Available online at www.sciencedirect.com
SCIENCE DIRECT
Virus Research 117 (2006) 133–144
Virus Research
www.elsevier.com/locate/virusres

Mimivirus and the emerging concept of "giant" virus

Jean-Michel Claverie [a,b,*], Hiroyuki Ogata [a], Stéphane Audic [a], Chantal Abergel [a], Karsten Suhre [a], Pierre-Edouard Fournier [a,b]

[a] *Information Génomique et Structurale, CNRS UPR 2589, IBSM, Parc Scientifique de Luminy, 163 Avenue de Luminy, Case 934, 13288 Marseille Cedex 9, France*
[b] *Faculté de Médecine, Université de la Méditerranée, 27 Blvd. Jean Moulin, 13385 Marseille Cedex 5, France*

Available online 15 February 2006

Abstract

The recently discovered *Acanthamoeba polyphaga* Mimivirus is the largest known DNA virus. Its particle size (750 nm), genome length (1.2 million bp) and large gene repertoire (911 protein coding genes) blur the established boundaries between viruses and parasitic cellular organisms. In addition, the analysis of its genome sequence identified many types of genes never before encountered in a virus, including aminoacyl-tRNA synthetases and other central components of the translation machinery previously thought to be the signature of cellular organisms. In this article, we examine how the finding of such a giant virus might durably influence the way we look at microbial biodiversity, and lead us to revise the classification of microbial domains and life forms. We propose to introduce the word "girus" to recognize the intermediate status of these giant DNA viruses, the genome complexity of which makes them closer to small parasitic prokaryotes than to regular viruses.

Keywords: Large DNA viruses; Mimivirus; Evolution; Genome

1. Introduction

The discovery of *Acanthamoeba polyphaga* Mimivirus (La Scola et al., 2003) and the analysis of its complete genome sequence (Raoult et al., 2004) sent a shock wave through the community of virologists and evolutionists. The size, gene content, and phylogenetic characterization of the virus genome, challenged many accepted ideas about what virus should look like, and where they might come from. Several comments have already been published on Mimivirus (Ghedin and Fraser, 2005; Desjardins et al., 2005; Koonin, 2005; Galperin, 2005; Moreira and Lopez-Garcia, 2005; Ogata et al., 2005a, 2005b). However, the consequences of the qualitative and quantitative gaps separating it from previously known DNA viruses are yet to be analyzed in depth.

Very large DNA virus genomes have accumulated steadily in the databases, since the spectacular achievement of Barrell's team sequencing the 230 kb of human cytomegalovirus (Human herpesvirus 5) as early as 1990 (Chee et al., 1990). Curiously, however, these incremental progresses failed to generate much emotion or trigger significant changes in the perception/notion of virus that prevails in the general community of biologists. In our collective subconscious mind, viruses are still thought of as highly optimized minimal "bags of genes", packaging just enough information to deal with host infection and to highjack the host machinery for multiplying tiny viral particles. Given the simplicity of a minimal particle (a capsid protein and a few more proteins for genome packaging), a viral genome is thus expected to carry less than a dozen of genes. In this context, a virus (or a phage) packing more than 200 genes already appears as an evolutionary absurdity, an "overkill".

Thus, if Mimivirus deserved some special attention, it is not primarily because it was larger than the previously largest virus before it. It was because it is the first virus the various dimensions of which (particle size and genome complexity) are significantly overlapping with those typical of parasitic cellular microorganisms (Table 1). This unique feature of Mimivirus invalidates the traditional size/complexity criteria, and prompt us to re-examine the fundamental question: what is a virus?

This question is not only philosophical, or related to speculative thinking about the origin of life. It has very practical implications. Particle size, for instance, was always central to virus isolation protocols, and still directly pertains to the design of the modern "metagenomic" studies aiming at assessing

* Corresponding author. Tel.: +33 4 91 82 54 20; fax: +33 4 91 82 54 21.
E-mail address: Jean-Michel.Claverie@igs.cnrs-mrs.fr (J.-M. Claverie).

0168-1702/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2006.01.008

Table 1
Top-20 virus genome sizes

Name	Genome size (kb)	Family	NCBI reference
Acanthamoeba Polyphaga Mimivirus	1181.4	*Mimiviridae*	NC_006450
Bacillus phage G	497.5	*Myoviridae*	(1)
Coccolithovirus EhV-86	407.3	*Phycodnaviridae*	NC_007346
Paramecium bursaria Chlorella virus NY-2A	368.7	*Phycodnaviridae*	(2)
Canarypox virus	359.9	*Chordopoxvirinae*	NC_005309
Paramecium bursaria Chlorella virus AR158	344.7	*Phycodnaviridae*	(2)
Ectocarpus siliculosus virus	335.6	*Phycodnaviridae*	NC_002687
Paramecium bursaria Chlorella virus 1	330.7	*Phycodnaviridae*	NC_000852
Paramecium bursaria Chlorella virus MT325	314.3	*Phycodnaviridae*	(2)
Shrimp white spot syndrome virus	305.1	*Nimaviridae*	NC_003225
Fowlpox virus	288.5	*Chordopoxvirus*	NC_002188
Pseudomonas phage phiKZ	280.3	*Myoviridae*	NC_004629
Bacteriophage KVP40	244.8	*Myoviridae*	NC_005083
Pongine herpesvirus 4	241.1	*Herpesviridae*	NC_003521
Melanoplus sanguinipes entomopoxvirus	236.1	*Entomopoxvirus*	NC_001993
Human herpesvirus 5 strain Merlin	235.6	*Herpesviridae*	NC_006273
Bacteriophage Aehl	233.2	*Myoviridae*	NC_005260
Amsacta moorei entomopoxvirus	232.4	*Entomopoxvirus*	NC_002520
Human herpesvirus 5 strain AD169	230.3	*Herpesviridae*	NC_001347
Murid herpesvirus 1	230.3	*Herpesviridae*	NC_004065

The "giant" viruses corresponding to the discontinuity in the distribution of genome sizes (Fig. 1A) are in gray. (1) Available at URL: pbi.bio.pitt.edu/; (2) available at URL: http://greengene.uml.edu/.

microbial biodiversity. Simply acknowledging the fact that all viruses might not be filterable through the typical "sterilizing" 0.2–0.3 μm-pore filters, already changes our interpretation of the currently available data, and call for significant changes to the protocols of future environmental sampling campaigns. Given the much higher relative abundance of viruses (including phages) over cellular organisms in most natural environments, it is not unlikely that a significant fraction of the already randomly sampled DNA sequences might originate from unknown large viruses. In this context, it is worth to notice that 15% of Mimivirus genes were found to exhibit their closest matches to environmental sequences (Ghedin and Claverie, 2005) despite the fact that the *Mimiviridae* is only distantly related to previously described virus families (Raoult et al., 2004).

The first part of this review article will focus on the genome size distribution of DNA viruses and propose that the largest of them might constitute a new type of microbial organisms, subject to their own, yet unknown, peculiar evolutionary constraints.

The interpretation of metagenomic data is also seriously challenged by the fact that there is now a complete overlap between the particle sizes, genome complexities, and the type of genes found in the largest DNA viruses and the smallest bacterial, archaeal or even eukaryotic organisms (Ward and Fraser, 2005). As we found more and more viral genes unrelated to the basic functions of DNA replication and particle synthesis (such as any kind of enzymes, components of signaling pathways, tRNAs, transcription and translation factors, etc.), what rigorous criteria remains at our disposal to reliably distinguish, using their sequences, viral genes from those belonging to the genome of a cellular organism? This question is addressed in the second part of this article.

