

Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms

Mickaël Boyer^{a,1}, Natalya Yutin^{b,1}, Isabelle Pagnier^a, Lina Barrassi^a, Ghislain Fournous^a, Leon Espinosa^a, Catherine Robert^a, Saïd Azza^a, Siyang Sun^c, Michael G. Rossmann^{c,2}, Marie Suzan-Monti^{a,3}, Bernard La Scola^a, Eugene V. Koonin^b, and Didier Raoult^{a,2}

^aUnité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, Centre National de la Recherche Scientifique, Unité Mixte de Recherche, Institut de Recherche pour le Développement 6236, Faculté de Médecine, Université de la Méditerranée, 13385 Marseille Cedex 5, France; ^bNational Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894; and ^cDepartment of Biological Sciences, Purdue University, West Lafayette, IN 47907

Contributed by Michael G. Rossmann, October 22, 2009 (sent for review September 1, 2009)

Giant viruses such as Mimivirus isolated from amoeba found in aquatic habitats show biological sophistication comparable to that of simple cellular life forms and seem to evolve by similar mechanisms, including extensive gene duplication and horizontal gene transfer (HGT), possibly in part through a viral parasite, the virophage. We report here the isolation of “Marseille” virus, a previously uncharacterized giant virus of amoeba. The virions of Marseillevirus encompass a 368-kb genome, a minimum of 49 proteins, and some messenger RNAs. Phylogenetic analysis of core genes indicates that Marseillevirus is the prototype of a family of nucleocytoplasmic large DNA viruses (NCLDV) of eukaryotes. The genome repertoire of the virus is composed of typical NCLDV core genes and genes apparently obtained from eukaryotic hosts and their parasites or symbionts, both bacterial and viral. We propose that amoebae are “melting pots” of microbial evolution where diverse forms emerge, including giant viruses with complex gene repertoires of various origins.

giant virus | horizontal gene transfer | nucleocytoplasmic large DNA virus | viral evolution

Definitions of viruses are commonly based on size criteria (1), and fine filters are routinely used for virus isolation. For this reason and also because virus research heavily focused on viruses infecting animals and plants, giant viruses have not been discovered until recently. Accordingly, viruses were generally regarded as small, specialized complexes of biomolecules rather than complex organisms (2). The concept of “giant virus” emerged with the discovery of phycodnaviruses, whose particle size is between 160 and 200 nm (i.e., *Paramecium bursaria Chlorella* virus) (3). Amoebae, as wild phagocytes, ingest any particles larger than 0.2 μm (4) and are therefore a potential source of giant viruses. Previous findings indicate that amoebae of the genus *Acanthamoeba* support multiplication of giant viruses such as Mimivirus and Mamavirus (5, 6) as well as the virophage Sputnik, a small virus parasite of the giant Mamavirus (7). Here we describe Marseillevirus, a giant virus isolated from the same host.

Results and Discussion

Structural Characterization of a Large Icosahedral Virus Isolated from Amoeba. Cocultivation experiments were performed between *A. polyphaga* and samples of water from a cooling tower located in Paris and monitored during 52 weeks, as previously described for Mamavirus isolation (7). Cell lysis was observed at 19 weeks of monitoring, and transmission electron microscopy showed the presence of virus particles of about 250 nm in diameter with an icosahedral capsid morphology (Fig. 1). Between 30 min and 1 h postinfection (p.i.), viruses were shown entering the amoeba (Fig. 1A); at later times p.i., a virus factory (VF) with a diffuse aspect was observed close to the amoeba nucleus (Fig. 1B), where both capsid assembly and viral DNA encapsidation seemed to occur simulta-

neously (Fig. 1C), leading to the formation of immature and mature viral particles (Fig. 1D). The Marseillevirus replication cycle was complete at 5 h p.i., an unusually rapid course of virus reproduction. Kinetics and quantification of the Marseillevirus replication cycle are presented in *SI Text*. A preliminary cryo-electron microscopy (cryo-EM) 3D reconstruction using images of purified virus showed that the virus has a roughly icosahedral shape with a diameter of about 250 nm. In addition, the virus possesses 12-nm-long fibers with globular ends on the surface (Fig. 1E and F). The capsid shell is ≈ 10 nm thick and is separated from the internal nucleocapsid by a gap of ≈ 5 nm. The nucleocapsid has a shape that roughly matches the external capsid structure and might be surrounded by a membrane (Fig. 1G).

