Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae

Defne Arslan[1], Matthieu Legendre[1], Virginie Seltzer, Chantal Abergel[2], and Jean-Michel Claverie[2]

Information Génomique et Structurale, Centre National de la Recherche Scientifique-Unité Propre de Recherche 2589, Aix-Marseille University, Institut de Microbiologie de la Méditerranée, Parc Scientifique de Luminy, Case 934, FR-13288 Marseille, France

Edited by James L. Van Etten, University of Nebraska, Lincoln, NE, and approved September 13, 2011 (received for review July 6, 2011)

Mimivirus, a DNA virus infecting acanthamoeba, was for a long time the largest known virus both in terms of particle size and gene content. Its genome encodes 979 proteins, including the first four aminoacyl tRNA synthetases (ArgRS, CysRS, MetRS, and TyrRS) ever found outside of cellular organisms. The discovery that Mimivirus encoded trademark cellular functions prompted a wealth of theoretical studies revisiting the concept of virus and associated large DNA viruses with the emergence of early eukaryotes. However, the evolutionary significance of these unique features remained impossible to assess in absence of a Mimivirus relative exhibiting a suitable evolutionary divergence. Here, we present *Megavirus chilensis,* a giant virus isolated off the coast of Chile, but capable of replicating in fresh water acanthamoeba. Its 1,259,197-bp genome is the largest viral genome fully sequenced so far. It encodes 1,120 putative proteins, of which 258 (23%) have no Mimivirus homologs. The 594 Megavirus/Mimivirus orthologs share an average of 50% of identical residues. Despite this divergence, Megavirus retained all of the genomic features characteristic of Mimivirus, including its cellular-like genes. Moreover, Megavirus exhibits three additional aminoacyl-tRNA synthetase genes (IleRS, TrpRS, and AsnRS) adding strong support to the previous suggestion that the Mimivirus/Megavirus lineage evolved from an ancestral cellular genome by reductive evolution. The main differences in gene content between Mimivirus and Megavirus genomes are due to (*i*) lineages specific gains or losses of genes, (*ii*) lineage specific gene family expansion or deletion, and (*iii*) the insertion/migration of mobile elements (intron, intein).

Mimiviridae | DNA virus phylogeny | girus | viral translation | tree of life

The discovery of the viral nature of *Acanthamoeba polyphaga* Mimivirus (1), followed by the determination of its outstanding genome sequence (1,182 kb) (2) led to an irreversible change in the way microbiologists looked at viruses (3–6), even reviving the debate on their classification as living microorganisms (7, 8). After Mimivirus, no obvious limit could be set anymore on the expected size of a viral particle, or the complexity of its gene content, both of them now largely overlapping with that of the simplest cellular organisms, such as parasitic bacteria (9). Moreover, the Mimivirus genome was found to encode a number of functions thought to be trademarks of cellular organisms, such as aminoacyl tRNA synthetases (AARS), threatening to invalidate the absence of a translation system as the last inviolate criterion separating viruses from the cellular world (1, 10). Subsequent studies revealed additional Mimivirus idiosyncrasies such as its unique virion unloading/loading mechanisms (11), singular gene transcription signaling such as the "hairpin rule" (12), and its susceptibility to infection by a new type of satellite virus, called "virophage" (13, 14). Beyond the initial excitation of their discovery, the next step is now to assess to what extent these features are anecdotal or, in the contrary, deeply linked to the emergence and mode of evolution of giant DNA viruses with genome sizes >1 Mb (hereby referred to as "Megaviridae"). Such a task requires comparing Mimivirus with relatives situated at optimal evolutionary distances, close enough to allow the unambiguous assignment of homologous features, but divergent enough to provide a clear illustration of the evolutionary forces at work. Several Mimivirus-related megaviridae have been mentioned in recent literature: Mamavirus (15) (nearly identical to Mimivirus) and few others more briefly described (Terra1, Terra2, Courdo, or Moumou) (16) and for which no genome sequence is available.

Through a campaign of random aquatic environmental sampling, followed by culturing on a panel of acanthamoeba species, we isolated Megavirus from sea water sampled close to the shore off the ECIM marine station in Las Cruces, Chile (*SI Materials and Methods*). Megavirus is now routinely cultured on *A. castellanii*. Its complete genome sequence was determined by using a combination of 454-titanium and Illumina HiSeq approaches. Here, we present an electron microscopy study of the Megavirus replication cycle in *A. castellanii* and an analysis of its genome. Despite their substantial evolutionary divergence, all of the unique features noticed in Mimivirus are conserved in Megavirus, and delineate a core set of cellular-like functions that might be fundamentally linked to the origin and evolution of Megaviridae.