2. Results and discussion

2.1. Giant viruses: a discontinuity in the distribution of virus genome sizes

As of January 6, 2006, 416 double stranded DNA virus complete genome sequences were available at the National Center for Biotechnology Information. This data is unevenly distributed, with a few virus clades accounting for a large proportion of the known genomes such the *Caudovirales* ("tailed" bacteriophages, 206 sequences), and 5 major animal-infecting virus families: *Papillomaviridae* (51), *Herpesviridae* (42), *Baculoviridae* (28), *Adenoviridae* (23), and *Poxviridae* (22). On the other hand, many families have less than a handful of representatives. Four of these under-represented families are associated to very large genomes, namely the *Mimiviridae* (*A. polyphaga* Mimivirus 1.18 Mb), the *Nimaviridae* (Shrimp white spot syndrome virus, 305 kb), the *Phycodnaviridae* (*Emiliania huxleyi* virus 86, 407 kb, *Paramecium bursaria* Chlorella virus 1, 331 kb, *Ectocarpus siliculosus* virus, 336 kb), and the *Polydnaviridae* (*Cotesia congregata* bracovirus, 567 kb, not discussed further in this paper because of its atypical very low coding density) (Espagne et al., 2004). Table 1 lists the top-20 largest viral genomes (publicly available as complete sequences). An immediate conclusion can be drawn from this simple list: it is that large-sized genomes are not specifically associated to a given virus family, host type, vector, or ecological niche. A diverse assortment of viruses infecting bacteria, invertebrate, vertebrate, algae, or amoeba is found among the top-sized genomes.

This lack of correlation prompted us to use an objective data-mining technique (Bougueleret et al., 1988) to search for remarkable features in the genome size distributions. A

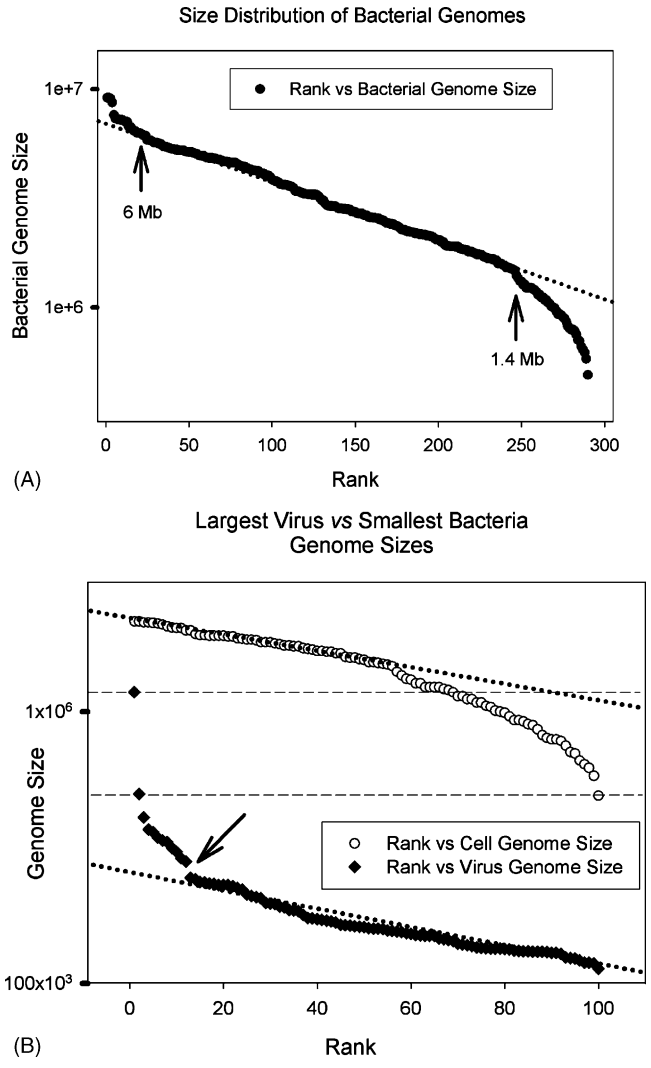

Fig. 1. Rank vs. size plots of the genome size distribution. (A) Complete bacterial genomes available from the National Center for Biotechnology Information (NIH, Bethesda, USA) as of January 2006. A logarithmic scale is used on the genome size axis. The arrows indicate two discontinuity at the high and low ends of genome sizes. The marked slope inflexion at 1.4 Mb coincides with obligate parasitic/symbiotic bacteria. (B) Top 100 largest viral genomes compared to the 100 smallest bacteria. The viral genome size distribution (black lozanges) exhibits a gap and a marked slope change for sizes above 280 kb (indicated by an arrow). The top ranked Mimivirus genome (1.2 Mb) allows viral genome sizes to overlap (as indicated by the dashed lines) with most obligate parasitic/symbotic bacteria (white circles) as listed in Table 2.

traditional heuristic approach is to use a two-dimensional data representation such as shown in Fig. 1. Here, the sizes of all sequenced bacterial genomes are plotted (using a logarithmic scale) against their rank values in the distribution (the largest genome being ranked 1, the second largest being ranked 2, etc.). It is expected that properties computed on a homogeneous category of "objects" (for instance microorganisms obeying similar evolutionary constraints) would be smoothly distributed in such a graph. Fig. 1A clearly shows this not to be the case. From approximately rank 25–250, the logarithm of the genome sizes follow a slowly decreasing quasi-linear distribution, encompassing bacterial genomes ranging from 6 to 1.4 Mb. This might be defined as the range of "regular bacteria". Below 1.4 Mb (approximately) the slope becomes much steeper, corresponding to anomalous bacteria exhibiting too small a genome for their rank in the distribution. Interestingly, all bacteria in that genome size range are obligate parasites (with the exception of the free-living marine bacterium *Pelagibacter ubique*, 1.31 Mb) (Rappe et al., 2002), and many of them intracellular. This sharp inflexion of the distribution suggests that bacteria having reached the minimal free-living genome size of 1.4 Mb, are forced into a parasitic life-style, a drastic change in the evolution pressure that allows the further shrinking of genome sizes. On the other end of the distribution, rank 1–25 encompasses "supersized" environmental bacteria, with the anomalous behavior of exhibiting a bigger genome than extrapolated from their rank. This might be interpreted as indicating that bacteria reaching genome size of the order of 6 Mb are pressured into acquiring even more genes, perhaps reflecting the larger proportion of genes devoted to the regulation of the expression of larger genomes (Konstantinidis and Tiedje, 2004). In conclusion, this application of the log (size) versus rank plot to the cellular prokaryote genomes indicates: (i) that the discontinuity exhibited in Fig. 1A is not generated by the mathematical representation; (ii) that such a discontinuity seems to coincide with the boundary between organisms subject to broadly different evolutionary constraints (or isolation protocols). It is now interesting to use the same approach to analyse the distribution of the genome sizes among ds-DNA viruses. As for the bacteria, the distribution of genome sizes (using a logarithmic scale) exhibits a smoothly decreasing, approximately linear region from ranks 12 to 100 (and below, not shown) corresponding to sizes of 244–110 kb (Fig. 1B). This part of the curve corresponds to viruses of various interspersed families (mostly bacteriophages, baculoviruses, herpesviruses, and poxviruses, see URL: www.giantvirus.org) the genome sizes (in a range of 110–244 kb) of which appear to results from similar evolutionary constraints. This region of the distribution might be defined as the one encompassing the "regular" large ds-DNA viruses. However, the 12 largest viruses (rank 1–12) are totally off scale, corresponding to much bigger sizes and size increments that is extrapolated from the rest of the distribution, predicting that the largest ds-DNA virus genome should be less than 270 kb. Despite the logarithmic scale, the size distribution of the top 12 ranking viruses is separated from the previous ones by a visible gap also coinciding with a large increase of the slope. With the caveat expressed below, we propose that such an abrupt change in the distribution might define a new type of "giant viruses", corresponding to genome sizes of 280 kb and larger. However, within this group of giant viruses, Mimivirus still appears to stand out (Fig. 1B).