Using 2D gel electrophoresis followed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Table S1), we identified 49 proteins in purified Marseillevirus virions (Fig. S1). The proteins detected in the virion represent diverse predicted functions, including bona fide structural proteins (e.g., capsid proteins) and some proteins potentially involved in the early stage of the virus cycle (e.g., an early transcription factor, a protein kinase, and an ankyrin repeat-containing protein). The detected virion proteins included products of some of the (nearly) universal nucleocytoplasmic large DNA virus (NCLDV) genes (8, 9), the most abundant ones being the capsid protein, a D6R-type helicase, and a S/T protein kinase, as well as products of genes that are conserved in subsets of the NCLDV, such as thioredoxin/glutaredoxin, RNase III, papain-like cysteine protease, and an ankyrin-repeat protein (Table S1). Western blot analysis with a mouse polyclonal antiserum against purified viral particles identified antigenic properties for 11 viral proteins, including products of four genes without detectable homologs (ORFans) (Fig. S1 and Table S1). Extensive posttranslational modification occurred during Marseillevirus protein synthesis: 10 of the 49 identified virion proteins were glycosylated and 19 were phosphorylated (Fig. S1 and Table S1). The virion also encapsidates some viral messenger RNAs similarly to Mimivirus (*SI Text*).

Author contributions: B.L.S. and D.R. designed research; M.B., N.Y., I.P., L.B., G.F., L.E., C.R., S.A., S.S., M.G.R., M.S.-M., and E.V.K. performed research; M.B., N.Y., I.P., L.B., G.F., L.E., C.R., S.A., S.S., M.G.R., M.S.-M., B.L.S., E.V.K., and D.R. analyzed data; and M.B., N.Y., S.S., M.G.R., M.S.-M., B.L.S., E.V.K., and D.R. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The Marseillevirus genome reported in this paper has been deposited in the GenBank database (accession no. GU071086).

¹M.B. and N.Y. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: mr@purdue.edu or didier.raoult@gmail.com.

³Present address: INSERM U912, 23 Rue Stanislas Torrents, 13006 Marseille, France.

This article contains supporting information online at www.pnas.org/cgi/content/full/0911354106/DCSupplemental.