Results and Discussion

Electron Microscopy Data. Megavirus and Mimivirus virion particles exhibit a very similar overall morphology, with a dense core nucleocapsid encased into an icosasedral-like capsid, itself covered by a layer of fibers. However, details make them readily recognizable, even in mixed culture (Fig. 1*A*): The Megavirus fibers are noticeably shorter (75 ± 5 nm compared with 120 ± 5 nm for Mimivirus) and cover a capsid slightly larger in diameter (440 ± 10 nm, compared with 390 ± 10 nm for Mimivirus). These dimensions refer to dehydrated virions as prepared for thin section transmission electronic microscopy (TEM), which induce an ≈20% shrinking of the capsids (17). Native Megavirus icosasedral capsids are thus 520 nm in diameter, corresponding to a total particle diameter of 680 nm. In addition, the "hair" of Megavirus virions often exhibits one or two patches of slightly longer and denser fibers (nicknamed "cowlicks") (Fig. 1*A Inset*). The Megavirus particles exhibit a clearly visible special vertex (Fig. 1*B*), corresponding to the "Stargate" already described for Mimivirus, a five-pronged star structure the opening of which triggers the release of the nucleocapsid into the host cell cytoplasm (11, 14). Upon protease treatments, Megavirus par-

Author contributions: C.A. and J.-M.C. designed research; D.A., M.L., V.S., C.A., and J.-M.C. performed research; D.A., M.L., V.S., C.A., and J.-M.C. analyzed data; and M.L., C.A., and J.-M.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. JN258408). A Megavirus genome browser is available at www.giantvirus.org/megavirus/.

[1]D.A. and M.L. contributed equally to this work.

[2]To whom correspondence may be sent. E-mail: jean-michel.claverie@univmed.fr or chantal.abergel@igs.cnrs-mrs.fr

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1110889108/-/DCSupplemental.

