Mimivirus

J.-M. Claverie (✉), C. Abergel, H. Ogata

Contents

Introduction 90
The Serendipitous Discovery of Mimivirus 90
Potential Source of Other *Mimiviridae* 91
Host Range and Pathogenicity 92
Genomics of Mimivirus 95
Overall Genome Structure 95
Mimivirus as a Bona Fide NCLDV 98
Mimivirus as a Unique Giant Virus: Virally Encoded Translation Components 100
Other Remarkable Features of the Mimivirus Genome 101
The Mimivirus Particle 104
Morphology and Ultrastructure 104
Proteomics of the Virion 111
Experimentally Validated Mimivirus Genes 112
Mimivirus-Encoded Components of the Translation Machinery 113
Mimivirus-Encoded Nucleotide Metabolism and DNA Replication Enzymes 113
Characterization of Other Mimivirus Gene Products 114
Mimivirus in the Tree of Life 115
Phylogenetic Analysis Using Components of the Translational Apparatus 115
Phylogenetic Analysis Using Clamp Loaders 116
Evidence of Lateral Gene Transfer Between Mimivirus and Corals 117

Abstract *Acanthamoeba polyphaga* Mimivirus, the first representative and prototype member of the *Mimiviridae*, is the latest addition to the menagerie of lesser-known big DNA viruses. Due to the size of its particle—a fiber-covered icosahedral protein capsid with a diameter of 0.7 μm—Mimivirus was initially mistaken for an intracellular parasitic bacteria. Its 1.2-Mb genome sequence was then found to encode more than 900 proteins, many of them associated with functions never before encountered in a virus, such as four aminoacyl-tRNA synthetases. The finding of Mimivirus-encoded central components of the protein translation apparatus thought to be the signature of cellular organisms revived the debate about the origin

J.-M. Claverie
Structural and Genomic Information Laboratory, Parc Scientifique de Luminy,
Case 934 13288 Marseille cedex 09, France
Jean-Michel.Claverie@univmed.fr

James L. Van Etten (ed.) *Lesser Known Large dsDNA Viruses.*
Current Topics in Microbiology and Immunology 328.

of DNA viruses and their possible role in the emergence of the eukaryotic cell. Despite the many features making it unique in the viral world, Mimivirus is nevertheless phylogenetically close to other large DNA viruses, such as phycodnaviruses and iridoviruses, and most likely share a common ancestry with all nucleocytoplasmic large DNA viruses. Postgenomic studies have now started in various laboratories, slowly shedding some light on the physiology of the largest and most complex virus isolated to date. This chapter summarizes our present knowledge on Mimivirus.

Introduction

The discovery of Mimivirus (for mimicking microbe virus) (La Scola et al. 2003), a double-stranded DNA virus infecting the common ameba *Acanthamoeba polyphaga*, followed by the sequencing and analysis of its genome (Raoult et al. 2004) sent a shock wave through the community of virologists and evolutionists. By its record particle size (750 nm in diameter) and genome length (1.2 million bp), the complexity of its gene repertoire (911 protein coding genes) as well as its particle composition (it contains products of more than 130 virus genes), Mimivirus blurred the established boundaries between viruses and parasitic cellular organisms. Beyond these quantitative aspects, Mimivirus has many types of genes never before encountered in a virus, most noticeably genes encoding central components of the protein translation machinery, previously thought to be the signature of cellular organisms. These exceptional genes include those encoding four aminoacyl-tRNA synthetases. The revolutionary finding of a partial protein translation apparatus in a virus, together with the presence of components of other pathways previously unique to cellular organisms, came at the right time to lend support to several bold theories linking ancestral viruses to the emergence of the eukaryotic domain (Claverie 2006). As more researchers are becoming involved in the study of Mimivirus, experimental information is now slowly accumulating, although very little is known about its physiology. This article reviews some of the recent progress, mostly including individual protein characterization, electron microscopy, and proteomics.

The Serendipitous Discovery of Mimivirus

In 1992, a pneumonia outbreak in the West Yorkshire mill town of Bradford (England), triggered an investigation for *Legionella* (a pneumonia causing intracellular parasitic bacterium) in the water of a nearby cooling tower. This investigation was conducted by Timothy Rowbotham, the officer in charge of Britain's Public Health Laboratory Service. Instead of the expected Gram-negative bacillus-like *Legionella*, he discovered a microorganism resembling a small Gram-positive coccus (initially called Bradfordcoccus) (Fig. 1).

After unsuccessful cultivation attempts and the failure of molecular identification using universal 16*S* rDNA bacterial primers, the mysterious sample was stored

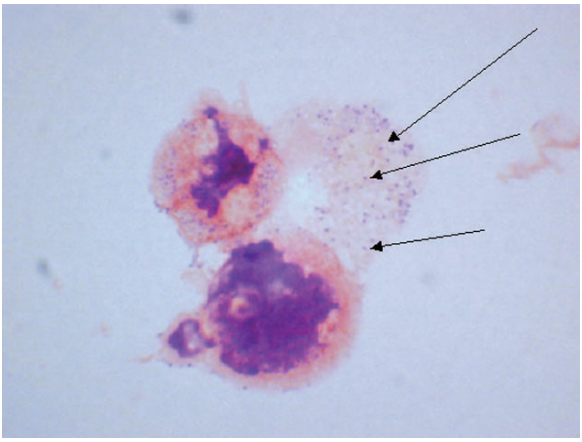

Fig. 1 Light microscopy appearance of Mimivirus particles in infected amoeba, following Gram coloration. The three *arrows* point to individual particles (*purple-blue*) within the ameba cytoplasm (*pink*). From La Scola et al. (2003)

in a freezer for approximately 10 years. It was then brought to the Rickettsia Unit at the school of Medicine in Marseille, France by Dr. Richard Birtle. There, following additional characterization attempts, electron microscopy of infected *Acanthamoeba polyphaga* cells provided the first hint that Bradfordcoccus was in fact a giant virus, with mature icosahedral particles approximately 0.7 μm in diameter, a size comparable to that of mycoplasma cells (La Scola et al. 2003). The viral nature of the agent was further established by the demonstration of an eclipse phase during its replication, and the analysis of several gene sequences exhibiting a clear phylogenetic affinity with nucleocytoplasmic large DNA viruses (NCLDV), a group of viruses including the *Poxviridae*, the *Iridoviridae*, the *Phycodnaviridae*, and the *Asfarviridae*. This new virus was named *A. polyphaga* Mimivirus and is now classified by ICTV as the first and prototype member species of the *Mimiviridae*, a new family within the NCLDV. The size of its particles makes Mimivirus the largest virus ever described. Mimivirus does not pass through a 0.3-μm pore filter, a usual experimental procedure to separate bacterial cells from viruses. Following such filtering steps, prevalent in environmental microbiology studies (for instance, metagenomics), Mimivirus is retained in the pool of prokaryotic cellular organisms (Ghedin and Claverie 2005).

Potential Source of Other Mimiviridae

Phylogenetic analyses of the most conserved genes common to all NCLDVs (the so-called NCLDV core genes, Iyer et al. 2001) consistently places Mimivirus in an independent lineage, between the *Phycodnaviridae* (algal viruses) and *Iridoviridae*

(predominantly fish viruses). This resemblance of Mimivirus with viruses found in aqueous environments prompted Ghedin and Claverie (2005) to search for evidence of other *Mimiviridae* in the environmental microbial DNA sequences gathered in the Sargasso Sea (Venter et al. 2004). This in silico search was successful since 15% of the Mimivirus 911-predicted protein sequences had their closest homologs in this metagenomic data set rather than in viral sequences of known origin. Furthermore, 43% of Mimivirus core genes had their closest homologs in the Sargasso Sea data set. It is thus very likely that other species of *Mimividae* remain to be isolated from the marine environment, probably infecting microalgae (Monier et al. 2008b) or heterotrophic protozoans. Interestingly, the Sargasso Sea data set where Mimivirus sequence homologs are detected correspond to bacteria-sized organisms that passed through 3-μm pore-sized filters and were retained by 0.2-μm pore-sized filters. Mimivirus-like particles (0.75 μm in diameter) are in this range. Our preliminary analysis of the latest oceanic metagenomic sequence data brought back by the *Sorcerer II* Global Ocean Sampling Expedition (Rusch et al. 2007) confirmed the presence of Mimivirus relatives at various oceanic locations around the globe (Monier et al. 2008a). These results thus predict that the marine environment contains Mimivirus relatives that are abundant enough to be randomly sampled from sea water. Thus it is only a matter of time before new *Mimiviridae* members are found in aqueous environments.

Host Range and Pathogenicity

Only cells from species of the *Acanthamoeba* genus have been productively infected by a cell-free viral suspension, among a large number of primary or established cell lines from vertebrates or invertebrates that were tested for their ability to support Mimivirus infection and replication (Suzan-Monti et al. 2006).

Upon infection of *A. polyphaga* cells, Mimivirus has a typical viral replication cycle with an eclipse phase until 5 h postinfection (p.i.), followed by the steady appearance of newly synthesized virions in the cytoplasm, leading to the clustered accumulation of viral particles filling up most of the intracellular space (Fig. 2A), until infected amebae start to lyse after 14 h p.i. The burst size is larger than 300 particles per cell. In a recent study, Suzan-Monti et al. (2007) described the assembly of the virion within and around very large cytoplasmic virus factories. Given that Mimivirus particles contain a rather complete transcription system (see Sect. 4.2) as do poxviruses, the entire replication cycle might occur in the cytoplasm. The presence of a highly conserved promoter motif in 50% of the Mimivirus genes (see Sect. 3.4) suggests that there are two categories of genes, some with promoters recognized by the viral RNA polymerase and some lacking the conserved promoter element that may be transcribed by the host RNA polymerase recruited from the ameba nucleus.

The mechanism of delivery of the particle content into the ameba cytoplasm also remains to be clarified, but seems to require the partial digestion of the sturdy fibril layer surrounding the mature particles (Figs. 2B and 3).

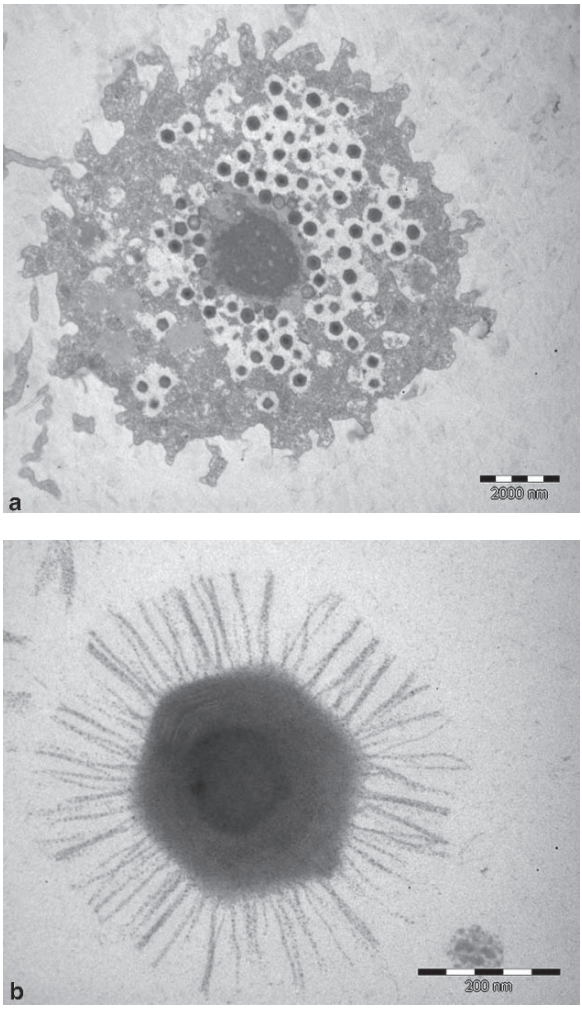

Fig. 2 Transmission electron microscopy of Mimivirus particles. **a** Mimivirus-infected *A. polyphaga* at 8 h p.i. shows intracytoplasmic accumulation of virus particles (Bar = 2 μ m). The *central dark nucleus-like region* is a cytoplasmic virus factory. **b** Mimivirus particles, purified from the supernatant of infected cells, appear as nonenveloped icosahedral virions surrounded by fibrils (Bar = 100 nm) (From Raoult et al. 2007)

The combination of genomic, proteomic, and ultrastructural analyses suggests the following infection scenario:

1. Free virus particles mimicking bacteria (by their size and perhaps a lipopolysaccharide [LPS]-like layer surrounding the capsid) are taken up as food by the ameba.