2.2. *Giant viruses encompass different families*

Interestingly, the select club of the giant viruses is no less diverse than the rest of the large ds-DNA viruses. Among its 12 members (see Table 1) one finds 2 poxviruses, 2 bacteriophages, 6 phycodnaviruses, the sole known Nimaviridae, and Mimivirus. This already suggests that, as for regular large ds-DNA viruses, genomic gigantism is not restricted to a specific host, a phylogenetic clade, or a given environmental niche. Various hypotheses

Table 2
Genome sizes of the largest viruses and of the smallest cellular organisms

Species name	Genome size (bp)	Domain	NCBI#
P. bursaria Chlorella virus NY-2A	368,683	dsDNA virus	(2)
Coccolithovirus EhV-86	407,339	dsDNA virus	NC_007346
Nanoarchaeum equitans	490,885	Archaea	NC_005213
Bacillus phage G	497,513	Phage	(1)
Guillardia theta nucleomorph	551,265	Eukaryota	NC_002751–53
Mycoplasma genitalium	580,074	Bacteria	NC_000908
Buchnera aphidicola str. Bp	615,980	Bacteria	NC_004545
Buchnera aphidicola str. Sg	641,454	Bacteria	NC_004061
Wigglesworthia glossinidia	697,724	Bacteria	NC_004344
Candidatus Blochmannia	705,557	Bacteria	NC_005061
Ureaplasma parvum	751,719	Bacteria	NC_002162
Mycoplasma mobile	777,079	Bacteria	NC_006908
Mesoplasma florum	793,224	Bacteria	NC_006055
Mycoplasma synoviae	799,476	Bacteria	NC_007294
Mycoplasma pneumoniae	816,394	Bacteria	NC_000912
Onion yellows phytoplasma	860,631	Bacteria	NC_005303
Mycoplasma hyopneumoniae	892,758	Bacteria	NC_006360
Borrelia garinii	904,246	Bacteria	NC_006156
Tropheryma whipplei	927,303	Bacteria	NC_004572
Mycoplasma pulmonis	963,879	Bacteria	NC_002771
Mycoplasma gallisepticum	996,422	Bacteria	NC_004829
Mycoplasma capricolum	1,010,023	Bacteria	NC_007633
Chlamydia trachomatis D/UW-3/CX	1,042,519	Bacteria	NC_000117
Chlamydia muridarum	1,072,950	Bacteria	NC_002620
Wolbachia endosymbiont strain TRS	1,080,084	Bacteria	NC_006833
Rickettsia prowazekii	1,111,523	Bacteria	NC_000963
Rickettsia typhi	1,111,496	Bacteria	NC_006142
Treponema pallidum	1,138,011	Bacteria	NC_000919
Chlamydophila abortus	1,144,377	Bacteria	NC_004552
Chlamydophila caviae	1,173,390	Bacteria	NC_003361
A. polyphaga Mimivirus	1,181,404	ds-DNA virus	NC_006450

Mimivirus genome is larger than the one of more than 30 cellular organisms (a number of nearly identical bacterial strains are not listed). These organisms are obligate parasites or symbionts, with the exception of *T. whipplei* for which axenic growth conditions have been described (Renesto et al., 2003). Only publicly available genome sequences are listed. The size indicated for Bacillus page G refers to the unique part of the genome sequence. The total packaged DNA is about 650 kb long. The coding density is about of one gene/1 kb for all the above organisms (ORF > 300 bp). The polydnavirus Cotesia congretata virus (568 kb) is not included because of its atypical low gene content. (1) Available at URL: pbi.bio.pitt.edu/; (2) available at URL: http://greengene.uml.edu/.

can be proposed that might account for the observed discontinuity in the genome size distribution. One may propose, for instance, that en block genome duplication might have occurred independently in the various families resulting in sporadic supersized members. Followed by the rapid functional diversification of the newly created paralogues, such an event might become evolutionarily advantageous, for instance by extending the host range (Mesyanzhinov et al., 2002). Despite a subsequent downsizing by gene loss, the genome size distribution might still reflect such discontinuous events. The discontinuity in the virus genome size distribution might also result from a trivial experimental bias. The change in slope might be interpreted as an indication that giant viruses (i.e. over 280 kb in genome size), although common in nature, might have a lesser probability to be isolated. Indeed, it is striking that the genome size around which the discontinuity occurs corresponds to a typical particle size of 200 nm in diameter. Sterilizing (i.e. bacteria removing) filters with 200–300-nm diameter pore sizes might typically retain (or damage) viruses in this size range, eventually hindering their isolation or their serendipitous discovery. Such filters are invariably used in environmental sampling protocols designed to a priori separate cellular organisms from viruses and would result in giant viruses material (e.g. DNA) to be misinterpreted as originating from cellular organisms (see Section 2.5). In conclusion, if the broken shape of the genome size distribution is simply due to this trivial experimental bias, giant viruses might be much more frequent than is presently thought. Many more giant viruses (including phages) might be still hiding in the category of so-called "uncultivable" environmental bacteria. The phycodnaviridae remains a promising source of giant viruses with several newly isolates exhibiting genomes (not yet sequenced) in the 460-kb range (Baudoux and Brussaard, 2005).

2.3. How big can a giant virus be?

If Mimivirus and the other giant viruses are not as exceptional as suggested by the distribution in Fig. 1B, how big can a virus be? Are there natural limits to the particle and genome sizes of a virus, and what are they? In the case of Mimivirus, we already estimated that the central core region corresponded to a volume large enough to pack its DNA at a core concentration of about 450 mg/ml, a usual level encountered by other viruses (Raoult et

al., 2004). Independent evaluations have proposed lower values and suggested that Mimivirus DNA density is in the lower range when compared to bacteriophages (Abrescia et al., 2004). Hence there is no evidence that Mimivirus particle size corresponds to a limit imposed by evolutionary or biophysical constraints. A 3D reconstruction of the 190 nm-diameter PBCV-1 particle, the largest to date, fixed the precise number of major capsid protein molecules VP54 to 5040 (Nandhagopal et al., 2002). Mimivirus particle consists into a capsid with a diameter of 500 nm covered by 125 nm long, closely packed fibers (Raoult et al., 2004). Its icosahedral capsid exhibits a pseudo-triangulation number of approximately 1180 (Xiao et al., 2005) and then is made of approximately 24,000 molecules of the major capsid protein. However, there are no obvious biophysical rules that would preclude even larger particles to exist.

In contrast, the typical linear dimension R of cellular organisms (e.g. bacteria) has to remain within a range dictated by the existence of a internal metabolic activity. This activity, roughly proportional to the cell volume (and thus to R^3) must be sustained by a flux of nutrient and energy dissipation through the surface of the cell envelope the area of which is proportional to R^2.

This can be described by the qualitative equation: $\alpha R^3 \approx \beta R^2$, and thus forces R within the possible (μm) range of β/α (flux/metabolic activity) (Kooijman, 2001).