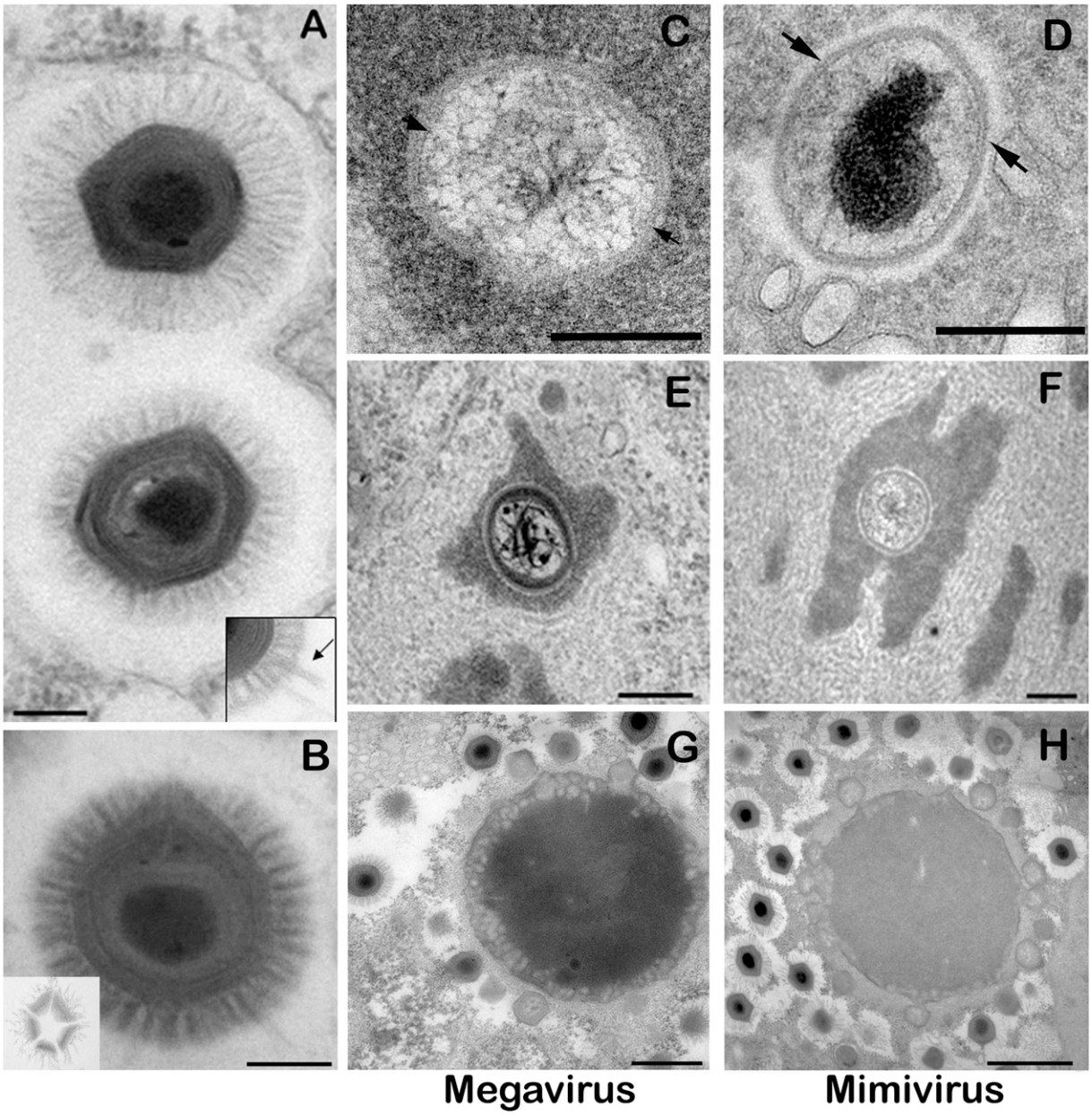

Fig. 1. Electron microscopy of Megavirus compared with Mimivirus. (*A*) Mimivirus (*Upper*) and Megavirus (*Lower*) particles in a same vacuole (coinfection). (*A*, *Inset*) Cowlicks (arrow) as often seen in the Megavirus fiber outer layer. (*B*) Megavirus stargate. (*B*, *Inset*) Transversal section of a Megavirus particle below an open stargate. Megavirus (*C*) and Mimivirus (*D*) seeds surrounded by a lipid membrane (arrows). Megavirus (*E*) and Mimivirus (*F*) early stages of the virion factories with the seeds at their centers. Megavirus (*G*) and Mimivirus (*H*) mature virion factories in full production. (Scale bars: *A*–*F*, 200 nm; *G* and *H*, 1 μm).

ticles appeared more fragile than the Mimivirus' ones, showing a larger proportion of damaged particles in TEM preparation.

The Megavirus isolate was initially tested for growth on *A. griffini*, *A. polyphaga,* and *A.castellanii,* and found to replicate in these three species. For comparison purpose, the replication of Megavirus was then studied in detail on *A. castellanii* (Neff American Type Culture Collection 30010). The infectious cycle of Megavirus lasts 17 h from the initial phagocytosis event to the time of maximal release of particles from fully mature virion factories, compared with 12 h for Mimivirus (MOI ≈ 10). The difference is mostly due to a slower progression from the "seed stage" (Fig. 1 *C* and *D*) to the fully bloomed virion factories (Fig. 1 *G* and *H*). Another macroscopic phenomenon specific of Megavirus is that ≈35% of the *A. castellanii* cells appear to die without undergoing productive infections. The reasons for this cytotoxicity is unknown but suggest that the laboratory *A. castellanii* strain behaves as a nonoptimal substitute for the unknown natural environmental host of Megavirus. We previously observed that Mimivirus induced a rounding of the infected *A. castellanii* cells, ≈6 h after infection. The same phenomenon, albeit slightly delayed, was observed with Megavirus. However, at variance with Mimivirus where rounded cells remained adherent, Megavirus infection caused most of them to come off their support, making the cultures more difficult to monitor by regular microscopy. *A. castellanii* cells infected by Megavirus exhibited three distinctive ultrastructural features in succession, as described for Mimivirus (11, 14, 18). First, the "seed," corresponding to the Megavirus core nucleocapsid extracted from the external particle layers, appeared clearly separated from the cell cytoplasm by a well-delineated lipid membrane (Fig. 1 *C* and *D*).

This membrane likely derives from the most internal of the two membranes visible inside the virus particles, as expected from the infection process (opening of the stargate, followed by the fusion of the first virus membrane with the vacuole membrane). These seeds then progressively turn into early virion factories recognizable as electron-dense nucleic acid-rich regions (18) isolated from the surrounding cytoplasm by an exclusion zone constituted of a mesh of fibrils 16 nm in diameter (Fig. 1 *E* and *F*). The biochemical nature of these fibrils remains to be characterized. Finally, fully mature virion factories release a large number of particles from their periphery where three stages of maturation are seen: empty assembled particles "budding" from the factory edges, "bald" particles filled with the nucleocapsid, and fiber-covered mature particles ready to exit the host cell (Fig. 1 *G* and *H*). The burst size of Megavirus-infected *A. castellanii* cells is approximately one-half the thousand virions released by those infected by Mimivirus.

Megavirus Overall Genome Structure and Gene Content. The genome of Megavirus is a linear double-stranded DNA molecule with a size of 1,259,197 bp (74.76% A+T), making it the largest described viral genome. The whole sequence was assembled and corrected at once from a dataset of 278,663 454-titanium reads combined with 42,288,396 Illumina Hiseq (paired-end) reads (*SI Materials and Methods*). We annotated 1,120 putative protein-coding sequences (CDSs) and 3 tRNAs (1 Trp, and 2 Leu). Megavirus CDSs range from 29 aa to 2,908 aa in length, for an average of 338.4 aa (median = 281.5 aa). The distance separating consecutive CDSs was short (114.5 nt in average), resulting into a very high coding density (90.14%). Eight hundred sixty-two of the 1120 (77%) predicted Megavirus CDS have homologs in Mimivirus, whereas 793 of the 979 (81%) Mimivirus CDSs (19) have homologs in Megavirus. Using the best reciprocal match criterion (*SI Materials and Methods*), Megavirus and Mimivirus were found to share 594 orthologous proteins exhibiting a broad distribution of similarity centered on an average of 50% identical residues (Table S1*A* and Figs. S1 and S2). Most likely inherited from a Megavirus/Mimivirus common ancestor, the corresponding gene set provides a minimal estimate of the core genome of ancestral Megaviridae.