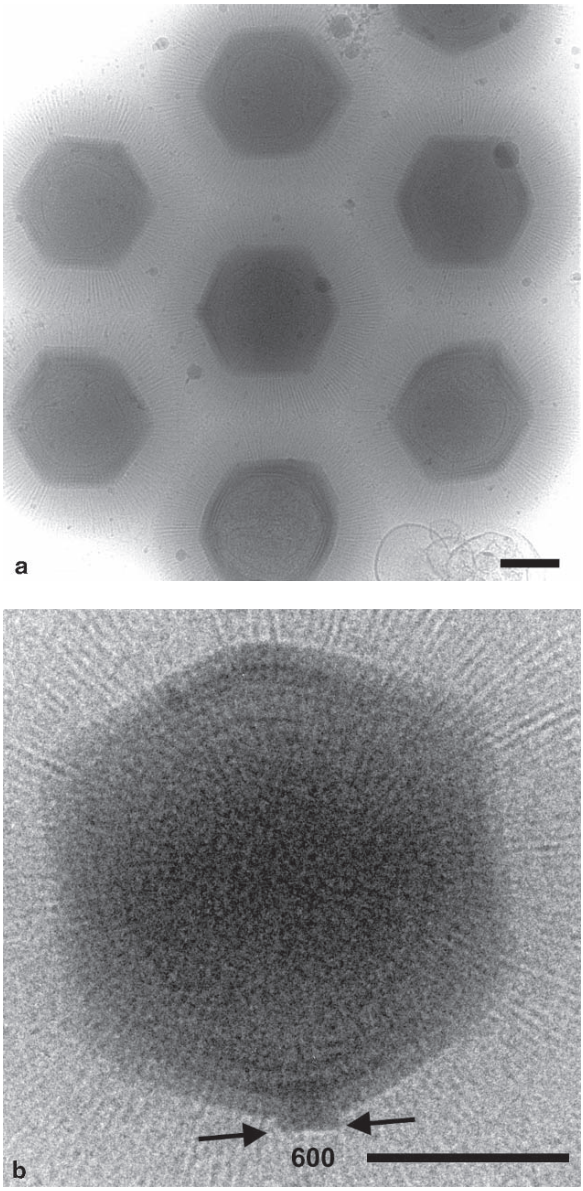

Fig. 3 Cryo-EM highquality images of Mimivirus particles. A Cluster of mature particles, exhibiting a solid and compact fiber layers. **b** Close-up of one particle (0.75 μm across) exhibiting a densely packed layer of crosslinked fibers and a single vertex. (From Xiao et al. 2005). Note the difference in fibril density with Fig. 2B

2. The LPS-like fibril layer is partially digested within the ameba endocytic vacuole, making the surface of the capsid accessible for interaction with the vacuole membrane.
3. The content of the capsid is then discharged into the ameba cytoplasm, probably through a fusion between the virus internal lipid membrane and the phagosome membrane, leaving the empty particle in the endocytic vacuole.
4. Transcription of early and late-early genes then occurs in the cytoplasm, most likely under the control of the Mimivirus highly conserved promoter using the virus-encoded and virion-packaged transcription machinery.

The experimentally determined narrow range of Mimivirus host cells, restricted to protozoans belonging to the *Acanthamoeba* genus, conflicts with reports suggesting that Mimivirus might be a human pathogen. La Scola et al. (2005) reported the presence of Mimivirus-specific antibodies in the sera of patients with community- or hospital-acquired pneumonia (see also Berger et al. 2006). In contrast, no evidence of Mimivirus infection was found in hospitalized children in Austria (Larcher et al. 2006), nor in a large CDC-led analysis of respiratory specimens from 496 pneumonia cases (Dare et al. 2008). An isolated case of laboratory infection of a technician by Mimivirus has been reported (Raoult et al. 2006). The patient's serum reacted strongly with several Mimivirus proteins. However, isolation of Mimivirus from the infected patient did not formally link the virus with the disease. Finally, mice experimentally inoculated (via intracardiac route) with Mimivirus developed histopathological features of pneumonia (Khan et al. 2006), but again, no virus was recovered from the lung tissues. In summary, it is not clear whether Mimivirus should be considered a potential pneumonia agent or is simply highly immunogenic (perhaps due to the unique LPS-like layer surrounding its protein capsid) or cross-reacts with a common bacterial species. As a precautionary measure, it is probably best to treat Mimivirus as a biosafety class 2 pathogen. In this context, it is worth remembering that Mimivirus particles remain infectious for at least 1 year when stored at 4–32°C in a neutral buffer.

Genomics of Mimivirus

Overall Genome Structure

Mimivirus genome sequencing revealed a single linear dsDNA molecule of 1,181,404 bp. A combination of bioinformatic methods led to the initial prediction that the virus had 911 protein-encoding genes and six tRNA genes. The exact number of protein-encoding genes may change slightly in the future due to the difficulty of identifying introns in some of the genes that lack relatives in the databases. Proteomic and resequencing data has already led to the correction of some annotations (see Sect. 4.2). With an overall coding percentage of 90.5% and an average intergenic

distance of 157 nt, Mimivirus exhibits the genome compaction observed in other DNA viruses. The large size of the Mimivirus genome is therefore not due to the accumulation of noncoding junk DNA. The overall nucleotide composition is 72% (A+T), leading both to strong positive bias in the usage of A+T-rich synonymous codons and to an increased abundance of amino-acid residues with A+T-rich codons (Raoult et al. 2004). For instance, isoleucine (9.9%), aspargine (8.9%) and tyrosine (5.4%) are twice as frequent in Mimivirus proteins as in ameba or human proteins. On the other hand, alanine (encoded by GCN codons) is a rare amino acid (3.1%). Such surprising flexibility, despite the constraints imposed by the necessity to maintain protein 3D structure, solubility and function, was previously noted for other viruses with even higher A+T (82.2%) contents such as *Amsacta moorei* entomopoxvirus (Bawden et al. 2000). The two strands of the Mimivirus DNA molecule encode roughly the same number of genes (450 R genes vs. 465 L genes). However, both the gene excess and the A+C excess profiles exhibit a clear slope reversal (around nucleotide position 400,000) as found in bacterial genomes; this reversal is usually associated with the origin (or terminus) of replication. Mimivirus genes are preferentially transcribed away from this location (578 leading strand ORFs vs. 333 lagging strand ORFs).

Mimivirus genome termini do not have the large terminal inverted repeats (up to 2 kb) found in Phycodnaviruses (Chlorovirus [Yamada et al. 2006] or *Ectocarpus siliculosus* virus [Delaroque et al. 2003]), its closest NCLDV relatives, but also a conserved feature in Poxviruses and Asfarviruses. The putative circular DNA molecules generated by pairing these repeats might be important during DNA replication. Interestingly, the closest phylogenetic relative of Mimivirus, *Emiliania huxleyi* virus (EhV) 86 appears to replicate as a circular molecule (Wilson et al. 2005). In place of inverted terminal repeats, the Mimivirus genome has a quasi-perfect (616/617) inverted repetition of a 617-bp sequence, beginning at nucleotide position 22,515 and its unique complementary counterpart near the end of the chromosome beginning at nucleotide position 1,180,529. As these regions are intergenic and are not flanked by paralogous genes, their extreme conservation suggests a strong functional constraint related to their perfect base-pairing (Fig. 4). Pairing these inverted repeats leads to a putative Q-like form for the Mimivirus genome, with a long (22,514-bp) and a short (259-bp) tail (Fig. 4). The short tail does not overlap with any ORFs. The long tail has a lower coding density than the rest of the genome (75% vs 90.5%), with larger intergenic distances (435 nt vs 157 nt in average).

This long tail region encodes 12 proteins as follows:

- R1 (795 aa) corresponds to a predicted replication origin binding protein (OBP), homologous to the herpesvirus core gene UL9. Its N-terminal DEAD-like helicase domain is 48% identical to the one found in the R8 (1052 aa). The R1 protein shares its C-terminal domain with the products of nearby genes R8 (38% identity), R9 (49% identity), and R10 (26% identity). This domain of unknown function is also found in other predicted viral OBPs (from Herpesvirus and Asfarvirus). These four proteins might be involved in the DNA replication priming process.

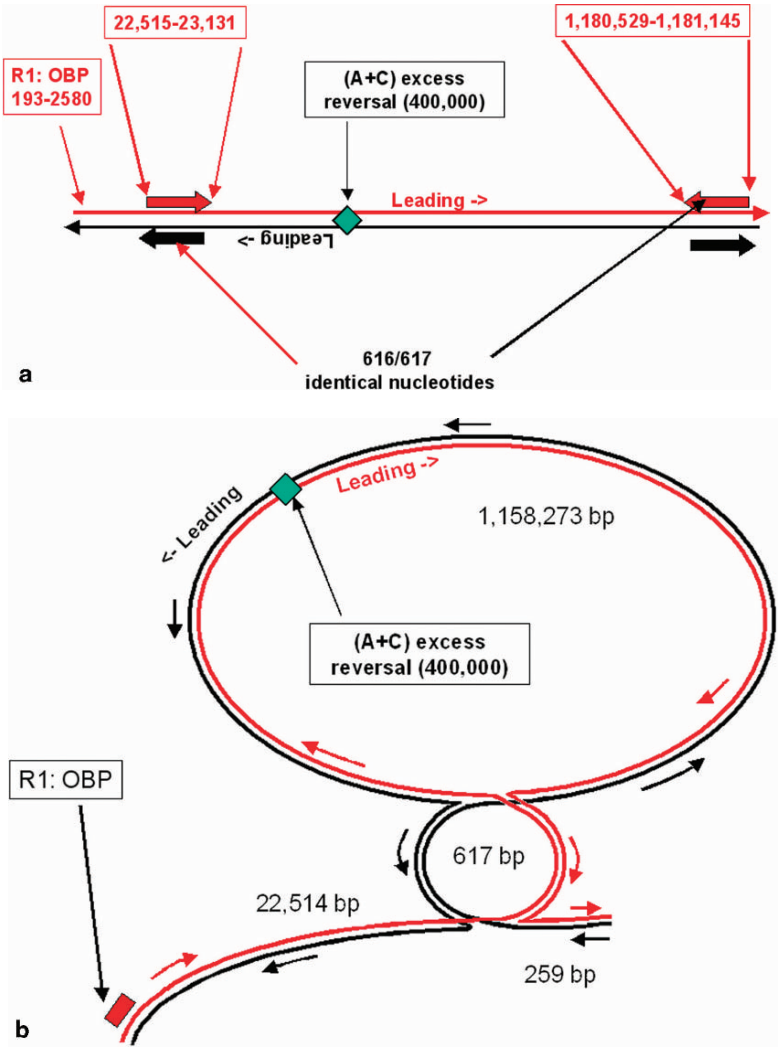

Fig. 4 Schematic structure of Mimivirus chromosome. **a** Remarkable features along the linear chromosome. **b** Putative circularized Q-like form obtained by pairing the largest perfect intergenic inverted repeat in the genome. The first ORF (R1) has a clear similarity to an origin of replication binding protein (OBP)

- L2 (246 aa) has a BRO family, N-terminal domain, which is associated with DNA binding.
- L3 (666 aa) encodes a homolog of chromosome segregation ATPases (COG1196).
- L4 (454 aa) encodes a predicted DNA binding protein of the N1R/P28 type.
- L5 (461 aa) encodes a protein with no functional attribute.
- L6 (218 aa) and L7 (155 aa) are 50% identical, but have no functional attribute or homolog in the database except for the Mimivirus L57 gene product.