In absence of metabolism, viral particles have no such constraints, and their volume are simply required to grow in direct proportion of the DNA to be packed in. The linear dimension R (i.e. diameter) only needs to increase very slowly, proportionally to the third root of the genome size. Thus, if Mimivirus can fit 1.2 Mb in a 0.5 μm-diameter particle, 6 Mb will fit in a 0.85 μm-diameter particle, and 10 Mb viral genome will only require a 1 μm-diameter particle. Of course, one expect a virus to remain small compared to the dimension of the host it infects (<1/30 of its host diameter for *A. polyphaga* Mimivirus) and its genome size must remain in proportion to its host genome size (so that the host biosynthetic DNA machinery and nucleotide pool can suit the need of its replication). In that respect, bacteriophage G (genome size ∼650 kb, and 200 nm in diameter) (Serwer and Hayes, 2001) infecting a bacillus (with genome size of about 5 Mb, 2 μm in size) may represent an extreme case. Accordingly, large amoebal protists could easily accommodate μm-sized, 10 Mb-genome viruses, given their own enormous genome size (hundreds of Gbp) and cell dimension (150–4000 μm). Thus, we must admit that the data at our disposal – possibly plagued by an experimental bias – do not suggest any clear limits on the genome size and complexity of the giant viruses that remain to be discovered.

2.4. Mimicking a bacteria as a positively selected trait for amoebal virus?

Mimivirus name originates from its initial misidentification as a Gram-positive bacteria, hence as a "microbe mimicking" virus. Two key factors were responsible for this mistake: a particle size allowing the virus to be easily visible with a light microscope, and its mild Gram-coloration (La Scola et al., 2003). These two properties might actually be central to the virus physiology. Electron microscopy study of amoeba, *A. polyphaga*, being infected by Mimivirus, suggests that the virus is initially taken up via the feeding phagocytosis pathway normally used by the amoeba to feed on bacteria. Among a large variety of tested primary or established cell lines from vertebrates or invertebrate, only *Acanthamoeba* species were found to be productively infected by a cell-free viral suspension (Suzan-Monti et al., in press). It is known that the initial step of phagocytosis is more efficiently triggered by particle sizes in the μm range. Using latex beads Korn and Weisman (1967) have shown that a transition toward a less active phagocytic behavior occurs for particles less than 0.6 μm in diameter. Interestingly, this is close to Mimivirus particle size. The formation of endocytic vesicules is also activated by the presence of a polysaccharide envelope on the engulfed particle, as typical for bacteria. This is mediated through the binding of amoebal lectin-type receptors. Electron microscopy and antibody binding patterns (La Scola et al., 2005) strongly suggest that Mimivirus particles are actually encased in a 125-nm thick polysaccharidic layer, making it even more palatable for its amoebal host. This is consistent with the presence of many key sugar-manipulating enzymes encoded in the virus genome, some of them being quite specific of the biosynthesis of cell-surface lipopolysaccharide material such as perosamine (Raoult et al., 2004). The Gram staining of the virus is likely due to its LPS-like layer. This is also consistent with the extreme sturdiness of the particle as observed in our preliminary proteomics studies (Raoult et al., 2004). It is likely that Mimivirus is probably locked in this spore-like structure, and that the digestion of this LPS-like envelope by the amoeba endocytic vacuole is a prerequisite to the bona fide viral infection, that occurs through the vacuole membrane. The virus host specificity might be simply dictated by the presence or absence of the required enzymes in the phagosomes of various amoebal species (Weekers et al., 1995). In conclusion, the properties that led to the initial misinterpretation of Mimivirus as an amoeba-infecting bacteria, might actually be central to the life-style of many other giant viruses infecting their cellular host via the phagocytic route. If this is true, protozoans normally feeding on bacteria should be particularly scrutinized in search of new giant viruses.

2.5. Reassessing metagenomics: what is a virus sequence?

As a consequence of the decreasing cost of sequencing DNA, the study of microbial biodiversity has now entered the genomic era, with the introduction of "metagenomics", defined as the culture-independent genomic analysis of an assemblage of microorganisms. Initial environmental sequencing projects targeted at 16S ribosomal RNA (rRNA) offered a glimpse into the phylogenetic diversity of uncultured organisms (reviewed in Riesenfeld et al., 2004; Delong, 2005). The high-throughput shotgun sequencing of environmental DNA samples was then introduced to provide a more global views of those communities (Venter et al., 2004; Tringe et al., 2005). The metagenomic approach is now being specifically applied to the study of viral communities, using a gene-centric approach (Culley et al., 2003;

Short and Suttle, 2002) or shotgun sequencing (Breitbart et al., 2002, 2003, 2004).

The existence of giant viruses comparable to small bacteria in terms of particle size and genome complexity makes the interpretation of metagenomic shotgun sequences much less straightforward than previously thought, on two counts. First, filtering steps are invariably used to separate the "bacterial" fraction (using a 0.1–0.3 μm pore size range) from the "viral" fraction. In consequences, non-filtering giant viruses will contribute sequences misinterpreted as part of the bacterial pool, while they will be missing from the survey of viral communities. Given the tendency of algal viruses to be large (e.g. phycodnaviruses, Van Etten et al., 1991, 2002; Claverie, 2005) and the fact that viruses might outnumber bacteria by an order of magnitude in some aquatic environments (Wommack et al., 1992; Wommack and Colwell, 2000), the results from these ecological surveys should be interpreted with caution. For instance, an exhaustive similarity search (Ghedin and Claverie, 2005) of all Mimivirus predicted proteins against all publicly available sequences identified many of their closest homologues among the "bacterial" pool of the Sargasso Sea environmental sequences (Venter et al., 2004). More detailed phylogenetic analyses strongly suggested that these environmental sequences do belong to unknown large viruses evolutionarily closer to Mimivirus than to any presently characterized viral species (Ghedin and Claverie, 2005).

A second – more fundamental – problem is that, in absence of the simple "filtering criteria", distinguishing giant virus genes from those found in cellular organisms becomes very tricky, when solely based on sequence similarity and predicted gene function. A funny example is given by the presence of an acetylcholinesterase-like gene in *A. polyphaga* Mimivirus. ORF L906 is 1737-bp long, encoding a 579-residue putative protein. The best BLASTP (Altschul et al., 1997) matching homologue of Mimivirus L906 in the nr database is *Torpedo californica* acetylcholinesterase (GenBank accession number CAA27169) exhibiting 30.1% identity over 508 residues. No viral sequence exhibits any significant similarity with Mimivirus L906. Captured in a metagenomic sampling, such a sequence will no doubt be classified as originating from a cellular organism, if not from a contamination from fish DNA! Yet this sequence is part of Mimivirus genome. Incidentally, many L906 homologues are found in the Sargasso Sea environmental data set, most of them from bacterial origin (probably).

The multi-alignment of MIMI_L906 (using T-Coffee, Poirot et al., 2004) with a set of carboxylesterases 3D structures and typical acetylcholinesterase sequences (Fig. 2) shows the presence of the expected catalytic triad involving a serine, a glutamate (or aspartate) and a histidine, as well as other motifs (Krejci et al., 1991). A phylogenetic analysis using MEGA3 (Kumar et al., 2001) clustered Mimivirus L906 with prokaryotic paranitrobenzyl carboxylesterases (Fig. 3) that are known to catalyze the hydrolysis of the *para*-nitrobenzyl esters of various β-lactam antibiotics. Although the exact role of the enzyme in Mimivirus is not known, it might play a role in the disruption of the amoeba phagocytic vacuole membrane, helping the virus to gain access to the host cytoplasm.

2.6. Translation apparatus genes: the final frontier between cells and viruses?

With more bacterial genomes being sequenced, the number of core genes strictly shared by all bacterial or archebacterial species, including the smallest parasitic ones, has been steadily decreasing. It is now down to 60 genes, including ribosomal proteins, aminoacyl-tRNA synthetases, and the core components of the transcription and DNA replication apparatus (Crapoulet et al., 2005). This indicates that parasitic bacteria have found evolutionary solutions to eventually dispense with most functions, except for those encoded by these 60 "core" genes. Of course, some of these cellular genes reputed irreplaceable today, might be found absent from a microbial genome sequenced in the future. Already, the nucleomorph (i.e. the enslaved algal nucleus) of the cryptomonad alga *Guillardia theta* does not appear to encode a complete set of ribosomal proteins, and might thus have developed an import mechanism for them (Douglas et al., 2001).