Fig. 2 illustrates at one glance both the overall similarity and divergence of the Megavirus and Mimivirus genomes. They display a large central region of colinearity extending from mg210 to mg804 in Megavirus, and L192 to R730 in Mimivirus. This region is solely disrupted by the inversion of a central 338-kb genome segment and the translocation of a 76-kb distal segment. Interestingly, one of the boundary of the inversion coincides with the slope reversal of the A+C excess profile, a position associated to the origin of replication in bacteria (Fig. S3). However, this quasiperfect colinearity abruptly vanishes at the two extremities of the Megavirus chromosome, respectively corresponding to 193 kb and 327 kb. Although these regions still encompass several hundred of homologous genes, their location and orientation appear extensively shuffled (Fig. 2). The process leading to the total loss of colinearity at the genome extremities remains unknown, because it is not correlated with a local enrichment in transposases, or to more divergent orthologous sequences. For instance, the highest conservation (93% identical residues) was found for a predicted cholinesterase (Mimivirus L906/Megavirus mg981) located at the extremity of the chromosomes.

Interestingly, the same pattern is observed when comparing two poxviruses exhibiting a level of sequence divergence comparable to the one between Megavirus and Mimivirus (e.g., DNA polymerases exhibiting 65% identical residues) (Fig. S4). This observation suggests that Poxviruses and Megaviridae, despite their considerable differences, might share a genome replication strategy (e.g., coupling replication with recombination) (20, 21)

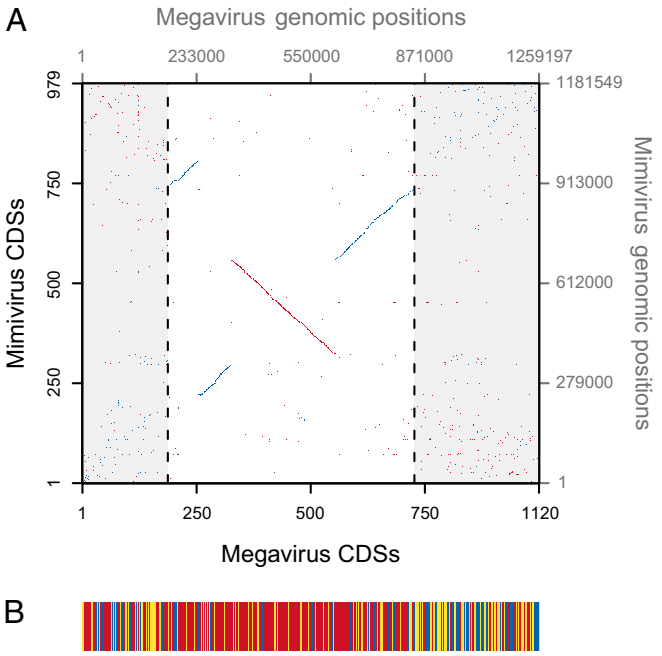

Fig. 2. Comparison of Mimivirus and Megavirus genomes. (*A*) Colinearity of Mimivirus and Megavirus CDSs. Each dot symbolizes the best BLAST match (*e* value $\leq 10^{-5}$) between the CDSs of the two viruses, in the same orientation (blue) or in reverse orientation (red). (*B*) The distribution of homologous CDSs. Megavirus 594 CDSs with Mimivirus orthologs are shown in red. The 268 additional Megavirus CDSs with a significant (nonreciprocal) match in Mimivirus CDSs are shown in blue. CDSs specific to Megavirus are shown in yellow. Orthologs clearly cluster in the central region, whereas the two other categories of CDSs (e.g., duplicated in or specific to Megavirus) tend to cluster at the extremities.

that favor the rearrangement, gain, or loss of genes at the extremities of viral chromosomes. Again, the dramatic 190-kb genome reduction exhibited by a recently described spontaneous Mimivirus mutant is mainly due to large deletions occurring at both ends of the genome (22).