- R8 (1052 aa), R9 (376 aa) and R10 (376 aa) are related to R1 and each other as described above.
- R11 (267 aa) encodes a protein of unknown function.
- L12 (487 aa) encodes a protein with no known function, but is 45% identical to L5.

Among the proteins encoded by these 12 genes, a putative function can be attributed to seven of them, all of which are related to DNA replication or binding. These statistics suggest that this clustering of genes encoding DNA replication components at one extremity of the viral chromosome may have functional significance. This makes the proteins of unknown function—L5, L6, L7, R11 and L12—all the more interesting because they may be involved in unknown DNA replication events.

Region		Identity	Overlapping ORFs
98,340–99,316	1,114,002–1,113,026	951 /977 bp	L79–R854 (transposase)
1,007,267–1,008,591	1,112,548–1,113,872	1320/1325 bp	L770–R854 (transposase)

Other remarkable regions of the genomes include two inverted repeats:

These regions may be the result of recent transposase-mediated duplications and thus may be devoid of topological significance.

Mimivirus as a Bona Fide NCLDV

Iyer and collaborators (Iyer et al. 2001) performed a detailed comparative analysis of the protein sequences encoded in the genomes of four families of large DNA viruses (collectively abbreviated as NCLDV) that replicate, completely or partly, in the cytoplasm of eukaryotic cells (poxviruses, asfarviruses, iridoviruses, and phycodnaviruses). They identified nine genes (class 1 core genes) that are shared by all these viruses and 22 more genes that are found in at least three of these four viral families (Class 2 and 3 core genes). Our analysis of the Mimivirus genome unambiguously identified homologs of the nine Class 1 core genes, and 17 of the 22 other core genes. The phylogenetic analysis of a concatenation of the protein sequences encoded by the Class 1 core genes robustly places Mimivirus in an independent lineage (the *Mimiviridae*) among the NCLDVs, in between the *Iridoviridae* and *Phycodnaviridae* (Fig. 5). Together with its morphological (icosahedral capsid) and ultrastructure (nonenveloped particle with an internal lipid membrane; see Sect. 4) characteristics, Mimivirus is clearly a bona fide member of the NCLDVs, despite being much larger and three times genetically more complex than the phycodnavirus EhV86, its closest relative and the NCLDV with the second largest genome.

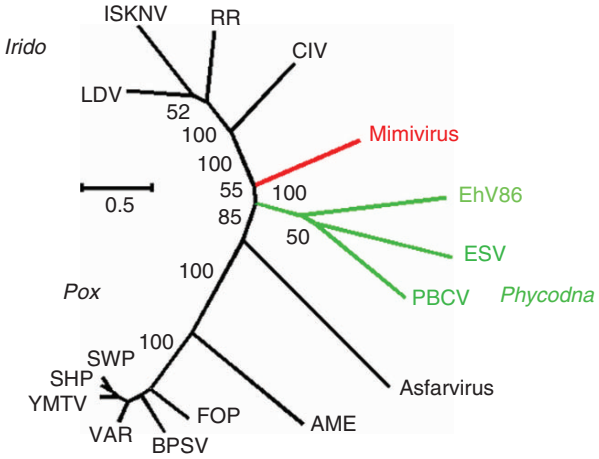

Fig. 5 Phylogenetic position of Mimivirus among established NCLDV families. Viral species are as follows: Iridoviridae (*CIV* chilo iridescent virus, *RR Regina ranavirus*, *LDV*, lymphocystis disease virus type 1, *ISKNV* Infectious spleen and kidney necrosis virus), *Poxviridae* (*SWP* swinepox virus, *SHP* sheeppox virus, *YMTV* Yaba monkey tumor virus, *VAR* variola virus, *BSPV* bovine popular stomatitis virus, *FOP* fowl pox virus, *AME Amsacta mooreii* entomopoxvirus), Asfarvirus: African swine fever virus, *Phycodnaviridae* (*PBCV Paramecium bursaria* chlorella virus 1, *ESV Ectocarpus siliculosus* virus, *EhV86 Emiliania huxleyi* virus 86). This tree was built using maximum likelihood and based on the concatenated sequenced of the proteins sequences encoded by the NCLDV Class 1 core genes. Bootstrap percentages are shown along the branches (except for the Pox families where they all are close to 100)

Interestingly, prior to the sequencing of EhV86, phycodnaviruses (algal virus, such as virus-infecting chlorella) were the closest relatives to Mimivirus. However, despite their large genomes (>300 kb) they lack a virus-encoded RNA polymerase, which made them unique among the NCLDVs (Van Etten 2003) and at odds with the plethoric gene content of Mimivirus. This apparent paradox is now alleviated with the finding of a complete transcription apparatus encoded in EhV86, the largest phycodnavirus genome sequenced to date. As it is unlikely that the many different genes required to constitute a functional transcription apparatus were independently acquired by EhV86, their presence strongly suggests that all extant NCLDV families (*Poxviridae*, *Asfarviridae*, *Iridoviridae*, *Phycodnaviridae*, and *Mimiviridae*) share a common ancestor (a mostly cytoplasmic DNA virus) that might have been even more complex than today's Mimivirus. The smaller phycodnavirus (Chloroviruses and Phaeoviruses) genomes probably underwent lineage-specific gene losses leading to the (still surprising) disappearance of their host-independent transcription apparatus. Similarly, Iridoviruses, Asfarviruses, and Poxviruses are all missing a few of the NCLDV core genes (Raoult et al. 2004), which may correspond to lineage-specific losses. Despite its huge genome, Mimivirus is not immune to this phenomenon, for instance with the puzzling

absence of a dUTPase homolog, a universal enzyme that is required to avoid the incorporation of deoxyuridine into its thymidine-rich DNA. An equivalent activity might be performed by the protein product of ORF L479 that has a MazG-type nucleotide pyrophosphohydrolase domain, probably derived from a bacterial source (Iyer et al. 2006). Similarly, the ATP-dependent DNA ligase present in all other NCLDVs is replaced by an NAD-dependent version of the enzyme in Mimivirus (R303). A succession of gene losses, some of them compensated by nonorthologous gene replacements via horizontal transfers, might explain the small number of recognizable orthologous genes shared by today's NCLDVs.

Mimivirus as a Unique Giant Virus: Virally Encoded Translation Components

If, on one hand, Mimivirus exhibits many of the features characteristic of previously described NCLDVs, on the other hand its 1.2-Mb genome encodes many unique genes not previously found in a virus (Raoult et al. 2004). For instance, Mimivirus possesses a complete set of DNA repair enzymes capable of correcting nucleotide mismatches as well as errors induced by oxidation, UV irradiation and alkylating agents. Mimivirus is also the only virus to encode the three major types of topoisomerases (the usual type IIA, the poxvirus-like type Ib, and the first viral type Ia). In addition, Mimivirus uniquely possesses a number of polysaccharide, amino-acid and lipid manipulating enzymes. Such metabolic capabilities, although covering a broader biochemical spectrum in Mimivirus, also exist in other NCLDVs, specially the phycodnaviruses (Van Etten 2003), where they often differ from one species to the next, suggesting their involvement in specific virus–host relationships. EhV86, for instance, encodes many components of the ceramide biosynthesis pathway (Wilson et al. 2005) that are thought to interfere with its host apoptosis-like cell death pathway.

Probably the most spectacular discovery in the Mimivirus genome was finding ten homologs of proteins with functions central to protein translation: four aminoacyl-tRNA synthetases, a mRNA cap-binding protein (eukaryotic initiation factor eIF4E, ORF L496), translation initiation factor eEF-1 (GTP-binding translocation factor, ORF R624), translation initiation factor SUI1/eIF1 (ORF R464), translation initiation factor eIF4A (an ATP-dependent RNA helicase, ORF R458), and peptide chain release factor eRF1 (ORF R726). In addition, Mimivirus encodes a homolog (ORF R405) of the tRNA (uracil-5-)-methyltransferase, the tRNA-modifying enzyme whose *Escherichia coli* counterpart catalyzes the methylation of the invariant tRNA uracil at position 54, thus defining the T-loop (TΨC arm) in all tRNAs. This region of the tRNA serves as a recognition site for the ribosome.

The four aminoacyl-tRNA synthetases (aaRS), all from Class I aaRSs, are specific for tyrosine (TyrRS), arginine (ArgRS), cysteine (CysRS) and methionine (MetRS). Finding these components of the translation apparatus in Mimivirus clearly violated the dogma that viruses rely entirely on the host translation machinery for protein synthesis.

Genes encoding tRNAs were previously described in a few viruses, including bacteriophage T4, herpes virus 4, and the chlorella viruses (Van Etten 2003). Similarly, Mimivirus encodes six tRNA-like genes, albeit mostly unrelated to the above aaRSs: three $tRNA_{leu}$ (2 TTA, 1 TTG), one $tRNA_{trp}$ (TGG), one $tRNA_{his}$ (CAC), and one $tRNA_{cys}$ (TGG). Although Mimivirus exhibits a codon usage that is fairly distinct from its ameba host, the above tRNAs or aaRSs are not related to the most conspicuous differences. One exception to this statement is TyrRS, which may help incorporate tyrosine into Mimivirus proteins where its frequency (5.4%) is twice that observed in ameba proteins.

The presence of many translation machinery components encoded in the genome of Mimivirus can be explained by two opposing hypotheses. On one hand, the traditional view that viruses capture genes from their environment and their host predicts that these translation components were acquired from cellular organisms (Moreira and Lopez-Garcia 2005). However, this hypothesis suffers from a lack of phylogenetic evidence (Ogata et al. 2005a). Also, the independent random acquisition (and retention) of so many translation-related genes is unlikely, given their lack of usefulness as individual components of an incomplete system. On the other hand, one may interpret the translation components found in Mimivirus as the remains of an even more complex ancestral genome that encoded a complete and functional translation apparatus, as occurs in cellular organisms. Such a genome reduction scenario is consistent with the hypothesis that NCLDVs originated from the primitive nucleus of ancestral eukaryotes (Claverie 2006). A genome reduction process, akin to the one observed in intracellular parasitic bacteria (Blanc et al. 2007) may have led to the present day Mimivirus. However, this evolutionary process appears to have stopped as Mimivirus shows no signs of ongoing genome degradation such as pseudogenes, repeat accumulation, or reduced coding density (Claverie et al. 2006). Presumably, there is a strong selective advantage for Mimivirus to retain its incomplete translation system, given the high evolutionary rates usually associated with viruses. In this context, assessing the biochemical and cellular function of Mimivirus-encoded translation components during the infection process becomes important (see Sect. 5).

