes were also found in dsDNA viruses like *Chlorella Viruses* (Van Etten and Meints, 1999), amino acyl-tRNA synthetases (R663, L124, L164, R639), translation initiation (R464, R496) and elongation (R624) factors, peptide release factor (R726) and a tRNA modifying enzyme (R405) are unique to *Mimivirus*. Moreover the transcripts of Arg-, Tyr- and Cys-tRNA synthetases were packaged within the viral particle (Table 1). The presence of this set of genes in the genome as well as genes coding for molecular chaperones (L254, L393, R260, R266, R445) or proteins involved in protein quality control (L605, L251) suggest they might have an important role during the viral replication cycle to allow efficient viral protein translation and assembly. *Mimivirus* also

Table 5
Mimivirus unique genes of non viral origin

Mimivirus ORF	Putative function	Category	Inferred origin
R663	Arg-tRNA synthetase	Translation tRNA modification	Eukaryotic
L124	Tyr-tRNA synthetase		Eukaryotic
L164	Cys-tRNA synthetase		Eukaryotic
R639	Met-tRNA synthetase		Eukaryotic
R726	Release factor eRF1		Eukaryotic
R624	Elongation factor eF-Tu		Eukaryotic
R464	Initiation factor SUI1		Eukaryotic
L496	Initiation factor 4E		Eukaryotic
R405	tRNA (5-uracil-)methylase		Eukaryotic
L359	Mismatch repair ATPase MutS	DNA repair	Bacterial
R693	Methylated-DNA-protein-methyltransferase		Bacterial
R406	Dioxygenase alkylated DNA repair		Bacterial
L315, L720	Formamidopyrimidine DNA glycosylase		Bacterial
R194, L221	Topoisomerase IB, IA		Bacterial
R480	Topoisomerase IIA		Eukaryotic
L687	Endonuclease UV-damaged DNA repair		Eukaryotic
L469	Polynucleotide kinase-phosphatase		Eukaryotic
L254, L393	HSP70	Protein folding	Eukaryotic
R260, R266, R445	DnaJ-like proteins		Eukaryotic
L605	Peptidyl-prolyl isomerase		Eukaryotic
L251	Lon-like protease		Eukaryotic
R418	Nucleoside diphosphate kinase	Nucleotide synthesis	Bacterial
L716	GMP synthase		Bacterial
R475	Asparagine synthase	Amino acid metabolism	Eukaryotic
R565	Glutamine synthase		Eukaryotic
L230	Lysyl hydroxylase	Protein modification	Eukaryotic
L906	Cholinesterase	Lipid metabolism	Eukaryotic
L808	Lanosterol-14-alpha-demethylase		Eukaryotic
R807	Dehydrocholesterol reductase		Eukaryotic
R689	NAG-1-phosphate uridylyltransferase	Polysaccharide metabolism	Bacterial
L136	Sugar transaminase		Bacterial
L780	Reductase		Bacterial
L612	Mannose-6-phosphate isomerase		Bacterial
L543	ADP-ribosylglycohydrolase		Bacterial

displays an unusual set of genes implied in different mechanisms of DNA repair that were never described before in dsDNA viruses and that are representative enzymes of the different repair pathways of DNA damages due to oxidization (L315, L720), methylation (R693), alkylation (R406), UV radiation (L687) and nucleotide mismatch (L359). More surprising was the first identification in a viral genome of the genes coding for the three topoisomerases IA (L221), IB (R194) and IIA (R480). Although topoisomerases IB and IIA could be found in NCLDVs, topoisomerase IA was never reported in a virus, and the presence of the three enzymes was only described in yeast and some environmental bacteria (Raoult et al., 2004).

In addition to the structural and regulatory viral genes necessary to perform the different steps of the replication cycle, large dsDNA viruses are known to have genes implied in various metabolic pathways such as sugar, lipid or amino acid metabolism. This also appeared to be the case in *Mimivirus* since genes coding for enzymes involved in nucleotide synthesis (R418, L716), amino acid metabolism (R475, R565), lipid metabolism (L906, L808, R807) and polysaccharide metabolism (R689, L136, L780, L612, L543) were identified. Among these enzymes, some are unique for a virus especially those intervening in the Glutamine metabolism (R475, R565, L716) and in the biosynthesis of oligosaccharides (listed above). The presence of these latter enzymes added to the presence of six glycosyltransferases in the *Mimivirus* genome might be linked to two of our observations: (i) the Gram-positive staining of *Mimivirus* might be due to the presence of an external layer of reticulated polysaccharides; (ii) the external fibrils are sensitive to protease and lyzozyme digestion (La Scola et al., submitted manuscript), suggesting that they might be composed of cross-linked glycoproteins.