On the other hand, giant virus genomes exhibit an increasingly large assortment of biosynthetic pathways and regulatory components, and most of them exhibit their own DNA replication and transcription apparatus. The variability of their gene contents is so large that any type of gene might eventually turn out in the next giant virus genome that will be sequenced.

The difference that was thought to unequivocally separate the cellular world from the one of viruses was the presence of a working translation apparatus. However, following the presence of tRNAs in the genome of many giant viruses, the discovery of many translation factors and several aminoacyl-tRNA synthetases encoded in Mimivirus genome was a serious blow to this last resort criteria. By now, finding several ribosomal protein genes – or even a ribosomal RNA-like sequence – in a future giant virus genome becomes much less unexpected.

With its size and genome complexity for the first time overlapping with those of many cellular organisms, Mimivirus shattered at once a century old portrait of what a bona fide virus should look like. This, however, is not a simple epistemological readjustment of no practical consequences. Mimivirus is probably the first representative of a long list of many more giant viruses, the proper classification of which will pose a serious and durable challenge to our definition of life forms. As an immediate concern, the interpretation of metagenomic data must now been reappraised. If there is no single gene that is common to all viral genomes (Edwards and Rohwer, 2005), any gene might also eventually turn out in a giant virus. In consequence, the attribution of remotely similar environmental sequences to a giant virus or a cellular organisms can not simply rely on the identity of its closest homologue, but must include a complete taxonomic assessment (e.g. using the best reciprocal match criteria for orthology) followed by a detailed phylogenetic analysis (Ghedin and Claverie, 2005). Even though, a doubt might still subsist about the origin of these sequences, in the case of low sequence similarity and/or insufficient coverage of certain clades. Furthermore, given the usually very large fraction of viral gene sequences exhibiting no similar-

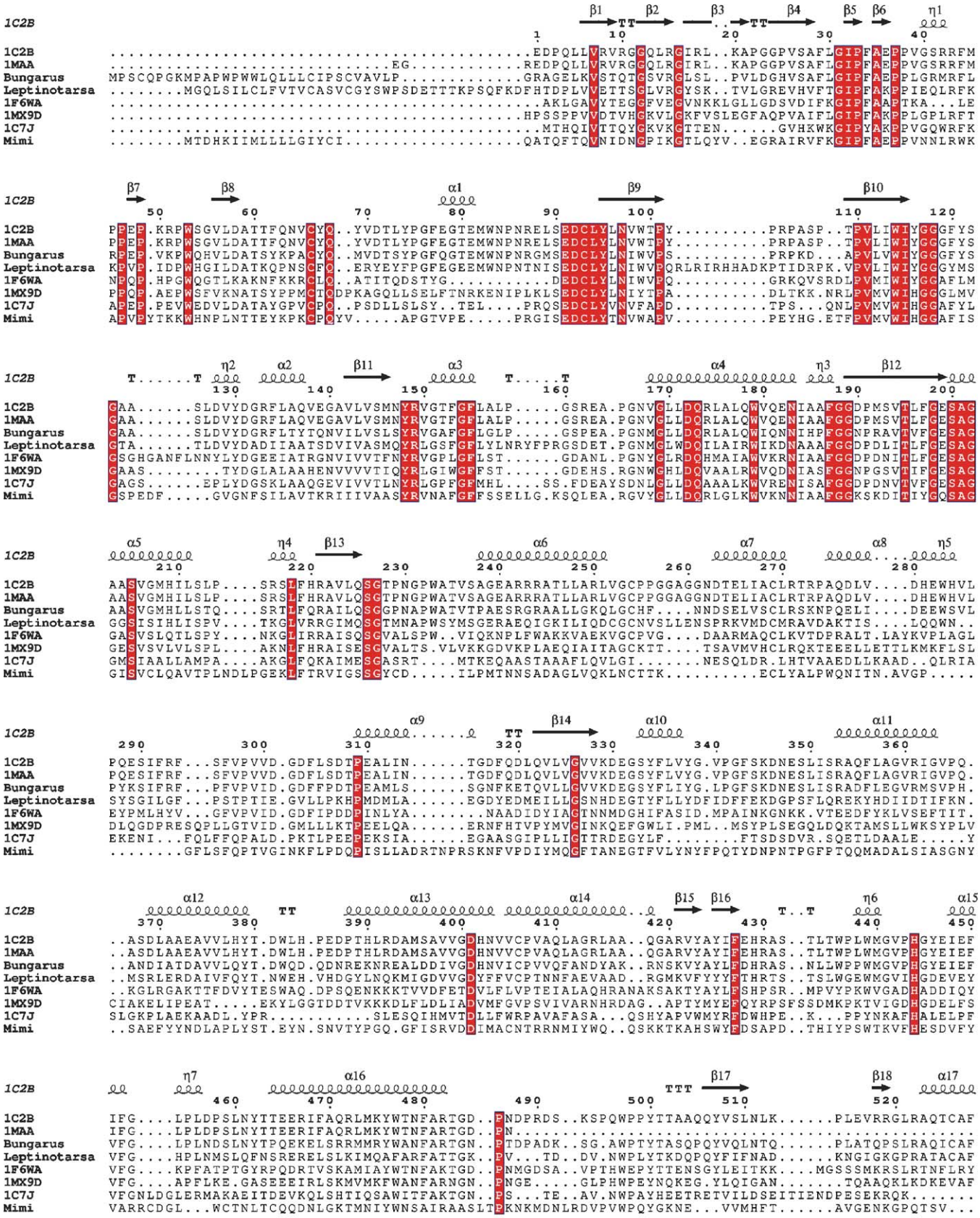

Fig. 2. Comparison of Mimivirus L906 putative protein sequence with typical carboxylesterase amino acid sequences, including proteins of known three-dimentional structures. The multiple alignment was generated with 3D-Coffee (Poirot et al., 2004). The highly conserved catalytic triad is composed of a serine at position 200, a glutamate at position 331, and a histidine at position 444. Critical/conserved residues are boxed. Approximately 40 C-terminal residues exhibiting little conservation across the different sequences are omitted.