Other Remarkable Features of the Mimivirus Genome

Intein and Introns

Inteins are protein-splicing domains encoded by mobile intervening sequences. They catalyze their own excision from the host protein. Although found in all domains of life (Eukarya, Archaea and Eubacteria) their distribution is sporadic. Mimivirus is one of the few dsDNA viruses containing an intein, inserted in its DNA polymerase B protein (Ogata et al. 2005b). The Mimivirus intein is closely related to one found in the DNA polymerase of *Heterosigma akashiwo* virus (HaV) a phycodnavirus that infects the single-cell bloom-forming raphidophyte

(golden brown alga) *H. akashiwo* (Nagasaki et al. 2005). Both inteins appear monophyletic to archaeal inteins. Two additional inteins have recently been reported in chlorovirus proteins (Fitzgerald et al. 2007). Type I introns are self-splicing intervening sequences that are excised at the mRNA level. One type IB intron has been identified in several chlorella viruses, but they are rare in viruses infecting eukaryotes. Mimivirus has six self-excising introns: one in the largest RNA polymerase subunit gene and the other three in the second-largest RNA polymerase subunit gene. Two introns were recently discovered in the gene encoding the major capsid protein (L425, now corrected in the UniProt Q5UQL7 entry). Given that introns are mostly detected when they interrupt the coding sequence of know proteins, additional introns located within anonymous Mimivirus ORFs might exist.

A Uniquely Conserved Promoter Signal

An exhaustive search for overrepresented "words" in the Mimivirus genome led to the discovery of an "AAAATTGA" octamer within the 150-nt upstream region of 403 of the 911 (45%) predicted protein-coding genes. A search for more sophisticated signals (Bailey and Gribskov 1998) led to a very similar result with 446 genes (49%) showing a conserved upstream motif. The location of this motif at positions ranging from –80 to –50 before the initiator codon is consistent with the short average size (157 ±113 nt) of the intergenic region in Mimivirus, and the compact promoter/5′ UTR structure (as well as 3′UTR) known for some ameba protists (Vanacova et al. 2003). Suhre et al. (2005) proposed that the AAAATTGA octamer might correspond to a TATA box-like core promoter element. The finding of such a strongly conserved sequence motif in front of nearly half of the Mimivirus genes is one more unique feature of this virus because eukaryotic (as well as viral) promoters usually lack clear consensus sequences.

There is a significant correlation between the upstream AAAATTGA motif and genes transcribed from the predicted leading strand (54% vs 40%). Finally, Suhre et al. (2005) noted that this motif was not common in the available ameba genome sequences. Applying this same analysis to the genomes of other large DNA viruses, confirmed that the homogeneity of this promoter sequence is unique to Mimivirus.

Based on the predicted function of the proteins encoded by the genes possessing the AAAATTGA motif in their upstream region, this putative promoter element appears to correlate with functions required for the early (or late-early) phase of viral infection. Suhre et al. (2005) also proposed that this Mimivirus TATA box-like signal might have co-evolved with the virus-encoded transcription preinitiation complex consisting of two RNA polymerase II subunits and a TFIID initiation factor homolog. This late-early promoter may thus be recognized by the Mimivirus encoded-transcription machinery, while the genes lacking this signal could be transcribed by the host RNA polymerase. This hypothesis received additional support from the proteomic analysis of Mimivirus particles showing that: (i) the virus-encoded transcription machinery is associated with the particle (and thus accessible

immediately after infection), and (ii) only a small fraction (approximately 10%) of the late genes encoding proteins associated with the virion have the AAAATTGA promoter element. Finally, it is worth noting that the exact same sequence was shown to function as a promoter for the *Chilo* iridescent virus DNA polymerase gene (Nalçacioglu et al. 2007).

Gene and Genome Duplication in Mimivirus

The Mimivirus genome is roughly 2.4 times larger than the second largest virus (Phage G) and ten times larger than the average DNA viruses (www.giantvirus.org). This observation raises the question of the mechanisms by which such a viral genome might have occurred, and of the evolutionary forces allowing such an anomalous genome to be maintained. DNA viruses vary widely in DNA content as well as in their genetic complexity. Larger genomes may result primarily from the accumulation of noncoding DNA. An extreme example is the *Cotesia congregata* Bracovirus, a *Polydnaviridae* with a 568-kb genome, but encoding a mere 156 proteins, for an overall 27% coding density (Espagne et al. 2004). Closer to Mimivirus, some sequenced members of the *Iridoviridae* (Zhang et al. 2004) or the *Baculoviridae* (Cheng et al. 2002) have coding densities below 69%.

With a coding density above 90%, the size of Mimivirus genome cannot be explained by a propensity to accumulate junk DNA. After a thorough analysis of the gene content, Suhre (2005) identified two main mechanisms contributing to the Mimivirus genome size. First, a segmental duplication of about 200 kb is at the origin of the telomeric regions of the Mimivirus linear genome. Second, many tandem gene duplications exist in various positions in the genome, sometimes generating large paralogous families of up to 66 members (these are ankyrin-domain containing ORFs). A perfect tandem expansion of 12 paralogous ORFs occurs from L174 to L185. Depending on the similarity threshold used, approximately 35% of the Mimivirus genes have at least one homolog in the virus's genome (E value $<10^{-5}$). This fraction lies well within the range of values encountered throughout the three domains of life: for example 17% for *Haemophilus influenzae*, 44% for *Mycoplasma pneumoniae*, 30% for *Saccharomyces cerevisiae*, and 65% for *Arabidopsis thaliana* (Suhre 2005 and references therein). From a different perspective, Ogata et al. (2005a) showed that horizontal gene transfer from its host, or other exogenous sources does not account for much of the Mimivirus genome. Despite their crude methodology, overestimating horizontal gene transfers, Filee et al. (2007) identified less than 10% of Mimivirus genes as originating from bacteria. Overall, these studies do not indicate that Mimivirus is quantitatively different from cellular organisms or other large DNA viruses (Monier et al. 2007) with respect to the various mechanisms leading to genome expansion. If a fraction of the Mimivirus genome can be attributed to segmental or tandem duplications, as well as to exogenous sources, this leaves roughly 400 unique genes that might be part of Mimivirus lineage, dating back to its origin.

The Mimivirus Particle

Morphology and Ultrastructure

Despite its unprecedented size, the icosahedral symmetry of the Mimivirus particle was good enough to allow Xiao et al. (2005) to make a computer-generated 3D reconstruction at a resolution of approximately 75 Å from series of cryo-electron microscopy (cryoEM) images. According to the reconstruction, the Mimivirus capsid has a diameter of approximately 0.5 μm, and is covered by 0.125 μm long, closely packed fibers (Fig. 3). The total diameter of a free particle is thus roughly 0.75 μm, consistent with its visibility in the light microscope (Fig. 1). Mimivirus has a pseudo-triangulation number of approximately 1180, predicting that the capsid contains approximately 70,000 individual molecules of the L425 ORF-encoded major capsid protein. Inside the 70-Å-thick protein shell, the cryoEM images reveal two 40-Å-thick lipid membranes, a structure also found in some other NCLDVs such as African swine fever virus (Asfarvirus); the phycodnaviruses and iridoviruses have a single membrane inside their capsids (Xiao et al. 2005).

The chemical nature of the fibers projecting from the outer layer of the particle is unknown. Mimivirus encodes eight large proteins with triple-helix forming collagen repeats: L71 (945 aa, seven repeats), R196 (1595 aa, nine repeats), R238 (441 aa, one repeat), R239 (939 aa, eight repeats), R240 (817 aa, six repeats), R241 (812 aa, three repeats), L668 (1387 aa, six repeats), L669 (1937 aa, ten repeats). Furthermore, Mimivirus possesses a homolog of the procollagen-lysine, 2-oxoglutarate-5-dioxygenase that catalyzes the posttranslational formation of hydroxylysine in X-Lys-Gly sequences, and one putative prolyl-4-hydroxylase (L593) that forms 4-hydroxyproline in -X-Pro-Gly sequences. Hydroxylysines are involved in the intermolecular crosslinking of collagen molecules, and hydroxyproline plays a central role in collagen folding and stability. In addition, a fraction of collagen hydroxylysine residues are the target of O-glycosylation. Consequently, it is tempting to speculate that the fiber layer surrounding the Mimivirus particle is made of a dense mesh of crosslinked and glycosylated collagen-like gene products.

Paradoxically, none of the proteins predicted to constitute the fiber layer (the above-mentioned ORF products with collagen repeats) were detected in the particle proteome (Table 1; Renesto et al. 2006). This result might be the consequence of their heavy crosslinking, making them irreversibly insoluble and excluding them from gel electrophoresis and subsequent mass-spectrometry analysis. The same explanation may apply to the three paralogs of the major capsid protein L425 (R439, R440, R441) that are not detected in the proteomic analysis. Another unique feature of Mimivirus particles is the presence of a pentagonal star-shaped structure centered at a single vertex of the icosahedral capsid (Fig. 6). This feature, nicknamed stargate, was proposed to play a central role in initiating the viral–phagosome membrane fusion by Dr. Nathan Zauberman (see www.weizmann.ac.il/ Organic_ Chemistry/minsky/nathan/mimivirus.shtml).