Another specific feature of *Mimivirus* genome is the presence of an Intein sequence within the DNA polymerase type B gene (R322) (Raoult et al., 2004; Ogata et al., 2005a). Inteins are mobile selfish DNA elements, with no known function, present in protein coding sequences that remove themselves by an autoproteolytic process following translation (Pietrokovski, 2001). They have been described in archaea,

bacteria and eukaryotes, but only once in an *Iridovirus*, the *Chilo Iridescent virus* (Pietrokovski, 1998). Detailed analysis of *Mimivirus* intein demonstrated canonical sequence motifs and the presence of a homing endonuclease domain, related to archeal inteins but belonging to a distinct subclass (Ogata et al., 2005a). Beside the presence of the Intein sequence, the *Mimivirus* DNA polymerase type B gene contains an exonuclease site and a polymerase site with conserved residues suggesting it is likely to be functional (Ogata et al., 2005a).

3.2.3. Expanded gene families

Throughout the *Mimivirus* genome analysis, it appeared that some genes might be members of families sharing conserved motifs which are a structural or a functional signature. The largest gene family is composed of 66 proteins expressing the 33-residue ankyrin-repeat. This motif, involved in protein-protein interaction, has been found in many proteins spanning a wide range of functions in the three domains of life, archaea, bacteria and eukaryotes, but also in a great number of viral genomes (Moavi et al., 2004). As a comparison, in the *PBCV-1* genome the ankyrin domain is shared in a six member family of genes (Van Etten, 2003). The second largest family is formed by more than 20 proteins with the BTB/POZ signature. This is an N-terminal motif found associated with other conserved domains such as KELCH repeats or Zinc fingers domains, implied in a wide range of homophilic or heterophilic protein–protein interactions (Aravind and Koonin, 1999). Besides these large families, four other might be of importance in the *Mimivirus* life cycle: one 14 member family of proteins with protein kinase motif, and three different 8 member families containing either the collagen triple-helix motif or the glucose–methanol–choline (GMC) oxidoreductase motif or similar to helicases. All these results stressed the level of complexity of the *Mimivirus* genome compared to those described until now and it should be guessed that it will increase further with our knowledge of orphan genes and of other family genes of uncharacterized paralogs (Koonin, 2005).

4. Origin and evolution of *Mimivirus*

4.1. Is Mimivirus a cellular organism or a virus?

By its size of about 650 nm, *Mimivirus* is not an ultrafiltrable agent as commonly defined for most viruses, but is rather comparable to the size of small intracellular bacteria such as *Rickettsia conorii* or *Tropheryma whipplei* (Raoult, 2005). The 1.2 Mb size of its genome is comparable to the genomes of these bacteria, 1.27 Mb for *R. conorii* (Ogata et al., 2001), and 0.9 Mb for *T. whipplei* (Bentley et al., 2003). Strand asymmetry for gene transcription was present in *Mimivirus* as in bacteria, which is generally associated to the origin of replication. This feature is common in bacteria but was only described in *β-herpesviruses* (Mrazek and Karlin, 1998). However, in *Mimivirus,* transcription of genes appeared to be similar from both strands.

Comparison of *Mimivirus* COGs numbers with genomes of similar size showed that they are fewer in *Mimivirus* than in different bacteria or small eukaryotes, but not within specific functional categories (Raoult et al., 2004). In small obligate intracellular bacteria, genes coding for nucleotide or amino acid synthesis are lacking. This makes these organisms dependent on their host cell products for their metabolism. These genes are present in *Mimivirus* genome as well as genes involved in other metabolic pathways. The major difference in *Mimivirus* genome with the genomes of small intracellular bacteria is the absence of genes coding for ribosomal proteins. Genes implied in energy production and conversion are under-represented in *Mimivirus* genome. Then, despite the presence of an unusually high number of genes involved in transcription and translation, *Mimivirus* is dependent on its host for at least the protein translation apparatus and energy metabolism.

Although the different steps of its replicative cycle are yet to be uncovered, the detailed genome analysis provided us useful informations on the complex apparatuses that might be involved in DNA replication and DNA repair; transcription which is independent of the host unlike other large DNA viruses like *PBCV-1* (Van Etten, 2003); but also in different pathways used for protein translation and folding, virion morphogenesis and intracellular transport. *Mimivirus* life cycle is strictly intracellular and it shares physical and functional characteristics with some small obligate intracellular bacteria. All this suggests a very complex life cycle but which remains, at least until now, rather a viral cycle than a bacterial one, since *Mimivirus* does not undergo division but replicate through the assembly of pre-formed subunits.