Unique Transcriptional Features in Megaviridae. Previous analyses of the Mimivirus genome uncovered two distinctive features within its noncoding moiety. The first one was the perfect conservation of the octameric motif "AAAATTGA" in front of 45% of Mimivirus CDSs (23). The second was the presence of unrelated palindromic sequences (capable of generating hairpins with a minimal stem length of 15 bp) at the 3′ end of 72% of all Mimivirus mapped transcripts (12). The AAAATTGA motif was later shown to be strongly correlated to early expressed transcripts (24), and the predicted hairpins were demonstrated to serve as polyadenylation signals (12, 24). We examined the 100-nt region upstream of the predicted start codon of the 1,120 Megavirus CDSs. Overall 446 (40%) were found to contain the exact AAAATTGA motif. This proportion was 33.8% (201/594) among Megavirus genes with orthologs in Mimivirus. For 170 of these (85%), the AAAATTGA sequence was also found upstream of Mimivirus orthologs. This ratio shows that (*i*) Megavirus and Mimivirus are using the same motif to specify early gene expression, and (*ii*) that the expression pattern of orthologous genes is globally well conserved. Detailed statistics on the distribution of early and late promoter elements are presented in Table S1*B*.

Similarly, we searched the Megavirus genome 3′ intergenic regions for palindromic sequences obeying the same constraints used to identify them in Mimivirus. Nine hundred fifty-four Megavirus CDSs (85%) genes were found to be followed by

a suitable predicted hairpin. The proportion was also 85% for Megavirus genes orthologous to Mimivirus genes exhibiting a hairpin. This correlation again suggests that these hairpins are the termination signal of Megavirus transcripts and that the transcript structures are well conserved between the two viruses. This prediction was experimentally verified by sequencing the 3′ end of the cDNA of Megavirus mg464, orthologous to Mimivirus major capsid (MCP) gene, L425. These two CDSs, 78% identical at the nucleotide level, encode two proteins sharing 79% of identical residues. As shown in Fig. S5, the polyadenylation of the Megavirus MCP mRNA occurs within the predicted hairpin, albeit 14 nucleotides upstream of the site used in Mimivirus. Given the low level of sequence similarity between these 3′ UTRs (<49% identical), this result provides the demonstration that the polyadenylation of Megaviridae transcripts is uniquely guided by secondary structure information, rather than a sequence signal.

In keeping with the cytoplasmic localization of its replication cycle, Megavirus possesses orthologs of all of the genes previously predicted to encode components of Mimivirus' transcription machinery: the two largest RNA polymerase subunits (mg373, mg339); 11 additional transcription factors: mg307, mg332, mg344, mg577, mg519, mg563, mg552, mg544, mg462, mg438, mg414, and one TATA-box binding like protein (mg435); and one polyA polymerase (mg561). All those genes are located in the middle of the colinear Megavirus/Mimivirus genome segments and exhibit a higher-than-average sequence similarity (64 ± 10% of identical residues).

A recurrent feature of viral genomes is the small proportion of genes for which a function can be predicted (e.g., Megavirus possesses 610 anonymous genes, 54.5% of its genome). It is then tempting to use the intensity of expression of a given gene as an indicator of its importance (i.e., essentiality), eventually to prioritize functional studies. Unexpectedly, this rationale appears to be false, because the expression level of Mimivirus genes exhibited no correlation with their conservation in Megavirus (Fig. S6).

Mimivirus Genes Unique Among dsDNA Viruses Are Conserved in Megavirus. Mimivirus was found to possess many genes never before identified outside of cellular genomes (Table S1*C*). The most unexpected were those related to protein translation, a trademark of cellular organisms. Determining whether these oddities were anecdotal (e.g., due to random horizontal gene transfers) or fundamentally linked to the origin and evolution of all Megaviridae required the identification of a Mimivirus relative at a suitable evolutionary distance. The Mimivirus genome exhibits eight components central to protein translation, including the first four AARS (ArgRS, CysRS, MetRS, and TyrRS) ever found in a virus (2). Megavirus orthologs were found for all of them (respectively encoded by mg804, mg807, mg771, mg907), strongly suggesting that they were present in the Megavirus/Mimivirus common ancestor. More unexpectedly, three additional AARS are found in the Megavirus genome. The AsnRS (mg743) is a member of the class-II AARS, whereas both the TrpRS (mg844) and the IleRS (mg358) are class-I AARS, like all of the ones found in Mimivirus. The significance of this finding is twofold: first, it demonstrates that viral AARS are not limited to class-I enzymes, and second, it makes the scenario of independent acquisition of these genes by HGT increasingly unlikely. Interestingly, the 730-kb genome of the Cafeteria roenbergensis virus, a very distant relative of Mimivirus, also encodes an IleRS (25). Both of these viral IleRS sequences are from the cytoplasmic archaeal/eukaryotic type, but do not exhibit phylogenetic affinity with a known eukaryotic clade (Fig. 3). The Megavirus AsnRS is branching at the origin of the mitochondrial enzyme type, before the radiation of the eukaryotes (Fig. S7*A*). The Megavirus TrpRS is of the archeal/eukaryotic type, again