Table 1 Mimivirus particle proteins

ORF	Protein annotation	Identification by Renesto et al. (2006)	Identification by this study (E-value)
R1	Replication origin binding protein	–	<0.1
L3	Unknown	–	<0.1
R10	Unknown	–	<0.1
L12	Unknown	–	<0.1
L18	Unknown	–	<0.1
L48	Unknown	2D	–
L56	Ankyrin-containing protein	–	<0.01
L65	Virion-associated membrane protein	1D	–
L66	Ankyrin-containing protein	–	<0.01
R69	Unknown	–	<0.01
L86	Ankyrin-containing protein	–	<0.1
L90	Unknown	–	<0.1
L98	Unknown	–	<0.1
L98b	New short ORF Unknown [37 aa; complement (124358.. 124471)] overlapping L98	–	<0.1
L116	Unknown	–	<0.1
L122	Ankyrin-containing protein	–	<0.1
R135	Choline dehydrogenase or related protein	1D–2D	<0.01
L137	Glycosyl-transferase domain	–	<0.1
L145	Unknown	–	<0.1
R160	Unknown	1D	–
R161	Unknown	1D	–
L164	Cysteinyl-tRNA synthetase	–	<0.1
L172	Unknown (L cluster)	–	<0.1
L173	Unknown (L cluster)	–	<0.1
L177	Unknown (L cluster)	–	<0.1
L180	Unknown (L cluster)	–	<0.1
L183	Unknown (L cluster)	–	<0.1
R186	Putative transposase	–	<0.1
R188	Unknown, similar to AAQ60770 from *Chromobacterium violaceum*	1D–2D	<0.01
R194	Topoisomerase I (pox-like)	1D	–
R195	Glutaredoxin (ESV128 type)	1D	–
L208	Unknown	1D	<0.1
L221	Topoisomerase I (bacterial type)	1D	<0.01
R225	Unknown	–	<0.1
L228	Unknown	–	<0.01
L230	Procollagen-lysine,2-oxoglutarate 5-dioxygenase	–	<0.01
L232	Protein kinase domain	–	<0.1
L235	RNA polymerase subunit 5	1D–2D	<0.01
L244	RNA polymerase II second Largest submit (Rpb2)	ID –	–
R253	Unknown	1D–2D	<0.01
L264	Unknown	1D	–
L269	Unknown	–	<0.1
L271	Ankyrin-containing protein	–	<0.1

(continued)

Table 1 (continued)

ORF	Protein annotation	Identification by Renesto et al. (2006)	Identification by this study (E-value)
L274	Unknown	1D	–
L279	Unknown	–	<0.1
L293	Unknown	1D	–
L294	Unknown	1D–2D	<0.01
URF130	Unknown (149 aa; 382733..383182)	–	<0.1
R301	Uncharacterized protein (Chilo iridescent virus 380R)	1D	–
R307	Protein phosphatase 2C domain	1D	–
L309	Unknown	1D	–
R311	BIR domain (Chilo iridescent virus 193R)	–	<0.1
L318	DNA polymerase family X	1D	–
R322	DNA polymerase (B family)	–	<0.01
R326	Unknown	1D	–
R327	Unknown	1D	–
L330	Unknown	1D–2D	<0.01
L334	Unknown	–	<0.1
R341	Putative polyadenylate polymerase	1D–2D	<0.01
R345	Unknown	1D–2D	<0.01
R347	Unknown	1D	–
R349	Unknown	–	<0.1
R350	Putative transcription termination factor, VV D6R helicase	1D	<0.01
L352	Unknown	1D	–
R355	Unknown	1D	–
L357	Unknown	–	<0.1
R362	Thioredoxin domain	1D–2D	<0.1
R366	Helicase domain	–	<0.01
L376	Unknown	1D–2D	<0.01
L377	Putative NTPase I	1D	–
R382	mRNA capping enzyme	1D	–
R383	Unknown	1D	–
R387	Unknown	1D–2D	<0.01
L389	Unknown	1D	<0.1
L394	Unknown	–	<0.1
R395	Similar to EsV-1–87 (*Ectocarpus siliculosus* virus)	–	<0.1
L396	VV A18 helicase	–	<0.1
R398	Calcineurin-like phosphoesterase domain	1D	–
L399	Unknown	1D	–
R400	S/T protein kinase, similar to PBCV–1 A617R	1D	–
R402	Unknown	1D–2D	<0.01
R403	Unknown	1D	–
R406	Alkylated DNA repair	–	<0.1
R407	tRNA (uracil–5-)-methyltransferase	2D	<0.1
L410	Similar to poxvirus P4B major core protein	1D–2D	<0.01
L417	Unknown	1D	–
R423	Unknown	–	<0.1
L425	Capsid protein 1, [SWISS-PROT: Q5UQL7], complement(join	1D–2D	<0.01

(continued)

Table 1 (continued)

ORF	Protein annotation	Identification by Renesto et al. (2006)	Identification by this study (E-value)
	(557530..559233,559658.. 559681,560873..560926))		
R429	PBCV1-A494R-like	–	<0.1
L437	VV A32 virion packaging ATPase	–	<0.1
L442	Unknown	1D–2D	<0.01
R443	Thioredoxin domain	1D–2D	<0.01
L446	Patatin-like phospholipase (463L)	–	<0.1
R449b	New short ORF Unknown (68 aa; 592148..592354) (overlapping R449)	–	<0.1
L452	Unknown	1D	–
L454	Unknown	1D	–
R457	Unknown	1D	–
R459	Unknown	1D–2D	<0.01
R463	Unknown	1D	–
R470	DNA-directed RNA polymerase subunit L	1D–2D	<0.01
R472	Unknown	1D	<0.01
R476	ATPase domain	–	<0.1
R480	Topoisomerase II	–	<0.1
L484	Ankyrin-containing protein	1D	–
L485	Unknown	2D	<0.01
R486	Two PAN domains	1D	–
L488	Unknown	1D	<0.1
R489	Unknown	1D–2D	<0.01
L492	Unknown	1D	<0.1
L498	Zn-dependent alcohol dehydrogenase	1D	–
R501	RNA polymerase II largest subunit (Rpb1)	1D	–
R510	Putative replication factor C subunit	–	<0.1
L515	Unknown	1D	<0.1
L516	Unknown	1D	–
R526	Putative triacylglycerol lipase	1D–2D	<0.01
R528	Unknown	1D	–
L532	Cytochrome p450 domain	2D	<0.1
L533	Unknown	1D	–
L538	Helicase conserved C-terminal domain (PFAM)	1D	–
L540	VVI8 helicase	1D	<0.1
L544	Transcription initiation factor TFIIB	1D	–
L550	Unknown	1D–2D	<0.01
R553	Unknown	1D–2D	<0.01
R557	Unknown	1D	–
R559b	New short ORF Unknown (90 aa; 750051..750323) overlapping R559	–	<0.01
R563	Helicase conserved C-terminal domain	1D	–
R566	Unknown	–	<0.01
L567	Unknown	1D–2D	<0.01
R571	Patatin-like phospholipase (similar to Chilo iridescent virus 463L)	–	<0.1

(continued)

Table 1 (continued)

ORF	Protein annotation	Identification by Renesto et al. (2006)	Identification by this study (E-value)
L581	Unknown	–	<0.01
R584	Unknown	1D–2D	<0.01
L585	Unknown	1D	–
L591	Unknown	1D–2D	–
R592	Helicase conserved C-terminal domain	–	<0.1
L593	Prolyl 4-hydroxylase	1D	–
R596	Thiol oxidoreductase E10R	1D–2D	<0.01
R607	Unknown	–	<0.01
R607b	New short ORF Unknown (31 aa; 801314..801409) overlapping R607	–	<0.1
R610	Proline-rich protein	1D–2D	<0.1
L611	Unknown	–	<0.01
L612	Mannose–6P isomerase	1D–2D	<0.01
R639	Methionyl-tRNA synthetase	–	<0.1
R641	Unknown	–	<0.01
R642	Unknown	–	<0.1
R644	Putative phosphatidylethanolamine-binding protein	1D	–
R646	Unknown	1D	–
L647	Unknown	1D–2D	–
R648	Unknown	1D–2D	<0.01
R653	Unknown	1D–2D	<0.01
R658	Unknown	1D	–
R661	Unknown	–	<0.1
R663	Arginyl-tRNA synthetase	–	<0.1
L670	Protein kinase domain and Cyclin N-terminal domain	–	<0.1
R679	Unknown	1D	–
URF277	Unknown (439 aa; complement [904734..906053])	–	<0.01
L687	Endonuclease for the repair of UV-irradiated DNA	1D	–
L688	Unknown	1D–2D	–
L690	Unknown	1D–2D	<0.01
R691	Unknown	1D	–
L692b	New short ORF Unknown (78 aa; complement [911840..912076]) overlapping R692	–	<0.1
R692	Unknown	1D–2D	<0.01
R695	Unknown	1D	–
L701	Unknown	1D	–
R705	Unknown	1D	–
R706	Unknown	1D	–
R710	Unknown	1D	<0.1
L720	Hydrolysis of DNA containing ring-opened N7 methylguanine	–	<0.1
R721	Similar to CheD, chemotaxis protein	1D–2D	<0.1
R722	Unknown	1D	–
L724	Unknown	1D–2D	<0.01
L725	Unknown	1D–2D	<0.01
R727	Unknown	1D	–
R741	Unknown	–	<0.01

(continued)

Table 1 (continued)

ORF	Protein annotation	Identification by Renesto et al. (2006)	Identification by this study (E-value)
R745	Unknown	–	<0.1
R753	Unknown	–	<0.01
L754	Unknown	–	<0.01
R756	Similar to predicted Fe-S-cluster redox enzyme	–	<0.01
L763	Unknown	–	<0.1
L766	Unknown	–	<0.1
L767	Unknown	–	<0.01
L774	Unknown	–	<0.1
R776	Unknown	–	<0.1
L778	Unknown	1D	–
R787	Ankyrin-containing protein	–	<0.1
L794b	New short ORF Unknown (66 aa; complement [1033747..1033947]) overlapping L694	–	<0.1
R811	Unknown	–	<0.01
R826	Two protein kinase domains	–	<0.01
L829	Unknown	1D–2D	<0.01
L834	Unknown	–	<0.1
R841	Ankyrin-containing protein	–	<0.01
R842	Unknown	–	<0.1
L851	Unknown	1D	–
L872	Unknown	1D–2D	<0.01
R877	Putative outer membrane lipoprotein	1D	–
R878b	New short ORF Unknown (72 aa; 1142194..1142412) overlapping R878	-–	<0.1
L893	Putative oxidoreductase (C-term)	1D	–
L894	Putative oxidoreductase (N-term)	1D	–
L899	Unknown	1D	–
R903	Unknown	–	<0.1
L909	Unknown	–	<0.01

Twenty-three new proteins identified at the high confidence level (E-value<0.01) are highlighted in grey.

Mimivirus encodes several enzymes usually involved in the synthesis of complex reticulated polysaccharides such as perosamine, found in the O-antigen moiety of the lipopolysaccharide (LPS) of various bacteria. The outer layer of the Mimivirus particle may resemble a bacterial cell wall, explaining its retention of the Gram stain. The presence of this polysaccharide layer could also make it palatable for its ameba host, the phagocytosis of which is both triggered by bacterial-sized particles (>0.6 μm) and enhanced by the recognition of surface sugar moieties (reviewed in Claverie et al. 2006). It is likely that Mimivirus is packaged in this spore-like structure, and that the digestion of the fiber outer layer by the ameba endocytic vacuole is a prerequisite to a productive infection. The virus-host specificity (*Acanthamoeba*) might be in part dictated by the presence of the necessary enzymes in the phagosomes of various ameba species (Weekers et al. 1995). Electron micrographs of Mimivirus in the phagocytic vacuole of its host suggest that a significant disruption of the particle's outer layer occurs at this time of the infection (Fig. 7).