4.2. Phylogeny and putative origin of Mimivirus

4.2.1. Linkage to NCLDVs and to a common ancestor

Mimivirus genome contains 26 of the 31 core genes (classes I–III) defined by Iyer et al. (2001), as belonging to a common ancestral virus. *Mimivirus* is phylogenetically related to NCLDVs as demonstrated by phylogenetic trees built on the ribonucleotide reductase small and large subunits and topoisomerase II sequences (La Scola et al., 2003) as well as on the concatenated sequences of the conserved eight class I core genes (Fig. 6A; Raoult et al., 2004). However, it appears to be different enough from any other of the NCLDVs families to constitute its own clade in the NCLDV tree. One hypothesis would be that originally *Mimivirus* contained the complete set of 61 core genes (classes I–IV), some of which having been lost during evolution and/or adaptation to the host. These results suggested the existence of a common ancestor originating before the emergence of the three domains of life and questioned the involvement of this large DNA ancestral virus in the emergence of eukaryotes. This hypothesis is a matter of debate since Koonin (2005) proposed that NCLDVs are not older than eukaryotes but that the common ancestral virus existed early in the evolution of eukaryotes before the radiation of the main eukaryotic lineages. An argument against

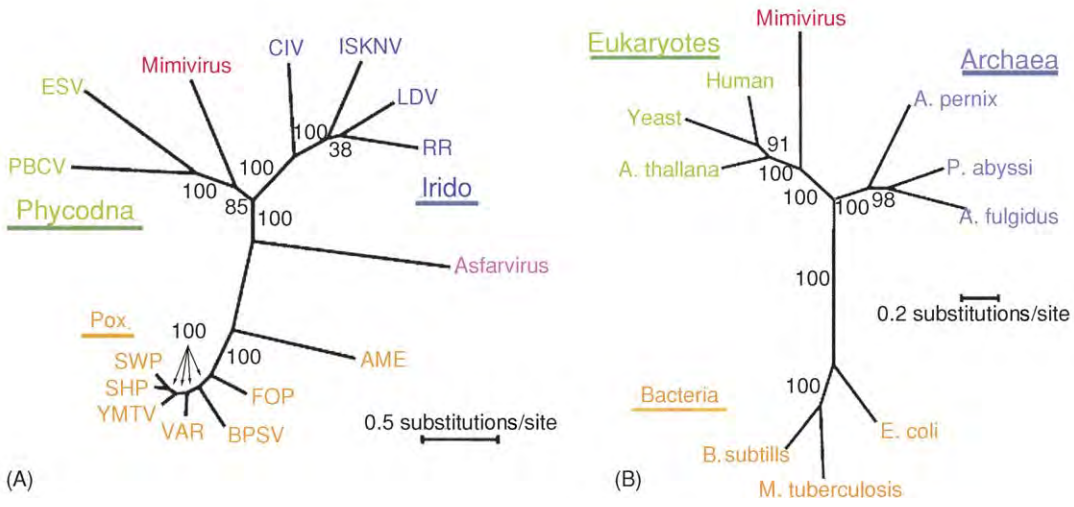

Fig. 6. Phylogeny of *Mimivirus*. Phylogenetic position of *Mimivirus* among the NCLDV families (A) and in the tree of species from the three domains of life (B). Both figures are from Raoult et al. (2004).

this theory and in favour of a very ancient common ancestor might come from the common architecture of the coat protein shared by viruses infecting different hosts in the three domains of life (Benson et al., 2004). Using structure-based modelling these authors showed that these viruses, with no sequence similarity, have "double-barrel trimer" coat proteins. Preliminary results seemed to indicate that this would also be the case for the *Mimivirus* capsid protein (unpublished results).

4.2.2. *Putative origin and evolution of Mimivirus*

Our current hypothesis is that *Mimivirus* and other large DNA viruses are of deep evolutionary origin close to the origin of the other domains of life. But phylogenetic analyses lead to alternative view points:

(i) *Mimivirus* as a "gene pickpocket" from its hosts? Phylogenetic trees based on a large sampling of selected protein sequences located *Mimivirus* within the Eukaryotes domain (Moreira and Lopez-Garcia, 2005). Another paper is in favour of this hypothesis based on the conservation of the core set genes of NCLDVS which was not described outside of eukaryotic viruses, the eukaryotic and bacterial evolutionary affinities of *Mimivirus* non viral genes and the presence of many paralogous families, suggesting that *Mimivirus* genome grew by gene duplication (Koonin, 2005).