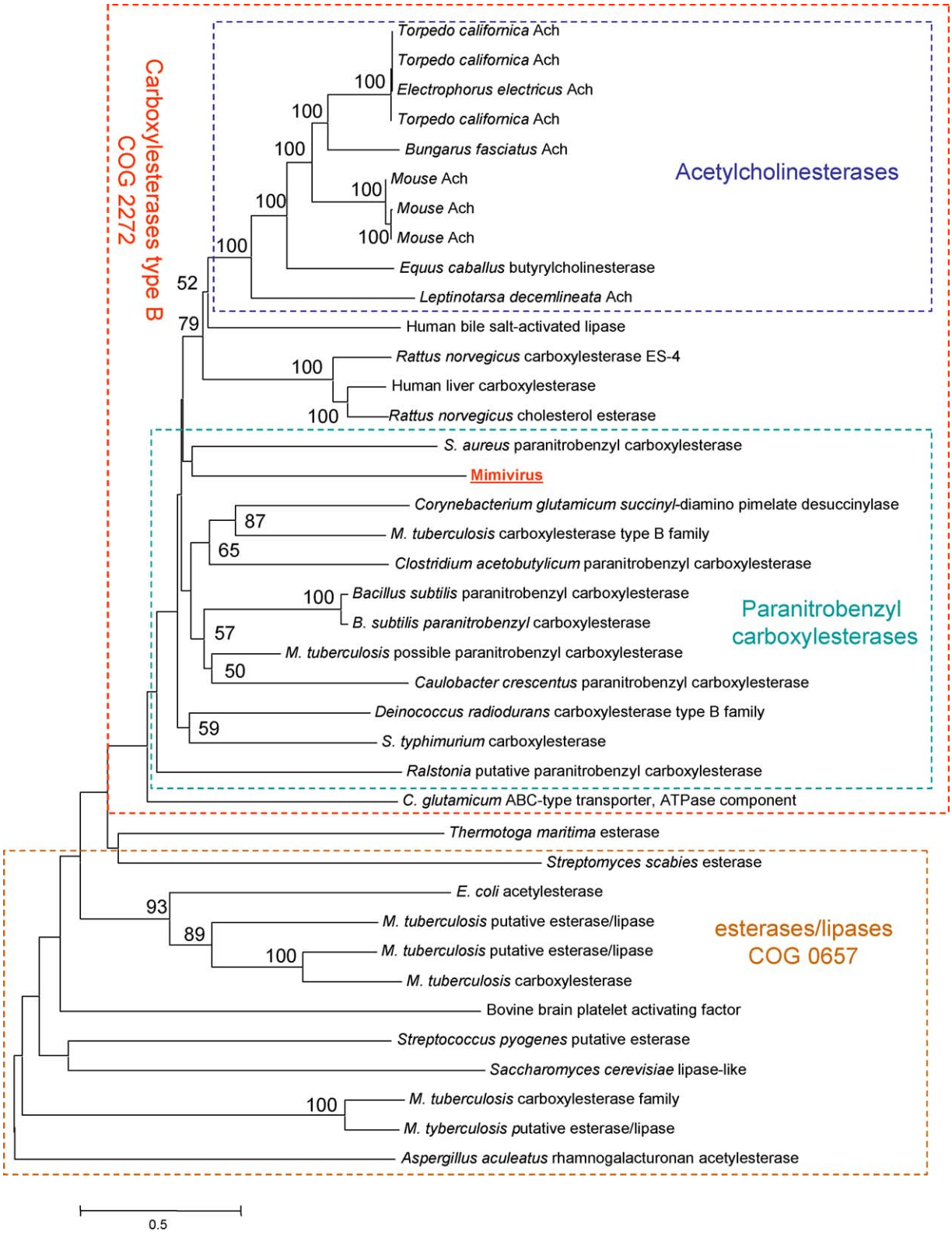

Fig. 3. Phylogenetic classification of the Mimivirus L906 encoded putative carboxylesterase among type B carboxylesterases and other esterases/lipases. The Neighbor-joining method with Poisson correction as provided by the MEGA3 software (Kumar et al., 2001) was used. Bootstrap values above 50% are indicated.

ity within the databases, one could suspect that a significant proportion of the "orphan" genes turning out in environmental sequencing projects might originate from unknown viruses or phages.

2.7. The mysterious origin of giant viruses

The evolutionary forces at the origin of giant viruses (loosely defined as those packing more than 300 genes in 200–300-nm

diameter particles) and the rationale behind their genome complexity are not understood. Various hypotheses can be proposed, from traditional to the most revolutionary. Prior to expose some of them, it is worth to notice that these viruses are truly more complex than their leaner counterparts (e.g. the typical 50 kb genome adenoviruses or phages): the increase in genome size is not due, for instance, to the accumulation of non-coding repeats, junk DNA, pseudo-genes, or the huge expansion of a few gene families. For instance, all the Mimivirus genes that have been successfully expressed as recombinant proteins exhibited their predicted activity or 3D structure (Abergel et al., 2005; Jeudy et al., 2005; Benarroch et al., 2006). These giant viruses also exhibit a more complex ultrastructure than their smaller counterparts as indicated by direct proteomic analyses (Raoult et al., 2004) or electron microscopy (Xiao et al., 2005). Nothing, in the recognizable gene content of giant viruses, appears to predispose them to the capture and accumulation of random DNA segments: in contrast with promiscuous bacteria, their genome is not particularly enriched with mobile elements, palindromic structures, or genes encoding the necessary enzymatic equipment (such as transposases, integrases, etc.).

Yet, viruses are traditionally seen as being prone to frequent lateral gene transfer from their host, and their genomes are considered like bags of randomly accumulated foreign genes, around the limited set of conserved core genes (Iyer et al., 2001) pertaining to essential functions.

We have shown elsewhere (Ogata et al., 2005b) that this picture, inherited from the world of retroviruses and transducing phages, does not agree with our analysis of Mimivirus genome: a sizable proportion (40%) of its predicted proteins does exhibit a significant similarity within the sequence databases, but very few (less than 1%) of them exhibit their closest match with a putative protist-related host such as *Entamoeba histolytica* the complete genome of which could be used (Ogata et al., 2005b), or against *Acanthamoeba castellani* (a host to Mimivirus) 16,000 expressed sequence tags (Fig. 4). Furthermore, Mimivirus sequences sharing the highest similarity with known sequences (hence presumably the most recently acquired ones), such as cholinesterase-like ORF L906, exhibit no phylogenetic affinity with the protist kingdom (Fig. 3).

Using a Bayesian method based on nucleotide word frequencies (Nakamura et al., 2004), we estimated that less than 9% of Mimivirus gene might have been recently acquired by lateral gene transfer. Of course, this value might be underestimated, given the fast evolving rate usually attributed to viruses. But, the relatively easy recognition of very ancestral genes within Mimivirus genome (Raoult et al., 2004; Iyer et al., 2001; Iyer et al., 2005) does not suggest it is a significant factor. Our studies based on phylogenetic tree inference suggest several putative horizontal gene transfers (http://www.giantvirus.org/mimitrees/). For instance, the Mimivirus mismatch repair ATPase (MutS) is most closely related to a homologue encoded in the mitochondrial genome of *Leptogorgia chilensis*. The highly conserved structure of the promoter regions of Mimivirus genes (49% of them exhibiting a unique and strictly conserved motif) (Suhre et al., 2005) is a strong argument against a mosaic structure of Mimivirus genome

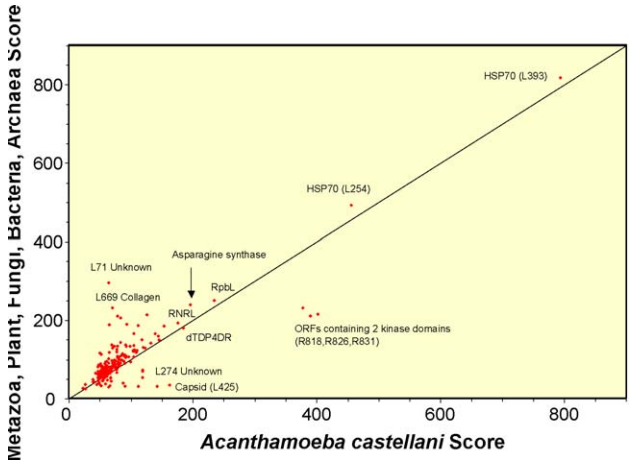

Fig. 4. Distribution of BLAST scores of Mimivirus ORFs against *Acanthamoeba castellani* sequences (vertical axis; TBLASTN) and non-protozoa sequences (horizontal axis; BLASTP). The *A. castellani* sequence set is composed of sequences downloaded through the NCBI Entrez system (18,433 sequences; 19.3 Mbp in total) and those obtained through the Protist EST Program (http://megasun.bch.umontreal.ca/pepdb/pep_main.html; 15,541 EST sequences). BLASTP scores against non-protozoa sequences were obtained using partial Mimivirus sequences (HSPs matching to *A. castellani* data). Abbreviations are as follows: dTDP4DR, dTDP-4-dehydrorhamnose reductase; HSP70, 70-kDa heat-shock protein; RNRL, ribonucleotide reductase large subunit; RpbL, RNA polymerase II largest subunit.

and a high prevalence of horizontal gene transfer from an amoebal host, where these promoter motifs are not found (Suhre et al., 2005). In conclusion, lateral gene transfer is unlikely to account for the huge increase of Mimivirus gene content compared to regular DNA viruses.