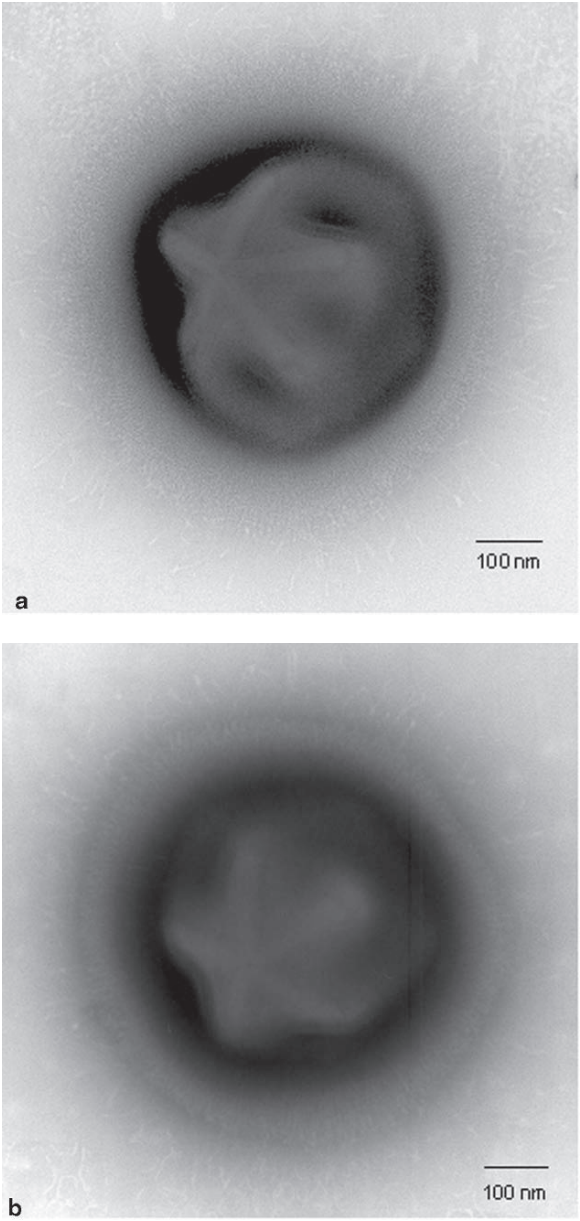

Fig. 6a,b TEM pictures (200 kV with FEI CM 200, ×50,000) of negatively stained (2% uranyl acetate) Mimivirus particles exhibiting a single pentagonal star-shaped structure. Also note the dense layer of fibers covering the protein capsid. Courtesy of Dr. Wai Li Ling and Dr. Jorge Navaza, Institut de Biologie Structurale, Grenoble

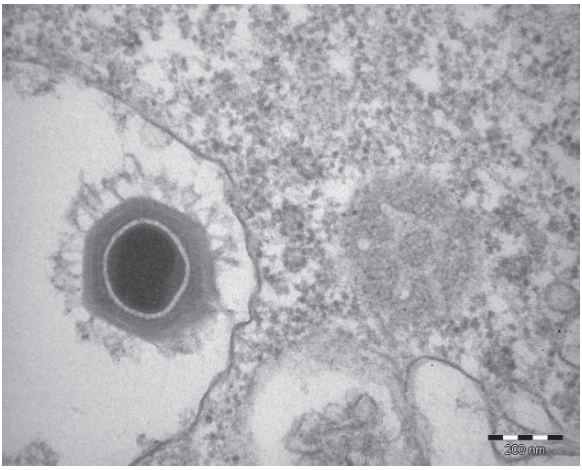

Fig. 7 Mimivirus particle within an *Acanthamoeba polyphaga* phagocytic vacuole. Note the disorganization of the fiber outer layer compared to its appearance around free particles (Fig. 2B). From Raoult et al. (2007)

Like bacteria, the LPS-like antigens at the surface of the virus particles might constitute the dominant Mimivirus epitope. Gold-labeled antibody molecules against Mimivirus bind to the outer layer of the virus (La Scola et al. 2005). In this context, the frequent seroconversion of pneumonia patients attributed to Mimivirus infection, although no Mimiviruses were isolated from these patients (Berger et al. 2006), could be the result of crossreactivity with a cell-wall antigen from a common bacterial human pathogen.

Proteomics of the Virion

The composition of purified virions was analyzed by using total extracts, and 1D and 2D gel electrophoresis followed by MALDI-TOF mass-spectrometry analysis of the in-gel trypsin digested bands. Proteins encoded by 114 Mimivirus genes were identified (Renesto et al. 2006). The function of over half of these virion-associated proteins are unknown. The 2D gels revealed numerous isoforms, probably due to posttranslational modifications such as glycosylation, acetylation and phosphorylation.

In addition to the expected major structural components (e.g., the major capsid protein L425 and core L410 protein), transcription enzymes and factors (12 gene products) constitute the largest functional category associated with the viral particles. This set includes all five predicted DNA-directed RNA polymerase subunits, two helicases (R350, L540), the mRNA capping enzyme, and four transcription factors (L377, L538, L544, R563), including a TATA box-like binding protein. The completeness of the transcription machinery components in the Mimivirus particles resembles the poxviruses (Zachertowska et al. 2005; Yoder et al. 2006; Resch et al. 2007) and is expected for a DNA virus that replicates predominantly or exclusively in the cytoplasm.

The next largest functional group contains nine gene products associated with oxidative pathways. These enzymes might help the virus cope with the oxidative stress generated by the host defense. The protein/lipid modification functional category is also well represented, including a phosphoesterase and a lipase, which are eventually used for digesting the cell (vacuole) membrane, two protein kinases, and a protein phosphatase. Finally, five proteins associated with DNA topology and damage repair are in the virion, including topoisomerases IA and IB and a DNA UV damage repair endonuclease (Renesto et al. 2006).

We used this review article to reanalyze the previously generated tryptic peptide mass lists from 487 spots in the 2D gels (Renesto et al 2006), using less stringent, albeit statistically sound criteria. Our main purpose is to eventually identify small exons (or ORFs) that might have been overlooked in the initial analysis of the Mimivirus genome. Each of the mass data was searched against a hybrid database containing all the previously annotated sequences (911 standard ORFs and 347 downgraded URFs) complemented with all other small ORFs (30–99 codons; 6,393 ORFs in total) delineated in the Mimivirus genome. The statistical significance of peptide identifications was assessed by randomizing every sequence in the hybrid database 100 times and repeating the same searches against this randomized database. We obtained an E-value (expected number of protein hits) for a given number of identified peptides in a protein of a given size range. Table 1 shows all 200 ORFs that were reported by Renesto et al. (2006) plus those identified here with an E-value less than 0.1; 23 new proteins identified at the high confidence level (E-value <0.01) are highlighted in grey. They include the B-type DNA polymerase (R322), several ankyrin repeat-containing proteins (L56, L66, R841), the procollagen-lysine hydroxylase (L230), a helicase (R366), a kinase (R826), 13 proteins of unknown functions, URF277, and a short ORF not previously annotated (overlapping with R559). Less confident identification includes the replication origin binding protein (R1), topoisomerase II (R480), putative replication factor C (R510), and three of the four tRNA synthetases (Met-, Arg-, CysRS). In addition, we putatively identified two URFs (URF130, URF277) and seven small ORFs (31–90 aa) not annotated in the original work. This analysis also allowed us to correct the sequence of the major capsid protein (L425, Uniprot Q5UQL7), two exons of which were previously overlooked (see Sect. 3.3.1).

Experimentally Validated Mimivirus Genes

Most of our knowledge on Mimivirus is derived from bioinformatic analyses of its genome sequence, proteomics, or electron microscopy studies, all of which are subject to overinterpretation. Following the initial excitement of the discovery of this exceptional virus, a few groups initiated postgenomic studies on the biochemistry and physiology of Mimivirus. The various genes for which molecular studies are in progress are summarized in this section.

Mimivirus-Encoded Components of the Translation Machinery

Genes encoding aminoacyl-tRNA synthetases were probably the most unexpected finding in Mimivirus. Therefore, they became the immediate focus of functional and structural studies in our laboratory. The status of these studies follows:

- The cysteinyl-tRNA synthetase (L164) was expressed in *E. coli* and the protein obtained in a soluble form.
- The methionyl-tRNA synthetase (R639) was expressed in *E. coli* and purified. Its enzymatic function has been characterized and it is specific for both Met and the eukaryotic $tRNA_{Met}$ (Abergel et al. 2007).
- Studies on the tyrosyl-tRNA synthetase (L124) are the most advanced. Following its production in *E. coli*, its enzymatic function was characterized and its specificity for the Tyr amino-acid and the eukaryotic $tRNA_{Tyr}$ validated using a panel of mutant $tRNA_{Tyr}$s. In addition, the crystal structure of the enzyme was determined at 2.2-Å resolution (Abergel et al. 2005, 2007). Mimivirus tyrosyl-tRNA synthetase has unique characteristics in its anti-codon recognizing regions and homodimer organization.

Mimivirus-Encoded Nucleotide Metabolism and DNA Replication Enzymes

Our laboratory also initiated a systematic characterization of the nucleotide metabolism enzymes. Enzymes being studied include:

- The deoxynucleotide monophosphate kinase (DNK, R512) has been expressed and purified. It behaves as a dimer. It is active with the two substrates dCMP and dGMP. Bacteriophage T4 DNK is the only member of this family of enzymes that recognizes three structurally dissimilar nucleotides: dGMP, dTMP and 5-hydroxymethyl-dCMP, while excluding dCMP and dAMP. The mimivirus homolog has 29% amino acid identity over a region of 120 residues with the T4 enzyme.
- The 584-aa R341 gene product, initially annotated as unknown, contains a polyadenylate polymerase domain at its N-terminus. The protein is associated with Mimivirus particles (Table 1) and African swine fever virus has a homolog (C475L). The polyadenylate polymerase is responsible for adding a poly A tail to the 3Î end of mRNA in eukaryotes. Its identification in Mimivirus adds to a number of virally encoded proteins involved in transcription. The R341 gene product has been expressed, purified and crystals diffracting at 4-Å resolution have been obtained.
- Studies on the nucleoside diphosphate kinase (NDK, R418), the first virus-encoded protein of its kind, are the most advanced. This enzyme usually catalyzes the synthesis of nucleoside triphosphates (NTPs) other than ATP. A detailed characterization of its enzymatic activity showed that the Mimivirus enzyme has a strong preference for deoxypyrimidine nucleotides (Jeudy et al. 2006). This

property might represent an adaptation to the production of the limiting TTP deoxynucleotide required for the replication of the large A+T-rich (72%) Mimivirus genome. The viral NDK might also assume a role in dUTP detoxification to compensate for the surprising absence of a Mimivirus dUTPase (deoxyuridine triphosphate pyrophosphatase), an important enzyme conserved in most viruses. The crystal structure of the enzyme in complex with various ligands has been obtained at 2.2-Å resolution (PDB: 2B8P, 2B8Q). Additional structural studies are in progress on the NDK to investigate the role of its shorter Kpn-loop and other specific Mimivirus features in the active site on its substrate specificity (Jeudy et al. 2005).

Two additional Mimivirus enzymes associated with DNA replication have been characterized outside our laboratory:

- The topoisomerase IB (R194) was characterized by Benarroch et al. (2006). The mimivirus enzyme was functionally more similar to the poxvirus enzyme than to its bacterial homolog, despite its greater sequence similarity to the latter.
- Instead of the ATP-dependent DNA ligase that is present in most NCLDVs, Mimivirus has a NAD^+-dependent DNA ligase (R303), which is found in bacteria and entomopoxvirus. Benarroch and Shuman (2006) validated the predicted function of the gene product, but found significant differences in its enzymatic behavior compared with both the bacterial and the entomopoxvirus enzymes. They proposed that the Mimivirus enzyme is an intermediate evolutionary stage between the bacterial and entomopoxvirus form of the NAD^+-dependent DNA ligase, suggesting a horizontal transfer in an ancestral ameba host.