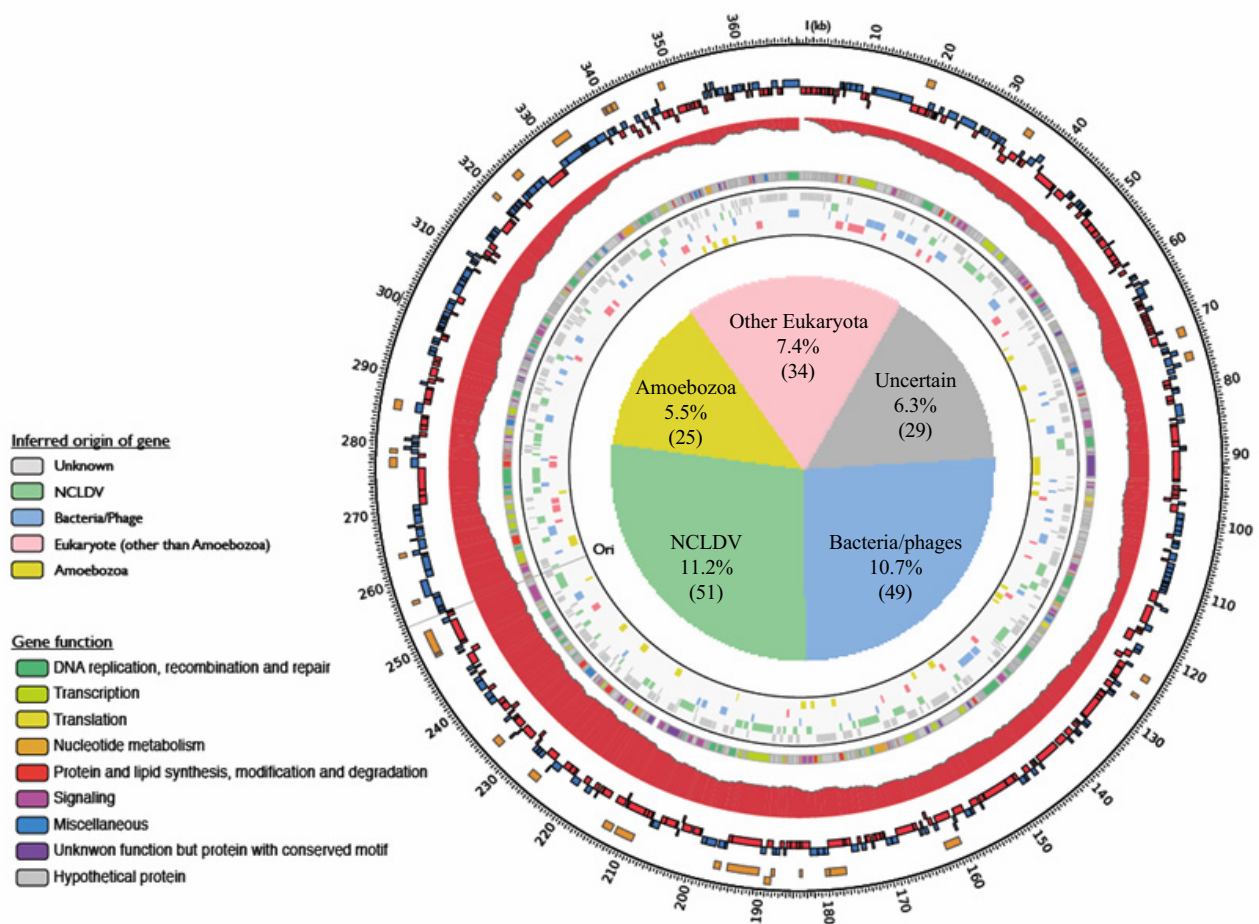


Fig. 2. Map of the Marseillevirus chromosome. Rings starting from outer to innermost correspond to (i) genome coordinates in kilobases; (ii) proteins identified through 2D mass spectrometry (orange); (iii) predicted protein-coding genes oriented in forward (blue) or reverse (red) strand; (iv) cumulative gene orientation skew; (v) predicted functions of proteins; and (vi) origin of each gene inferred from sequence comparison and phylogenetic analyses (light gray background). The pie chart inside the ring represents taxonomic breakdown of Marseillevirus genes by probable origins inferred by phylogenetic analysis or sequence conservation (Table S4). "Ori" indicates putative origin of replication deduced from the position of slope reversal (around position 253,000) of the cumulative gene orientation skew.

1 and *Cotesia plutellae* bracovirus, which do not belong to the NCLDV, have been shown to encode histone-like proteins (15, 16). The histone-like proteins of Marseillevirus were detected in the viral particle (Table S1), suggesting that these proteins could condense DNA to facilitate viral DNA packaging.