Another striking feature of Mimivirus genome is the absence of the sign of reductive genome evolution or pseudogenes. The genome is packed with genes (with an average intergenic region size of 150 nt), none of them showing any indication of degradation. Fig. 5 shows that Mimivirus genome exhibit a higher ORF density for any size ranges than the genomes of *Rickettsia prowazekii* or *Mycobacterium leprae* known to contain a large proportion of "junk" DNA and pseudo-genes due to ongoing

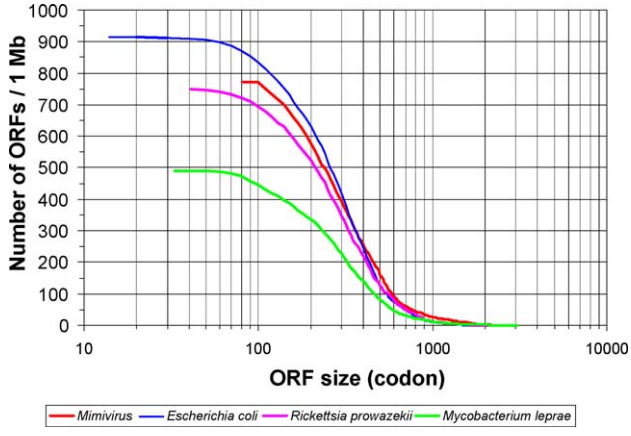

Fig. 5. Comparison of ORF size cumulative distributions. The graph shows the numbers of ORFs (*y*-axis) (per 1 Mb), that are longer than or equal to X-codons (*x*-axis) annotated in the genomes of Mimivirus, *Escherichia coli, Rickettsia prowazekii* and *Mycobacterium leprae*.

reductive evolution (Fig. 5). The Mimivirus ORF density is close to that of *Escherichia coli*, with a higher proportion of ORF larger than 400 residues. The absence of decaying genes argues against Mimivirus large genome being the result of frequent acquisitions of more or less random host genes. It is also in contrast with the clear tendency toward genome reduction/gene degradation observed for most intracellular parasitic bacteria, with similar or smaller genome sizes (such as *Rickettsia, Buchnera*, or *Mycoplasma*, Fig. 1B) (Ogata et al., 2001; Andersson and Andersson, 2001; Moran, 2002).

In contrast to these cellular organisms appearing to be irreversibly evolving toward an increasingly host-dependent lifestyle, Mimivirus and other giant viruses appear to be in an evolutionary steady state, showing no tendency toward reducing their size. On the contrary, Mimivirus exhibits some large families of paralogues originating from relatively recent multiplication/duplication events (Suhre, 2005). A more ancient segmental genome duplication event of about 200,000 bp was also detected. Some of the largest families of recently duplicated genes correspond to tandem duplications of up to 11 copies in a row (genes LI75 to LI85). A phylogenetic analysis of these genes indicates that they are not the result of a burst multiplication of the same ancestral gene, but that they were derived from distinct duplication events, and evolved independently following their creation (Suhre, 2005). These events may thus be regarded as neo- or sub-functionalization events. Using remote homology detection methods (Soding, 2004), a number of these gene families can be linked to functions such as transcription control, cell signaling and protein ubiquitination. We therefore speculate that these genes may play a role in recently acquired and/or diversified host adaptation functions. If any, Mimivirus genome shows more signs toward expansion than to reduction!

At variance with the classical "incremental bag of genes" view of DNA virus, we have proposed (Raoult et al., 2004; Ogata et al., 2005b) that the Mimivirus lineage might have emerged before the individualization of cellular organisms from the three domains of life prior to the Darwinian threshold (Woese, 2002), eventually participating to the mixing of bacterial and archeal genes that led to the emergence of the eukaryotic cell (reviewed in Pennisi, 2004; Ogata et al., 2005b). In that context, different viral genes are not expected to exhibit entirely consistent phytogenies, and the similarity between extant orthologues in cellular or other viral species is expected to be very low.

Hendrix et al. (2000) proposed an ingenious scenario for the origin of bacteriophages based on the concept of "morons", i.e. laterally transferred gene incrementally increasing the fitness of the host cell at the prophage stage. A important feature of this scenario is the central position given to the capsid protein and the gene encoding it. Indeed, the capsid protein, the origin of which is unclear, is the sole component shared by most ds-DNA viruses and unique to the viral world. Unfortunately, if the "moron accretion" scenario fits well with the highly mosaic structures found in phage genomes, it does not apply to eukaryotic ds-DNA viruses that do not integrate in the host genome. Also, this hypothesis does not address the fact that bacteriophages (and eukaryotic DNA viruses) do come in such a broad range of size and complexity.

At the opposite of the "moron accretion" scenario, large DNA viruses can also be seen as the result of a progressive reductive evolution akin to what is experienced by intracellular parasitic bacteria, but along a fundamentally different pathway that is not preserving the integrity of the translation apparatus, and of at least one energy production system. Free-living organisms might thus become intracellular obligate parasite by following the mutually exclusive "cellular" or viral reductive evolutionary pathways. Among the few arguments in favor of this scenario is the coincidence between the genome sizes of the largest known DNA virus and of the largest intracellular parasitic bacteria, as if both started their reductive evolution from an ancestor harboring a minimal free-living genome of about 1.4 Mb (Fig. 1). This scenario does account for the presence of some relics of the translation system, beside a complete DNA replication and transcription system, and miscellaneous metabolic enzymes in the largest viruses, seen as the least advanced (slowest evolving) along their parasitic evolutionary pathway. However, the main difficulty with this scenario is to figure out a mechanisms allowing the transition from the cell-division reproduction mode to the viral multiplication mode.

Clearly, more genome sequences, functional genomics studies, and the determination of the 3D structures of many viral gene products (specially those bearing no similarity to any other protein), are needed to explore the origin of giant viruses, or "girus". In the meantime, we also have to identify the kind of evolutionary constraints that allowed these giant viruses to retain, or achieve, their anomalous genome size and complexity.

Acknowledgements

The authors wishes to acknowledge the support of CNRS, of the French National Genopole Network, and of Marseille-Nice Génopole.

References

Abrescia, N.G., Cockburn, J.J., Grimes, J.M., Sutton, G.C., Diprose, J.M., Butcher, S.J., Fuller, S.D., San Martin, C., Burnett, R.M., Stuart, D.I., Bamford, D.H., Bamford, J.K., 2004. Insights into assembly from structural analysis of bacteriophage PRD1. Nature 432, 68–74.

Abergel, C., Chenivesse, S., Byrne, D., Suhre, K., Arondel, V., Claverie, J.-M., 2005. Mimivirus TyrRS: preliminary structural and functional characterization of the first amino-acyltRNA synthetase found in a virus. Acta. Cryst. F61, 212–215.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Andersson, J.O., Andersson, S.G., 2001. Pseudogenes, junk DNA, and the dynamics of *Rickettsia* genomes. Mol. Biol. E 18, 829–839.

Baudoux, A.C., Brussaard, C.P., 2005. Characterization of different viruses infecting the marine harmful algal bloom species Phaeocystis globosa. Virology 341, 80–90.

Benarroch, D., Claverie, J.-M., 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.

Bougueleret, L., Tekaia, F., Sauvaget, I., Claverie, J.-M., 1988. Objective comparison of exon and intron sequences by means of 2-dimensional data analysis methods. Nucleic Acids Res. 16, 1729–1738.