Characterization of Other Mimivirus Gene Products

ORF L276 is predicted to encode the first viral mitochondrial substrate carrier. In eukaryotic organisms, these proteins are located in the inner mitochondrial membrane or are integral to the membrane of other eukaryotic organelles. Monné et al. (2007) produced milligram quantities of the L276 gene product in *Lactococcus lactis* that were used for a detailed functional characterization. The protein transports dATP and dTTP, suggesting that Mimivirus might target its host mitochondria for obtaining the necessary deoxynucleotide required for replication of its A+T-rich genome.

We successfully produced the R355 gene product, predicted to encode a polyprotein protease potentially involved in the regulation of sumoylation. Crystals (diffracting up to 1.5-Å resolution) of the recombinant protein have been obtained.

The L678 gene product (homologous with a histone methyltransferase) has also been produced, but not in a soluble form. Finally, the product of the L222 gene, a member of a family of 12 close Mimivirus paralogs, has been produced in a soluble form. This protein, of unknown function, is predicted to interact with RNA.

Finally, the predicted DNA glycosylase activities (Endonuclease VIII) of ORF L315 and ORF L720 gene products were experimentally characterized (Bandaru et al. 2007).

Mimivirus in the Tree of Life

Phylogenetic Analysis Using Components of the Translational Apparatus

Building the Tree of Life, in other words, reconstructing the phylogenetic relationships among all living organisms, is one of the fundamental challenges in biology. Numerous attempts to derive such a tree have been published (see Delsuc et al. 2005). They all involve a comparison of universal genes present in all organisms from the three domains Eubacteria, Eukarya and Archaea. Besides the two major subunits of the DNA-directed RNA polymerase, these universal genes all belong to the protein translation apparatus including ribosomal proteins (~20), a handful of transcription factors, and the aminoacyl-tRNA synthetases (~20). Given the absence of virus-encoded translational machinery, as well as the nonuniversality of virus-encoded RNA polymerases, DNA viruses were always excluded from Tree of Life constructions. However, the numerous components of the translational machinery coded by Mimivirus allow, for the first time, a DNA virus to be included in a Tree of Life analysis. Including the same set of genes used for cellular organisms a phylogenetic analysis indicates that Mimivirus branches near the origin of the Eukarya domain (Fig. 8).

Considering that the central position that Mimivirus occupies among the NCLDVs (Fig. 5), it is tempting to speculate that the ancestor of these large DNA

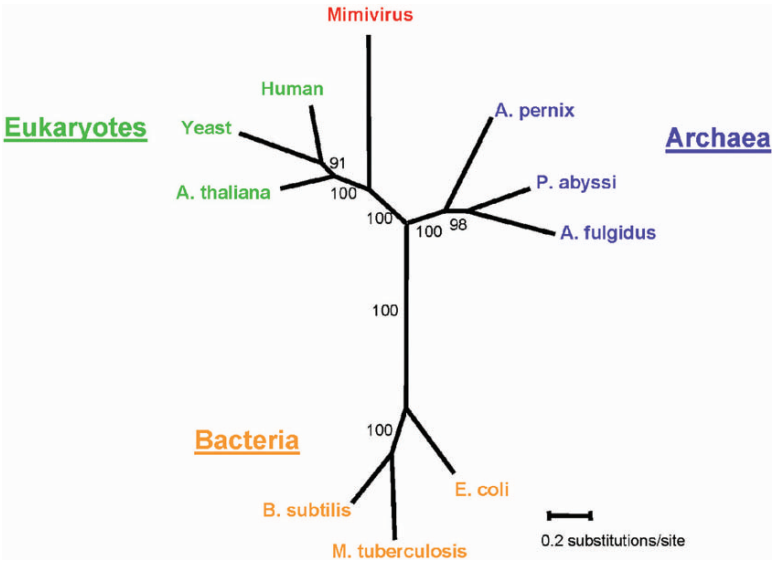

Fig. 8 Mimivirus in the Tree of Life. This tree was built using the concatenated sequences of seven universally conserved protein sequences (ArgRS, MetRS, TyrRS, the two RNA polymerase subunits, PCNA and a 5'-3' exonuclease. From Raoult et al. (2004)

viruses infecting eukaryotes could predate the radiation of the various eukaryotic kingdoms, or even the emergence of eukaryotes. Such a proposition, highly controversial at the time it was first proposed (Raoult et al. 2004), has since gained some acceptance since it fits nicely with previously published hypotheses linking primitive viruses to the origin of the eukaryotic nucleus (Villarreal and DeFilippis 2000; Takemura 2001; Bell 2001). An alternative hypothesis taking into account the reversibility and flexibility of the gene flow between ancestral viruses and the genome of primitive eukaryotes was more recently proposed (Claverie 2006). According to this scenario, ancestral viruses and pro-nucleus might have exchanged their roles iteratively during the predarwinian era, accounting for the diversity of extant DNA virus families and their partial monophyletic character.

Phylogenetic Analysis Using Clamp Loaders

Despite its high bootstrap values, a common criticism of the phylogenetic tree shown in Fig. 8 is the long branch connecting Mimivirus to the tree trunk, indicating that the sequences used to build the tree are quite divergent. We identified the clamp loader (replication factor C subunits) proteins as an alternative set of sequences exhibiting minimal divergence across the three domains of life, and present in a few DNA viruses, including Mimivirus. Clamp loaders use ATP hydrolysis to load the ring-shaped sliding-clamp made of PCNA subunits around the DNA molecule at the time of replication, promoting processivity. Clamp loader homologs remain remarkably similar in sequence (>25% identity over more than 250 residues) across the three domains of life. Mimivirus encodes its own PCNA molecule, and is again unique among viruses in possessing four clamp loader small subunits (R395, L499, R510, L478) and one large subunit (R411), as found in cellular eukaryotes. In contrast, the archaeal functional homolog (from which the eukaryote clamp loader is thought to have evolved) is usually composed of one small subunit and one large subunit (with the exception of *Methanosarcina acetivorans* that has two similar small subunits and one large subunit) (Chen et al. 2005). Eubacterial clamp loaders are made of two different small subunits and one large subunit (Majka and Burgers 2004). Robust phylogenetic trees encompassing the three domains of life can be made from the multiple alignment of one Mimivirus clamp loader paralogs with its most similar homologs (reciprocal best match) in cellular organisms, as shown in Fig. 9.

In this reconstruction using the small subunit R395 clamp loader protein sequence, Mimivirus is positioned near its fellow NCLDV *Ectocarpus siliculosus* virus (EsV), both of them at the very root of the branch leading to all eukaryotes. Note that the Mimivirus and EsV sequences are positioned in between the Eubacteria and Prokarya domains. Similarly, the phage SPM2 sequence is positioned in between the Archea and Eubacteria domains. Taken at face value, these positions definitely suggest that ancestors of these DNA viruses were present at the time the three major forms of cellular organisms were individualized and are consistent with the hypothesis that they provided the DNA biochemistry and the necessary replication machinery to emerging cellular microorganisms (Forterre 2006; Claverie 2006).

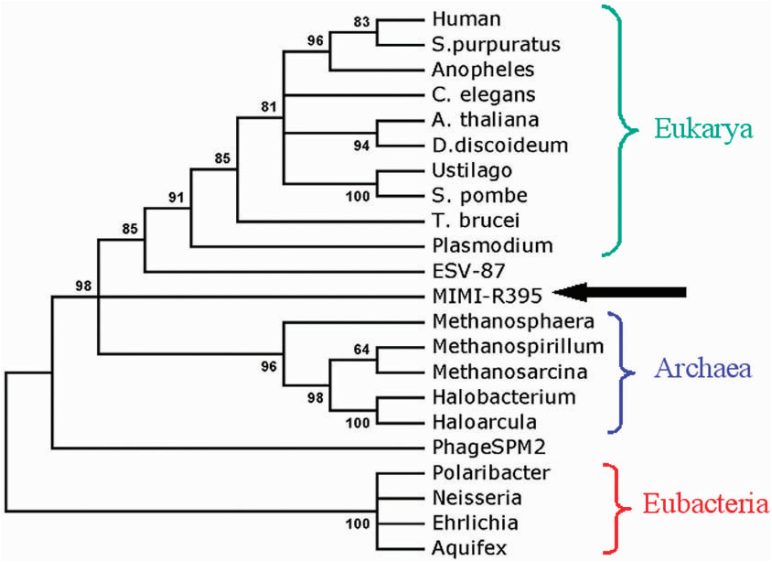

Fig. 9 Mimivirus in the Tree of Life. The tree was built using an alignment (145 positions retained) of the sequence of the Mimivirus clamp loader protein (R395) with its best reciprocal homologs in the indicated species. The server at www.phylogeny.fr was used with defaults parameters (rooting at midpoint)

Evidence of Lateral Gene Transfer Between Mimivirus and Corals

As noted above, a large fraction of the genes now constituting the Mimivirus genome have an ancestry predating the emergence of eukaryotes. However, it is also clear that several genes found their way into the Mimivirus genome through horizontal transfers, probably facilitated by the concentration of bacteria or other viruses within the ameba host during its normal feeding process. Together with the metagenomic analyses, identifying the putative origin of the laterally transferred genes might provide hints on ecological niches to look for additional Mimiviridae members. The MutS (mismatch repair) gene provides an intriguing example.

Mimivirus is the sole virus known to have a MutS family protein gene (L359). MutS proteins function in DNA mismatch repair and recombination. These enzymes are ubiquitous in bacteria and eukaryotes and are also found in several Archaea (e.g., *Halobacterium* spp. and *Methanosarcina mazei*). Eukaryotes have at least six major paralogous groups of MutS proteins (MSH1 to MSH6/7), as well as an additional isolated paralogous group encoded in the gorgonian coral mitochondrial DNA (mtMSH). Gorgonians are the sole known eukaryotes exhibiting mitochondrial DNA-encoded mtMutS family proteins (Pont-Kingdon et al. 1998). Interestingly, the Mimivirus MutS sequence is clearly related to the gorgonian mtMSH (Fig. 10). Furthermore, *Sulfurimonas denitrificans* (formerly *Thiomicrospira denitrificans*),

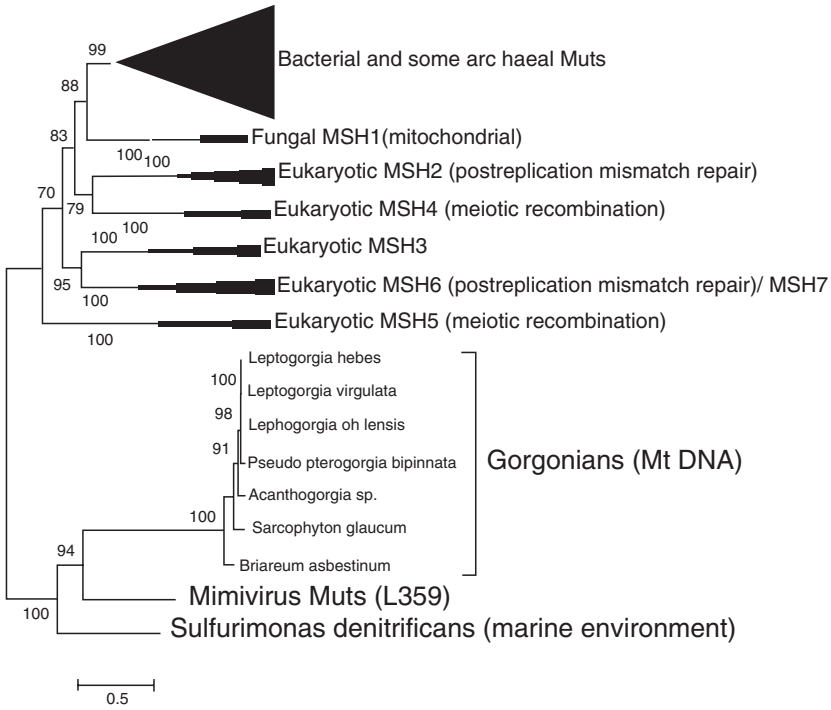

Fig. 10 Phylogenetic tree of the MutS family proteins. The tree was built using the PhyML server (www.phylogeny.fr) with the option of T-coffee for the alignment program. Bootstrap percentages above 70% are indicated

the marine environmental epsilonproteobacterium, has a MutS homolog related to the Mimivirus MutS and the mtMSH. These results strongly suggest that horizontal transfers of the MutS protein genes occurred (eventually within an ameba host) between an ancestor of Mimivirus, corals and environmental epsilonproteobacteria. Together with the identification of Mimivirus relatives in the Sargasso Sea data set, this last result is one more incentive to look for new *Mimiviridae* species in association with marine protists.