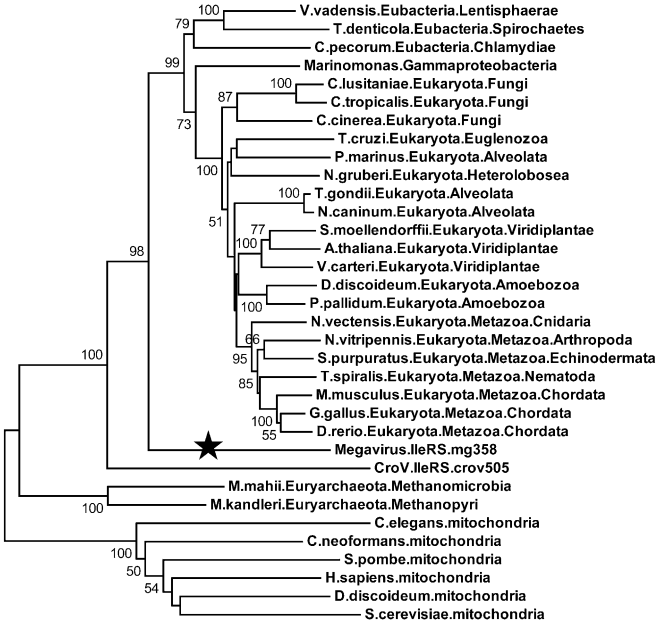

Fig 3. Phylogenetic reconstruction of IleRS sequences. The midpoint-rooted neighbor-joining tree was generated from a 381-aa alignment of conserved positions. The tree topology and bootstrap values were very similar when using different alignment programs, reconstruction methods, and substitution models) (*SI Materials and Methods* and *Materials and Methods*). The Megavirus (star) and CroV IleRS sequences are branching off the eukaryote domain before the radiation of the cytoplasmic IleRSs.

branching out before the radiation of the eukaryotes (Fig. S7*B*). In our opinion, these three additional Megavirus AARS were part of an ancestral Megaviridae genome and were subsequently lost in the Mimivirus lineage. These seven AARS were most likely the remnant of a complete set of 20 AARS inherited from an ancestral cellular genome.

The Mimivirus genome was also found unusually packed with DNA repair enzymes, capable of correcting damages caused by UV light, ionizing radiation, or chemical mutagens (2). Megavirus exhibits orthologs to all these previously identified genes, including a specific type of mismatch repair enzyme MutS (mg543) common to large DNA viruses and specifically abundant in the marine environment (26). In addition, Megavirus exhibits a DNA photolyase (mg779), an enzyme using the energy of light to repair thymidine dimers. A remnant of this gene, interrupted by a transposase, can be found in Mimivirus (R853-R855). The presence of this functional photolyase in Megavirus, among its many other DNA repair enzymes, might participate to its increased resistance to UV irradiation: At a level sufficient to totally inactivate Mimivirus (30 min, 20 cm under a 30 W UV lamp, 253 nm), close to 100% of Megavirus particles remained infectious. One hour of irradiation at the same intensity was required to cause a 90% decrease in Megavirus infectivity.

The three types of topoisomerases found in Mimivirus: one of type II (R480), a bacterial-like type I (L221), and a pox-like type I (R194) are also found in Megavirus (mg403, mg323, mg859). However, Megavirus mg859 is a much closer homolog (50% identical) to the topoisomerase 1b found in CroV (crov152), suggesting that the Mimivirus R194 gene has a distinct origin.

Mimivirus exhibited a number of enzymes related to protein folding, and various sugar and amino acid manipulating enzymes not usually found in viruses. With the exception of two enzymes of the cholesterol biosynthetic pathways, all these genes have well conserved homologs in Megavirus (Table S1*C*), showing that they are part of the Megaviridae core gene set.

Finally, Megavirus mg431 encodes a uridine monophosphate kinase, an enzyme previously undescribed in a DNA virus. This enzyme is 44% identical to its closest homologs in bacteria where it is the rate-limiting enzyme of the pyrimidine salvage pathway (Table S1*C*).

Lineage-Specific Genome Features. The progressive accumulation of point mutations (random drift) appears to be the main cause of the divergence between the Megavirus and Mimivirus genomes. The orthologous CDSs, as well as the "orthologous" intergenic regions, exhibit an average similarity of 65% identical nucleotides, higher than the 50% identical residues shared by orthologous proteins (Fig. S1).

Two hundred fifty-eight Megavirus CDSs exhibit no obvious homolog in Mimivirus and, reciprocally, 186 Mimivirus CDSs have no homolog in Megavirus. More than 85% of these lineage-specific CDSs correspond to proteins without functional predictions. They appear to cluster at the ends of the Megavirus chromosome (Fig. 2*B*). These CDSs might also correspond to fast evolving proteins pushed below our similarity threshold ($e < 10^{-5}$). More likely, they correspond to lineage specific losses along the Mimivirus or Megaviruses branches. Genome reduction is a universal, fast, and irreversible process among intracellular parasitic microorganisms (27, 28) that might also apply to the evolution of Megaviridae from their more complex ancestors (3). An alternate scenario would attribute the 258 Megavirus "private" genes to horizontal gene transfers (HGT) that occurred after the divergence of the Megavirus/Mimivirus branches. However, when screened against the NR protein sequence database, these Megavirus private genes exhibited a much lower percentage of matches (17% vs. 52%) than the genes with orthologs in Mimivirus. Moreover, the few matching genes (44/258) exhibited no peculiar affinity with potential gene donors. This result argues against recent HGTs (at least from known viruses or cellular organisms) as the major cause in the difference in gene content between Megavirus and Mimivirus.