(ii) *Mimivirus* as a source of new genes for its hosts? All the known unicellular genomes from the three domains of life shared 63 COGs, of which 7 were described in the *Mimivirus* genome: 3 amino acyl-tRNA synthetases, β and β′ subunits of RNA polymerase, the sliding clamp subunit of DNA polymerase and a 5′–3′ exonuclease. A phylogenetic tree built on the concatenated sequences of these seven genes placed *Mimivirus* in a branch near the origin of Eukaryotes (Fig. 6B; Raoult et al., 2004). The tree topology was invariant after several methodological changes. Presence of an archeal intein was described in the eukaryotic-related DNA polymerase ORF (Ogata et al., 2005a). Most of the *Mimivirus* ORFs have higher sequence similarities to eukaryotic genes. We recently developed new arguments opposed to the "gene pickpocket" hypothesis based on complete genome comparison between *Entamoeba hystolytica,* representative of *Mimivirus* hosts, and *Mimivirus*. This analysis revealed 87 reciprocal best matching ORFs between the two genomes and the distribution of similarity scores demonstrated that only five genes were more similar to *Entamoeba histolytica* than to other eukaryotic kingdoms. Furthermore, only 8.3% of *Mimivirus* ORFs, homologs to other organisms, are likely to have been recently acquired. These results suggested that horizontal gene transfer between the host and *Mimivirus* was negligible (Ogata et al., 2005b).

The discovery of *Mimivirus* forced back the frontiers between viruses and cellular organisms since the only differences between *Mimivirus* and small obligate intracellular bacteria, at least until now, are the absence of ribosomal proteins and of proteins involved in energy metabolism and multiplication by assembly of pre-formed subunits. Extensive studies are necessary to uncover the totality of *Mimivirus* genome and life cycle and to increase our knowledge of this new field in virology. Confirming the hypothesis of the deep evolutionary origin of *Mimivirus* and other large dsDNA viruses would need isolation and description of either other members of the *Mimiviridae* family from amoebae, as suggested by Ghedin and Fraser (2004) or of new large dsDNA viruses families.

Acknowledgements

Lina Barassi, Bernard Campagna, Christophe Fraisier, Amélie Honstettre, and Catherine Robert are gratefully acknowledge for their expert technical assistance.

References

Afonso, C.L., Tulman, E.R., Lu, Z., Oma, E., Kutish, G.F., Rock, D.L., 1999. The genome of *Melanoplus sanguinipes* entomopoxvirus. J. Virol. 73, 533–552.

Aravind, L., Koonin, E.V., 1999. Fold prediction and evolutionary analysis of the POZ domain: structural and evolutionary relationship with the potassium channel tetramerization domain. J. Mol. Biol. 285, 1353–1361.

Ayres, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M., Possee, R.D., 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virology 202, 586–605.

Benson, S.D., Bamford, J.K., Bamford, D.H., Burnett, R.M., 2004. Does common architecture reveal a viral lineage spanning all three domains of life? Mol. Cell 16, 673–685.

Bentley, S.D., Maiwald, M., Murphy, L.D., Pallen, M.J., Yeats, C.A., Dover, L.G., Norbertczak, H.T., Besra, G.S., Quail, M.A., Harris, D.E., von Herbay, A., Goble, A., Rutter, S., Squares, R., Squares, S., Barrell, B.G., Parkhill, J., Relman, D.A., 2003. Sequencing and analysis of the genome of the Whipple's disease bacterium *Tropheryma whipplei*. Lancet 361, 637–644.

Bresnahan, W.A., Shenk, T., 2000. A subset of viral transcripts packaged within human cytomegalovirus particles. Science 288, 2373–2376.

Brookes, S.M., Dixon, L.K., Parkhouse, R.M.E., 1996. Assembly of African swine fever virus: quantitative structural analysis in vitro and in vivo. Virology 224, 84–92.

Dolan, A., Jamieson, F.E., Cunningham, C., Barnett, B.C., McGeoch, D.J., 1998. The genomes sequence of herpes simplex type 2. J. Virol. 72, 2010–2021.

Galperin, M.Y., Koonin, E.V., 2004. Conserved hypothetical" proteins: prioritization of targets for experimental study. Nucleic Acids Res. 32, 5452–5463.

Ghedin, E., Fraser, C., 2004. A virus with big ambitions. Trends Microbiol. 13, 56–57.