Acknowledgements We thank Jim Van Etten for his correction of the original version of this chapter, as well as Dr. Wai Li Ling and Dr. Jorge Navaza Jorge for allowing us to present micrographs of the stargate structure (Fig. 6) prior to publication.

References

Abergel C, Chenivesse S, Byrne D, Suhre K, Arondel V, Claverie JM (2005) Mimivirus TyrRS: preliminary structural and functional characterization of the first amino-acyl tRNA synthetase found in a virus. Acta Crystallograph Sect F Struct Biol Cryst Commun. 61:212–215

Abergel C, Rudinger-Thirion J, Giege R, Claverie JM (2007) Virus-encoded aminoacyl-tRNA synthetases: structural and functional characterization of Mimivirus TyrRS and MetRS. J Virol81:12406–12417

Bailey TL, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14:48–54

Bandaru V, Zhao X, Newton MR, Burrows CJ, Wallace SS (2007) Human endonuclease VIII-like (NEIL) proteins in the giant DNA Mimivirus. DNA Repair 6:1629–1641

Bawden AL, Glassberg KJ, Diggans J, Shaw R, Farmerie W, Moyer RW (2000) Complete genomic sequence of the *Amsacta moorei* entomopoxvirus: analysis and comparison with other poxviruses. Virology 274:120–139

Bell PJ (2001) Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? J Mol Evol 53:251–256

Benarroch D, Shuman S (2006) Characterization of mimivirus NAD+-dependent DNA ligase. Virology 353:133–143

Benarroch D, Claverie JM, Raoult D, Shuman S (2006) Characterization of mimivirus DNA topoisomerase Ib suggests horizontal gene transfer between eukaryal viruses and bacteria. J Virol 80: 314–321

Berger P, Papazian L, Drancourt M, La Scola B, Auffray JP, Raoult D (2006) Ameba-associated microorganisms and diagnosis of nosocomial pneumonia. Emerg Infect Dis 12:248–255

Blanc G, Ogata H, Robert C, Audic S, Suhre K, Vestris G, Claverie JM, Raoult D (2007) Reductive genome evolution from the mother of rickettsia. PLoS Genet 3:e14

Chen YH, Kocherginskaya SA, Lin Y, Sriratana B, Lagunas AM, Robbins JB, Mackie RI, Cann IK (2005) Biochemical and mutational analyses of a unique clamp loader complex in the archaeon *Methanosarcina acetivorans*. J Biol Chem 280:41852–1863

Cheng CH, Liu SM, Chow TY, Hsiao YY, Wang DP, Huang JJ, Chen HH (2002) Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the Baculoviridae. J Virol 76:9024–9034

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

Claverie JM, Ogata H, Audic S, Abergel C, Suhre K, Fournier PE (2006) Mimivirus and the emerging concept of "giant" virus. Virus Res 117:133–144

Dare RK, Chittaganpitch M, Erdman DD (2008) Screening pneumonia patients for mimivirus. Emerg Infect Dis 14:465–467

Delaroque N, Boland W, Muller DG, Knippers R (2003) Comparisons of two large phaeoviral genomes and evolutionary implications. J Mol Evol 57:613–622

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 6:361–375

Espagne E, Dupuy C, Huguet E, Cattolico L, Provost B, Martins N, Poirie M, Periquet G, Drezen JM (2004) Genome sequence of a polydnavirus: insights into symbiotic viral evolution. Science 306:286–289

Filee J, Siguier P, Chandler M (2007) I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. Trends Genet 23:10–15

Fitzgerald LA, Graves MV, Li X, Feldblyum T, Nierman WC, Van Etten JL (2007) Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect Chlorella NC64A. Virology 358:472–484

Forterre P (2006) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc Natl Acad Sci U S A 103:3669–3674

Ghedin E, Claverie JM (2005) Mimivirus relatives in the Sargasso sea. Virol J 2:62

Iyer LM, Aravind L, Koonin EV (2001) Common origin of four diverse families of large eukaryotic DNA viruses. J Virol 75:11720–11734

Iyer LM, Balaji S, Koonin EV, Aravind L (2006) Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res 117:156–184

Jeudy S, Coutard B, Lebrun R, Abergel C (2005) *Acanthamoeba polyphaga* mimivirus NDK: preliminary crystallographic analysis of the first viral nucleoside diphosphate kinase. Acta Crystallograph Sect F Struct Biol Cryst Commun. 61:569–572

Jeudy S, Claverie JM, Abergel C (2006) The nucleoside diphosphate kinase from mimivirus: a peculiar affinity for deoxypyrimidine nucleotides. J Bioenerg Biomembr 38:247–254

Khan M, La Scola B, Lepidi H, Raoult D (2006) Pneumonia in mice inoculated experimentally with *Acanthamoeba polyphaga mimivirus*. Microb Pathog 42:56–61

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

La Scola B, Marrie TJ, Auffray JP, Raoult D (2005) *Mimivirus* in pneumonia patients. Emerg Inf Dis 11:449–452

Larcher C, Jeller V, Fischer H, Huemer HP (2006) Prevalence of respiratory viruses, including newly identified viruses, in hospitalised children in Austria. Eur J Clin Microbio Infect Dis 25:681–886

Majka J, Burgers PM (2004) The PCNA-RFC families of DNA clamps and clamp loaders. Prog Nucleic Acid Res Mol Biol 78:227–260

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

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

Monier A, Larsen JB, Sandaa RA, Bratbak G, Claverie JM, Ogata H (2008b) Marine mimivirus relatives are probably large algal viruses. Virol J 5:12

Monne M, Robinson AJ, Boes C, Harbour ME, Fearnley IM, Kunji ER (2007) The mimivirus genome encodes a mitochondrial carrier that transports dATP and dTTP. J Virol 81:3181–3186

Moreira D, Lopez-Garcia P (2005) Comment on "The 1.2-megabase genome sequence of Mimivirus". Science 308:1114

Nagasaki K, Shirai Y, Tomaru Y, Nishida K, Pietrokovski S (2005) Algal viruses with distinct intraspecies host specificities include identical intein elements. Appl Environ Microbiol 71:3599–3607

Nalçacioglu R, Ince IA, Vlak JM, Demirbag Z, van Oers MM (2007) The Chilo iridescent virus DNA polymerase promoter contains an essential AAAAT motif. J Gen Virol 88:2488–2494

Ogata H, Abergel C, Raoult D, Claverie J-M (2005a) Response to Comment on "The 1.2-megabase genome sequence of Mimivirus". Science 308:1114

Ogata H, Raoult D, Claverie JM (2005b) A new example of viral intein in Mimivirus. Virol J 2:8

Pont-Kingdon G, Okada NA, Macfarlane JL, Beagley CT, Watkins-Sims CD, Cavalier-Smith T, Clark-Walker GD, Wolstenholme DR (1998) Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. J Mol Evol 46:419–431

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

Raoult D, Renesto P, Brouqui P (2006) Laboratory infection of a technician by mimivirus. Ann Intern Med 144:702–703

Raoult D, La Scola B, Birtles R (2007) The discovery and characterization of Mimivirus, the largest known virus and putative pneumonia agent. Clin Infect Dis 45:95–102

Resch W, Hixson KK, Moore RJ, Lipton MS, Moss B (2007) Protein composition of the vaccinia virus mature virion. Virology 358:233–247

Renesto P, Abergel C, Decloquement P, Moinier D, Azza S, Ogata H, Fourquet P, Gorvel JP, Claverie JM (2006) Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. J Virol 80:11678–11685

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S et al (2007) The *Sorcerer II* Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol 5:e77

Suhre K (2005) Gene and genome duplication in Acanthamoeba polyphaga Mimivirus. J Virol 79:14095–14101

Suhre K, Audic S, Claverie JM (2005) Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. Proc Natl Acad Sci U S A 102:14689–14693

Suzan-Monti M, La Scola B, Raoult D (2006) Genomic and evolutionary aspects of Mimivirus. Virus Res 117:145–150

Suzan-Monti M, La Scola B, Barrassi L, Espinosa L, Raoult D (2007) Ultrastructural characterization of giant volcano-like virus factory of *A. polyphaga* Mimivirus. PloS ONE 3: e328

Takemura M (2001) Poxviruses and the origin of the eukaryotic nucleus. J Mol Evol 52:419–425

Vanacova S, Liston DR, Tachezy J, Johnson PJ (2003) Molecular biology of the amitochondriate parasites, *Giardia intestinalis*, *Entamoeba histolytica* and *Trichomonas vaginalis*. Int J Parasitol 33:235–255

Van Etten JL (2003) Unusual life style of giant chlorella viruses. Annu Rev Genet 37:153–195

Venter JC. Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66–74

Villarreal LP, DeFilippis VR (2000) A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. J Virol 74:7079–7084

Weekers PH, Engelberts AM, Vogels GD (1995) Bacteriolytic activities of the free-living soil amoebae, *Acanthamoeba castellanii*, *Acanthamoeba polyphaga* and *Hartmannella vermiformis*. Antonie Van Leeuwenhoek 68:237–243

Wilson WH, Schroeder DC, Allen MJ, Holden MT, Parkhill J, Barrell BG, Churcher C, Hamlin N, Mungall K, Norbertczak H, Quail MA, Price C, Rabbinowitsch E, Walker D, Craigon M, Roy D, Ghazal P (2005) Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. Science 309:1090–1092

Xiao C, Chipman PR, Battisti AJ, Bowman VD, Renesto P, Raoult D, Rossmann MG (2005) Cryo-electron microscopy of the giant Mimivirus. J Mol Biol 353:493–496

Yamada T, Onimatsu H, Van Etten JL (2006) Chlorella viruses. Adv Virus Res 66:293–336

Yoder JD, Chen TS, Gagnier CR, Vemulapalli S, Maier CS, Hruby DE (2006) Pox proteomics: mass spectrometry analysis and identification of Vaccinia virion proteins. Virol J 3:10

Zachertowska A, Brewer D, Evans DH (2005) Characterization of the major capsid proteins of myxoma virus particles using MALDI-TOF mass spectrometry. J Virol Methods 132:1–12

Zhang QY, Xiao F, Xie J, Li ZQ, Gui JF (2004) Complete genome sequence of lymphocystis disease virus isolated from China. J Virol 78:6982–6994