Marseillevirus Genome Encompasses a Complex Repertoire of Genes of Various Origins. Among the predicted Marseillevirus proteins, 59, 57, 70, and 2 showed the highest sequence similarity to homologs from viruses, bacteria, eukaryotes, and archaea, respectively (Fig. 2 and Table S4). We hypothesize that the genome repertoire of Marseillevirus consists of genes derived from several distinct sources, in large part via HGT. The presence of numerous genes apparently derived from eukaryotes on different time scales is common in NCLDV (9, 17, 18). The presence of numerous genes of probable bacterial origin seems to be a distinctive feature of those NCLDV that infect unicellular eukaryotic hosts, in particular, the Mimivirus and Marseillevirus reproducing in amoebae, and algal Phycodnaviruses (18) (Tables S1 and Table S4). In addition to the NCBI databases, we searched the draft genome of *Acanthamoeba castellanii*, the host of Marseillevirus, for possible homologs of viral genes. Altogether we identified *Acanthamoeba* homolog to 80 Marseillevirus genes; for eight of these genes, the homolog from *Acanthamoeba* showed the closest similarity to the corresponding Marseillevirus protein, suggesting relatively recent HGT.

A notable feature of Marseillevirus is the presence of 17 genes shared with Mimivirus/Mamavirus but absent in other NCLDV. When a tree of the NCLDV was constructed by comparison of gene repertoires (19), Marseillevirus confidently grouped with Mimivirus/Mamavirus (Fig. 4), in contrast to the phylogenetic tree of the universal genes, which puts Marseillevirus together with Iridoviruses and Ascoviruses (Fig. 3). Thus, the gene repertoires of the two families of amoebae viruses are related, in all likelihood, as a result of interviral HGT. Moreover, eight Marseillevirus genes either had detectable homologs only in the Mimivirus/Mamavirus and *Acanthamoeba* or formed a distinct branch in the respective phylogenetic trees (Table S4), an indication of multiple gene exchanges between amoebae and its viruses.

To characterize the origins of Marseillevirus genes more precisely, we performed a comprehensive phylogenetic analysis. Phylogenetic trees were constructed for 89 Marseillevirus proteins with homologs in diverse organisms and sufficient number of phylogenetically informative sites in the respective multiple alignments; *Acanthamoeba* was included in the tree whenever a homolog of the respective Marseillevirus gene was detected (Table S4 and Fig. S4). The results of this analysis, combined with the information on genes uniquely shared with different taxa, yielded the final breakdown of the Marseillevirus genes by their probable evolutionary origin (Fig. 2 and Table S4). Altogether, Marseillevirus contains 51 genes of NCLDV origin (including those exclusively shared with Mimivi-



Fig. 3. A maximum-likelihood tree based on concatenated alignments (1,849 positions) of five NCLDV core proteins: D5 type ATPase, DNA polymerase B, A32 ATPase, major capsid protein, and A1L/VLTF2 transcription factor. The tree was built using TreeFinder (WAG[,]:G[Optimum]:4, 1,000 replicates, Search Depth 2).

ruses and Phycodnaviruses), 49 genes of probable bacterial or bacteriophage origin, and 85 genes of apparent eukaryotic origin. For 25 of the “eukaryotic” genes of Marseillevirus, an origin from *Acanthamoeba* was strongly supported, and in 22 of these cases, the respective branch of the tree encompassed Marseillevirus, Mimiviruses, and *Acanthamoeba*, with the implication of multiple gene transfers. In addition, three genes seemed to originate in other *Amoebozoa* (Table S4 and Fig. 2). In contrast, no gene could be traced to the known bacterial parasites of amoebae such as *Legionella* or *Parachlamydia*.

Notably, the genes for histone-like proteins, as well as four of the five genes encoding translation system components, were apparently acquired in amoeba, in agreement with the recent observations on the origin of Mimivirus genes with similar functions (17) but not with the hypothesis on the ancestral nature of the translation apparatus components in giant viruses (20). There seems to be a nonrandom connection between the functions of Marseillevirus genes and their inferred origins; many of the genes encoding defense and repair functions—in particular, nucleases—appear to be of bacterial or bacteriophage origin (often shared with other NCLDV), genes for metabolic enzymes and proteins implicated in protein and lipid modification or degradation are of mixed bacterial and eukaryotic origins, whereas genes related to signal transduction are primarily of eukaryotic extraction (Table S4 and Fig. S4.1–S4.82). In addition to the genes that appear to have common origin in Marseillevirus and Mimiviruses, we detected several cases where related genes (e.g., the gene for deoxynucleotide monophosphate kinase) were apparently acquired by these viruses from independent sources (Fig. S4.23). This finding suggests that HGT is common enough to translate into a nonnegligible chance of convergent acquisition of genes that confer a selective advantage onto recipient viruses.