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J.M., Nulton, J., Salamon, P., Rohwer, F., 2004. Diversity and population structure of a near-shore marine-sediment viral community. Proc. Biol. Sci. 271, 565–574.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., Rohwer, F., 2003. Metagenomic analyses of an uncultured viral community from human feces. J. Bacteriol. 85, 6220–6223.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., Rohwer, F., 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. U.S.A. 99, 14250–14255.

Chee, M.S., Bankier, A.T., Beck, S., et al., 1990. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. 154, 125–169.

Claverie, J.-M., 2005. Giant viruses in the oceans: the 4th Algal Virus Workshop. Virol. J. 2, 52.

Crapoulet, N., Renesto, P., Dumler, J.S., Suhre, K., Ogata, H., Claverie, J.-M., Raoult, D., 2005. *Tropheryma whipplei* genome at the beginning of the post-genomic era. Curr. Genom. 6, 195–205.

Culley, A.I., Lang, A.S., Suttle, C.A., 2003. High diversity of unknown picorna-like viruses in the sea. Nature 424, 1054–1057.

Desjardins, C., Eisen, J.A., Nene, V., 2005. New evolutionary frontiers from unusual virus genomes. Genom. Biol. 6, 212.

Delong, E.F., 2005. Microbial community genomics in the ocean. Nat. Rev. Microbiol. (Advance Online Publication, May 10).

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.T., Wu, X., Reith, M., Cavalier-Smith, T., Maier, U.G., 2001. The highly reduced genome of an enslaved algal nucleus. Nature 410, 1091–1096.

Edwards, R.A., Rohwer, F., 2005. Opinion: viral metagenomics. Nat. Rev. Microbiol. (Advance Online Publication, May 10).

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

Galperin, M.Y., 2005. Life is not defined just in base pairs. Environ. Microbiol. 7, 149–152.

Ghedin, E., Claverie, J.-M., 2005. Mimivirus relatives in the Sargasso Sea. Virol. J. 2, 62.

Ghedin, E., Fraser, C.M., 2005. A virus with big ambitions. Trends Microbiol. 13, 56–57.

Hendrix, R.W., Lawrence, J.G., Hatfull, G.F., Casjens, S., 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8, 504–508.

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

Iyer, L.M., Koonin, E.V., Leipe, D.D., Aravind, L., 2005. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res. 33, 3875–3896.

Jeudy, S., Coutard, B., Lebrun, R., Abergel, C., 2005. *Acanthamoebapolyphaga* mimivirus NDK: preliminary crystallographic analysis of the first viral nucleoside diphosphate kinase. Acta. Cryst. F61, 569–572.

Konstantinidis, K.T., Tiedje, J.M., 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proc. Natl. Acad. Sci. U.S.A. 101, 3160–3165.

Kooijman, S.A., 2001. Quantitative aspects of metabolic organization: a discussion of concepts. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 331–349.

Koonin, E.V., 2005. Virology: Gulliver among the Lilliputians. Curr. Biol. 15, R167–R169.

Korn, E.D., Weisman, R.A., 1967. Phagocytosis of latex beads by Acanthamoeba. II. Electron microscopic study of the initial events. J. Cell. Biol. 34, 219–227.

Krejci, E., Duval, N., Chatonnet, A., Vincens, P., Massoulie, J., 1991. Cholinesterase-like domains in enzymes and structural proteins: functional and evolutionary relationships and identification of a catalytically essential aspartic acid. Proc. Natl. Acad. Sci. U.S.A. 88, 6647–6651.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

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. Infect. Dis. 11, 449–452.

Mesyanzhinov, V.V., Robben, J., Grymonprez, B., Kostyuchenko, V.A., Bourkaltseva, M.V., Sykilinda, N.N., Krylov, V.N., Volckaert, G., 2002. The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. J. Mol. Biol. 317, 1–19.

Moran, N.A., 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108, 583–586.

Moreira, D., Lopez-Garcia, P., 2005. Comment on "the 1.2-megabase genome sequence of mimivirus". Science 308, 1114a.

Nakamura, Y., Itoh, T., Matsuda, H., Gojobori, T., 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nat. Genet. 36, 760–766.

Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X., Baker, T.S., Graves, M.V., Van Etten, J.L., Rossmann, M.G., 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. Proc. Natl. Acad. Sci. U.S.A. 99, 14758–14763.

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 *R. prowazekii*. Science 293, 2093–2098.

Ogata, H., Raoult, D., Claverie, J.-M., 2005a. A new example of viral intein in Mimivirus. Virol. J. 2, 8.

Ogata, H., Abergel, C., Raoult, D., Claverie, J.-M., 2005b. Response to comment on "the 1.2-megabase genome sequence of mimivirus". Science 308, 1114b.

Pennisi, E., 2004. The birth of the nucleus. Science 305, 766.

Poirot, O., Suhre, K., Abergel, C., O'Toole, E., Notredame, C., 2004. 3DCoffee@igs: a webserver for combining sequences and structures into a multiple sequence alignment. Nucleic Acids Res. 32, W37–W40.

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.

Rappe, M.S., Connon, S.A., Vergin, K.L., Giovannoni, S.J., 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418, 630–633.

Renesto, P., Crapoulet, N., Ogata, H., La Scola, B., Vestris, G., Claverie, J.-M., Raoult, D., 2003. Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. Lancet 362, 447–449.

Riesenfeld, C.S., Schloss, P.D., Handelsman, J., 2004. Metagenomics: genomic analysis of microbial communities. Annu. Rev. Genet. 38, 525–552.

Serwer, P., Hayes, S.J., 2001. Partially condensed DNA conformations observed by single molecule fluorescence microscopy. Biophys. J. 81, 3398–3408.

Short, S.M., Suttle, C.A., 2002. Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. Appl. Environ. Microbiol. 68, 1290–1296.

Soding, J., 2004. Protein homology detection by ITMM-ITMM comparison. Bioinformatics 21, 951–960.

Suhre, K., 2005. Gene and genome duplication in *Acanthamoeba polyphaga* Mimivirus. J. Virol. 79, 14095–14101.

Suhre, K., Audic, S., Claverie, J.-M., 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., in press. Genomic and evolutionary aspects of Mimivirus. Virus Res. Available online 21 September 2005.

Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., Bork, P., Hugenholtz, P., Rubin, E.M., 2005. Comparative metagenomics of microbial communities. Science 308, 554–557.

Van Etten, J.L., Lane, L.C., Meints, R.H., 1991. Viruses and virus like particles of eukaryotic algae. Microbiol. Rev. 55, 586–620.

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

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S.,

Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., Smith, H.O., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.

Ward, N., Fraser, C.M., 2005. How genomics has affected the concept of microbiology. Curr. Opin. Microbiol. 8, 564–571.

Weekers, P.H., Engelberts, A.M., Vogels, G.D., 1995. Bacteriolytic activities of the free-living soil amoebae, *Acanthamoeba castellanii, Acanthamoeba polyphaga* and *Hartmannella vermiformis*. Antonie Van Leeuwenhoek 68, 237–243.

Woese, C.R., 2002. On the evolution of cells. Proc. Natl. Acad. Sci. U.S.A. 99, 8742–8747.

Wommack, K.E., Colwell, R.R., 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64, 69–114.

Wommack, K.E., Hill, R.T., Kessel, M., Russek-Cohen, E., Colwell, R.R., 1992. Distribution of viruses in the Chesapeake Bay. Appl. Environ. Microbiol. 58, 2965–2970.

Xiao, C., Chipman, P.R., Battisti, A.J., Bowman, V.D., Renesto, P., Raoult, D., Rossmann, M.G., 2005. Cryo-electron microscopy of the giant Mimivirus. J. Mol. Biol. 353, 493–496.