The difference in the gene content of the two Megaviridae is also due to the differential expansion or reduction of large paralogous families. Six hundred fifty-two of the Megavirus predicted proteins are in one copy (not matching elsewhere in the genome at $e < 10^{-5}$), representing 58.2% of the gene content. The corresponding number is 585 (59.8%) for Mimivirus. The genome of Megavirus is thus truly more complex than the one of Mimivirus and not simply repeat-rich or more redundant. Although the distributions of single- vs. multiple-copy genes are globally similar for the two viruses, specific protein families experienced lineage specific expansions. For instance, the third largest Mimivirus gene cluster (referred to as the N172 L cluster; ref. 29) corresponding to 14 paralogues is represented by a single copy in Megavirus. Conversely, the set of 10 FNIP repeat-containing proteins constituting the N165 cluster in Mimivirus (29) is inflated to 55 members (mg34 paralogues) in Megavirus.

Finally, only a few differences between the Megavirus and Mimivirus genomes resulted from lineage-specific movements of mobile elements. Megavirus exhibits five segmented CDSs. The major capsid protein (mg464, with two introns), the largest RNA polymerase subunit (mg373, with one intron), and the DNA polymerase (mg582, with one intein) exhibit the same exact topology than their Mimivirus orthologs. Orf mg500, encoding a HSP70 chaperonin-like protein, contains a type I intron, whereas its Mimivirus ortholog (L393) does not. The second largest RNA polymerase subunit (mg339 orthologous to Mimivirus L244) exhibited the most intricate rearrangements with changes in the numbers and locations of introns, and the insertion of an intein (30) (Fig. S8). The duplication and movement of transposases (five in Mimivirus; only two in Megavirus) also caused the disruption of a Mimivirus photolyase (R853-R855) and a second photolyase paralogue in Megavirus (mg400).

Conclusion

We present the analysis of a Mimivirus relative isolated from a marine environment. With a genome 77 kb (6.5%) larger than Mimivirus, Megavirus shows that the limit is not yet reached in the complexity of giant DNA viruses infecting acanthamoeba or other phagotrophs from yet-uncharted protozoan clades.

The potential origin of giant mimivirus-like genomes has been hotly debated, basically opposing two views. One is depicting Mimivirus as an extremely efficient gene "pickpocket," explaining its large genome as the result of considerable HGTs from its host, bacteria, or other viruses (7, 20). This scenario has been criticized in detail elsewhere (4, 8, 14). The opposite view claims that the level of HGT remained marginal (10%) and that most of the Mimivirus genes originated from an even more complex viral ancestor, itself eventually derived from an ancestral cellular genome (4, 8). The origin of the many cell-specific functions uniquely encoded by Mimivirus is central to this debate.

Thanks to their optimal evolutionary distance, the comparison between the Mimivirus and Megavirus genomes allowed us to delineate a common gene set most likely derived from their common ancestor. This ancestral gene set was found to include most of the Mimivirus cell-like key functions, in particular the ones associated to protein translation. Moreover, three additional AARS were identified in Megavirus, to our opinion ruling out HGT (i.e., seven independent gene acquisitions) as the origin of these genes. In contrast, our analyses corroborate the scenario whereby the last Megaviridae common ancestor originated from a cellular organism (thus endowed of a translation apparatus), from which today's Megaviridae (Fig. S7*C*) mostly derived by a number of lineage specific genome reduction events. The analysis of other distant Megaviridae genomes will provide an increasingly clearer picture of this evolutionary process.

Materials and Methods

Megavirus chilensis was produced in *A. castellanii* and purified as described for Mimivirus (12) where the discontinuous gradient of CsCl was replaced by a sucrose gradient (10/20/30/40%), the viral pellet resuspended in PBS was layered on, and was centrifuged at 5,000 × *g* for 30 min.

DNA Extraction. The purified Megavirus pellet was resuspended in 50 mM Tris·HCl at pH 7.5 and incubated for 30 min at 37 °C in the presence of 1.25 mg/mL lysozyme. Lysis was performed by adding 0.1 mg/mL proteinase K incubated 30 min at 55 °C, 1% laurylsulfate 30 min at room temperature, and 10 mM DTT overnight at room temperature. After phenol-chloroforme extraction and alcohol precipitation, the recovered DNA was resuspended in RNase-Dnase free water.