Viruses of amoeba are characterized by a large size and chimeric genomes, with a gene repertoire acquired from a variety of distinct sources. These viruses harbor a conserved core of NCLDV genes that encodes key proteins responsible for viral genome replication and virion morphogenesis, an additional group of genes shared by amoebal viruses, and a broad variety of genes acquired from bacteria, selfish elements, and eukaryotes. Evidence of direct derivation from *Acanthamoeba* or its known parasites or symbionts

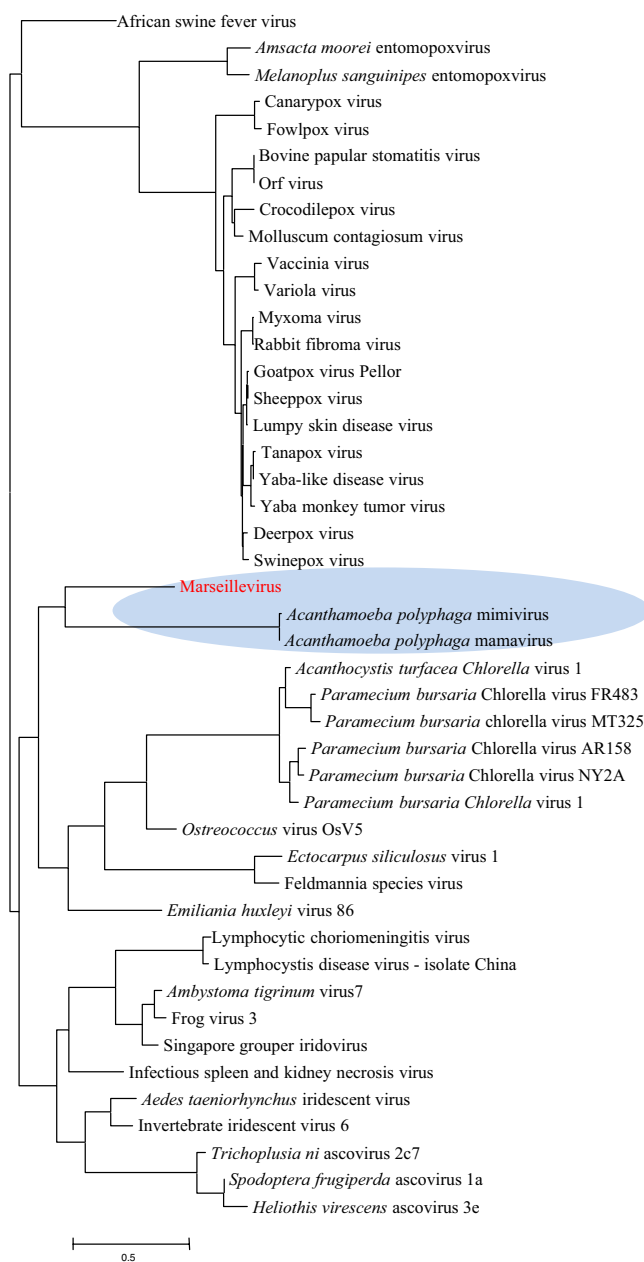


Fig. 4. Neighbor-joining clustering of NCLDV by phyletic pattern. The phyletic patterns of the orthologous sets of NCLDV genes (8, 9) indicating the presence/absence of the respective gene in each virus were used for the construction of the neighbor-joining tree (phylip3.66) after adding the Marseillevirus orthologs.

was obtained for a relatively small number of genes. Although the current repertoire of viruses and bacteria infecting amoebae is far from being complete, the relative paucity of genes of *Acanthamoeba* origin in Marseillevirus suggests that the virus might have changed hosts, perhaps more than once, in the course of its evolution. Indeed, a recent report indicates that relatives of the Mimivirus could infect marine animals such as sponges and corals (21).