Greijer, A.E., Dekkers, C.A.J., Middeldorp, J.M., 2000. Human cytomegalovirus virions differentially incorporate viral and host cell RNA during the assembly process. J. Virol. 74, 9078–9082.

Iyer, L.M., Aravind, L., Koonin, E.V., 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. 75, 11720–11734.

Koonin, E.V., 2005. Virology: Gulliver among the Lilliputians. Curr. Biol. 15, R168–R169.

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.-M., Raoult, D., 2003. A giant virus in amoebae. Science 299, 2033.

La Scola, B., Marrie, T.J., Auffray, J.P., Raoult, D., 2005. *Mimivirus* in pneumonia patients. Emerg. Inf. Dis. 11, 449–452.

Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G.F., Rock, D.L., Van Etten, J.L., 1997. Analysis of 74 kb of DNA located at the right end of the 330 kb chlorella virus PBCV-1 genome. Virology 237, 360–377.

Moavi, L.K., Cammett, T.J., Desrosiers, D.C., Peng, Z.-Y., 2004. The ankyrin repeat as molecular architecture for protein recognition. Protein Sci. 13, 1435–1448.

Moreira, D., Lopez-Garcia, P., 2005. Comment on «The 1.2-Megabase genome sequence of *Mimivirus*». Science 308, 1114a.

Moss, B., 2001. *Poxviridae*: the viruses and their replication. In: Knipe, D.M., Holey, P.M. (Eds.), Fields Virology. Lippincott Williams and Wilkins, Philadelphia, pp. 2849–2883.

Mrazek, J., Karlin, S., 1998. Strand compositional asymmetry in bacterial and large viral genomes. Proc. Natl. Acad. Sci. U.S.A. 95, 3720–3725.

Novoa, R.R., Calderita, G., Arranz, R., Fontana, J., Granzow, H., Risco, C., 2005. Virus factories: associations of cell organelles for viral replication and morphogenesis. Biol. Cell 97, 147–172.

Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M., Raoult, D., 2001. Mechanisms of evolution in *Rickettsia conorii* and *Rickettsia prowazekii*. Science 293, 2093–2098.

Ogata, H., Raoult, D., Claverie, J.-M., 2005a. A new example of viral intein in *Mimivirus*. Virol. J. 2, doi:10.1186/1743-422X-2-8.

Ogata, H., Abergel, C., Raoult, D., Claverie, J.-M., 2005b. Reply to Comment on «The 1.2-Megabase genome sequence of *Mimivirus*». Science 308, 1114.

Pietrokovski, S., 1998. Identification of a virus intein and a possible variation in protein-splicing reaction. Curr. Biol., R634–R635.

Pietrokovski, S., 2001. Intein spread and extinction in evolution. Trends Genet. 17, 465–472.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.-M., 2004. The 1.2-Megabase genome sequence of *Mimivirus*. Science 306, 1344–1350.

Raoult, D., 2005. The journey from Rickettsia to *Mimivirus*. ASM News 71, 278–284.

Salas, M.L., 1999. African swine fever (*Asfarviridae*). In: Granoff, A., Webster, R.G. (Eds.), Encyclopedia of Virology. Academic Press, San Diego/London, pp. 30–31.

Sciortino, M.-T., Suzuki, M., Taddeo, B., Roizman, B., 2001. RNAs extracted from herpes simplex virus 1 virions: apparent selectivity of viral but not cellular packaged in virions. J. Virol. 75, 8105–8116.

Senkevich, T.G., Bugert, J.J., Sisler, J.R., Koonin, E.V., Darai, G., Moss, B., 1996. Genome sequence of human tumorigenic poxvirus: prediction of specific host response–evasion genes. Science 273, 813–816.

Tidona, C.A., Darai, G., 1997. The complete DNA sequence of lymphocystis disease virus. Virology 230, 207–216.

Van Etten, J.L., 2003. Unusual life style of giant chlorella viruses. Annu. Rev. Genet. 37, 153–195.

Van Etten, J.L., Meints, R.H., 1999. Giant viruses infecting algae. Annu. Rev. Microbiol. 53, 447–494.

Van Etten, J.L., Graves, M.V., Müller, D.G., Boland, W., Delaroque, N., 2002. *Phycodnaviridae*—large DNA algal viruses. Arch. Virol. 147, 1479–1516.

Williams, T., 1996. The Iridoviruses. Adv. Virus Res. 46, 347–412.

Yanez, R.J., Rodirguez, J.M., Nogal, M.L., Yuste, L., Enriquez, C., Rodriguez, J.F., Vinuela, E., 1995. Analysis of the complete genome sequence of African swine fever virus. Virology 208, 249–278.