Electron Microscopy. The *A. castellanii*-infected cells were washed in PBS and resuspended in 2% paraformaldehyde 0.5% glutaraldehyde in PBS for 1 h at room temperature. After three washes in PBS, cell pellet was fixed in 2% osmium tetroxyde, washed once in PBS, dehydrated in 70, 95, and 100% ethanol, and embedded in Epon-812. Ultrathin sections were poststained with 4% uranyl acetate and lead citrate and were observed by using FEI Tecnai TEM operating at 120 kV.

SI Materials and Methods contains the details of the procedures used for the virus isolation and its genome sequencing, assembly, and annotation.

ACKNOWLEDGMENTS. We thank A. Bernadac and P. Bergam for help with the electron microscopy, D. Byrne and A. Lartigue for sequence validations, A. Chambouvet for early culture attempts, and Dr. Moriah Szpara for reading the manuscript. DNA sequencing was performed by the University of Oklahoma's Advanced Center for Genome Technology (Prof. B. Roe) and GATC Biotech. This work was supported by the Centre National de la Recherche Scientifique, Agence Nationale de la Recherche Grant ANR-BLAN08-0089, and a fellowship from the Direction Générale de l'Armement (to D.A.). The sampling expedition was sponsored by the ASSEMBLE initiative (European Commission's seventh framework program).

1. La Scola B, et al. (2003) A giant virus in amoebae. *Science* 299:2033.
2. Raoult D, et al. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306: 1344–1350.
3. Claverie JM (2006) Viruses take center stage in cellular evolution. *Genome Biol* 7:110.
4. Claverie JM, Abergel C (2010) Mimivirus: The emerging paradox of quasi-autonomous viruses. *Trends Genet* 26:431–437.
5. Raoult D, Forterre P (2008) Redefining viruses: Lessons from Mimivirus. *Nat Rev Microbiol* 6:315–319.
6. Villarreal LP, Witzany G (2010) Viruses are essential agents within the roots and stem of the tree of life. *J Theor Biol* 262:698–710.
7. Moreira D, López-García P (2009) Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol* 7:306–311.
8. Claverie JM, Ogata H (2009) Ten good reasons not to exclude giruses from the evolutionary picture. *Nat Rev Microbiol*, 7:615, author reply 615.
9. Claverie JM, et al. (2006) Mimivirus and the emerging concept of "giant" virus. *Virus Res* 117:133–144.
10. Abergel C, Rudinger-Thirion J, Giegé R, Claverie JM (2007) Virus-encoded aminoacyl-tRNA synthetases: Structural and functional characterization of mimivirus TyrRS and MetRS. *J Virol* 81:12406–12417.
11. Zauberman N, et al. (2008) Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. *PLoS Biol* 6:e114.
12. Byrne D, et al. (2009) The polyadenylation site of Mimivirus transcripts obeys a stringent 'hairpin rule'. *Genome Res* 19:1233–1242.
13. La Scola B, et al. (2008) The virophage as a unique parasite of the giant mimivirus. *Nature* 455:100–104.
14. Claverie JM, Abergel C (2009) Mimivirus and its virophage. *Annu Rev Genet* 43:49–66.
15. Colson P, et al. (2011) Viruses with more than 1000 genes: Mamavirus, a new Acanthamoeba castellanii mimivirus strain, and reannotation of mimivirus genes. *Genome Biol Evol* 3:737–742.
16. La Scola B, et al. (2010) Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. *Intervirology* 53:344–353.
17. Xiao C, et al. (2009) Structural studies of the giant mimivirus. *PLoS Biol* 7:e92.
18. Mutsafi Y, Zauberman N, Sabanay I, Minsky A (2010) Vaccinia-like cytoplasmic replication of the giant Mimivirus. *Proc Natl Acad Sci USA* 107:5978–5982.
19. Legendre M, Santini S, Rico A, Abergel C, Claverie JM (2011) Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. *Virol J* 8:99.
20. Filée 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.
21. Esteban DJ, Hutchinson AP (2011) Genes in the terminal regions of orthopoxvirus genomes experience adaptive molecular evolution. *BMC Genomics* 12:261.
22. Boyer M, et al. (2011) Mimivirus shows dramatic genome reduction after intra-amoebal culture. *Proc Natl Acad Sci USA* 108:10296–10301.
23. Suhre K, Audic S, Claverie JM (2005) Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. *Proc Natl Acad Sci USA* 102: 14689–14693.
24. Legendre M, et al. (2010) mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res* 20:664–674.
25. Fischer MG, Allen MJ, Wilson WH, Suttle CA (2010) Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc Natl Acad Sci USA* 107: 19508–19513.
26. Ogata H, et al. (2011) Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. *ISME J* 5:1143–1151.
27. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D (2009) Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol Direct* 4:13.
28. Blanc G, et al. (2007) Reductive genome evolution from the mother of Rickettsia. *PLoS Genet* 3:e14.
29. Suhre K (2005) Gene and genome duplication in Acanthamoeba polyphaga Mimivirus. *J Virol* 79:14095–14101.
30. Perler FB (2002) InBase: The intein database. *Nucleic Acids Res* 30:383–384.