Unlike most other host cells, amoebae are commonly infected by numerous, taxonomically diverse microorganisms (22). As we show in a direct experiment, amoeba cells can be simultaneously and productively infected with Marseillevirus and two bacterial parasites (Fig. 5 and *SI Text*). This promiscuity probably results in extended coexistence of multiple parasites and/or symbionts within the same amoeba and so might make the amoeba a veritable factory

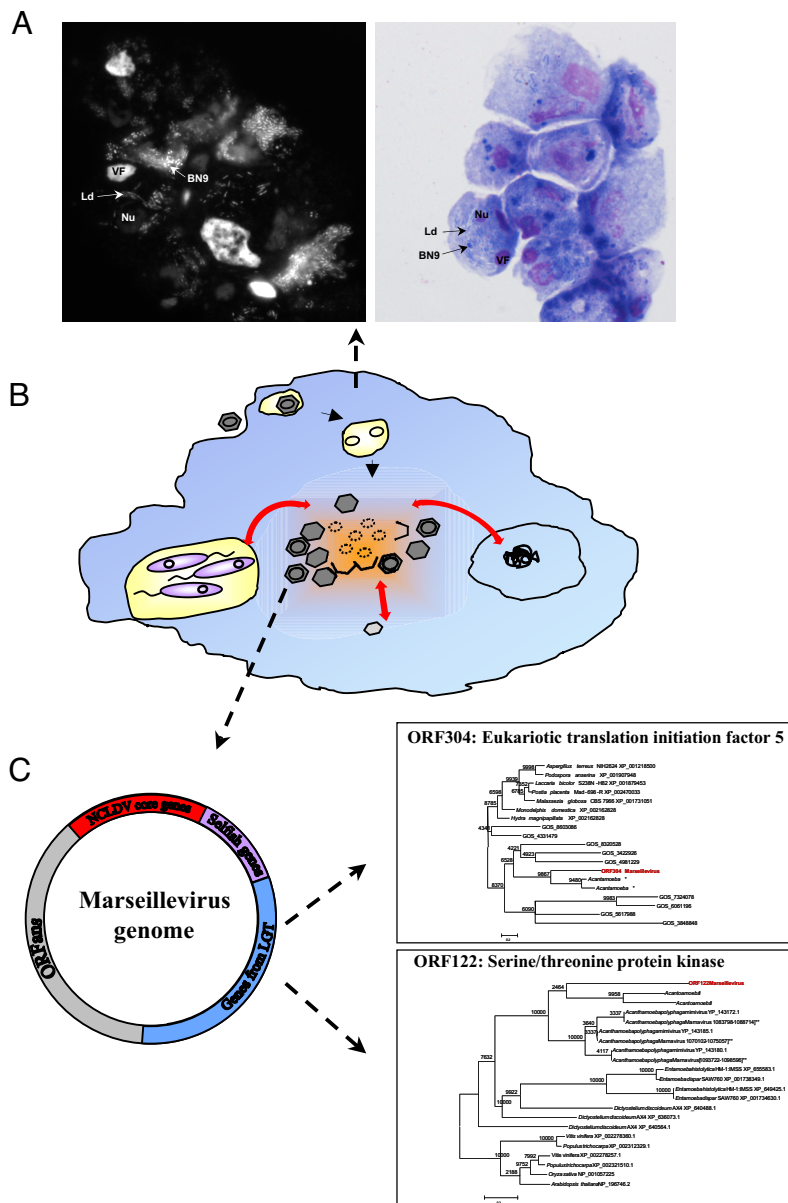


Fig. 5. (A) DAPI (Left) and Hemacolor (Right) staining of *A. castellanii* (nucleus, Nu) coinfecting with *Legionella drancourtii* (Ld), *Parachlamydia* strain BN9 (BN9), and Marseillevirus (VF). Amoeba cells containing the three microorganisms were observed at 16 h and 24 h p.i. The DAPI and Hemacolor-stained microorganisms were controlled by performing amoeba infection with each microorganism alone. Marseillevirus was detected by the characteristic morphology of its VF. (B) Schematic representation of multiple intracellular microorganisms (bacteria in purple, Marseillevirus in dark gray, and its VF in orange, and other viruses in light gray) infecting amoeba. Lateral gene exchange (red arrow) could occur during microorganism multiplication. (C) Schematic representation of Marseillevirus genome with some examples of gene probably acquired by lateral HGT. *Marseillevirus homolog sequence was detected in *Acanthamoeba castellanii* Neff draft genome and included in the phylogenetic studies (Fig. S4). **Numbers in brackets indicate the position of Marseillevirus homolog sequence in *Acanthamoeba polyphaga* Mamavirus genome.

for gene mixing between the eukaryotic host, its various viruses, and bacterial parasites and symbionts. The amoebal genetic melting pot seems to produce organisms with complex, chimeric genomes such as the giant viruses. The very preponderance of giant viruses in amoebae might be explained by the action of an HGT ratchet in the host's intracellular environment where viruses are constantly bargaged with DNA from diverse sources. The possibility that giant viruses shuttle between different eukaryotic hosts further expands the gene pool to which they are exposed. Given the diversity of phagocytic unicellular eukaryotes (23, 24), it seems certain that the discovery of Mimivirus, Mamavirus, and Marseillevirus is only the first narrow window into a wondrous world of giant viruses, some of which could be even bigger and more complex than the current record holder, Mamavirus.

Methods

Isolation. At the start of the study, metal pieces were introduced into a cooling tower located in Paris. One piece was removed weekly to monitor biofilm formation, together with water samples to check microbiological evolution. Adherent biofilm was homogenized into sterile water and filtered through a 0.22- μm -pore-sized filter. Water samples were filtered as well. Filters were then shaken into sterile Page's amoebal saline (PAS), and each suspension was inoculated onto *Acanthamoeba polyphaga* microplates, as previously described (7). Cocultures were screened for cytopathic effects at day 3, and subcultured onto fresh amoebal microplates. Marseillevirus was purified using the end-point dilution method.

Electron Microscopy and Immunofluorescence. Experiments were performed as previously described (25).

Cryo-EM. Marseillevirus particles were flash-frozen on holey grids in liquid ethane. Images were recorded at 39 K magnification with a CM200 FEG microscope with electron dose levels of $\approx 20 \text{ e}^-/\text{\AA}^2$. All micrographs were digitized at $3.175 \text{ \AA pixel}^{-1}$ using a Nikon scanner.

Sequencing and Analysis of Marseillevirus Genome. Marseillevirus genome was pyrosequenced on 454-Roche GSFLX as described (26). The raw data (6.3 Mbp) were assembled by the GSFLX gsAssembler. Protein-coding genes were predicted using GeneMark.hmm 2.0 (27). The translated protein sequences were searched against the NCBI Refseq and env_nr (environmental nonredundant) protein sequence databases using BLASTP (28). Conserved domains were identified by searching the Conserved Domain Database (CDD version 2.13) (29) using RPS-BLAST. Multiple alignments of protein sequences were constructed using MUSCLE (30). Similarity-based clustering of protein sequences was performed using BLASTCLUST with subsequent manual curation. Maximum-likelihood (ML) phylogenetic trees were constructed using TreeFinder (31). Detailed methods for genome analysis are provided in *SI Text*.

Proteomic Analysis. Experiments were performed as previously described (7).

RNA Extraction from Marseillevirus Virions and RT-PCR Analysis. Experiments were conducted as previously described (6). Specific primers used in this study were provided in *SI Text*.

ACKNOWLEDGMENTS. We are grateful to Bernard Campagna for his technical assistance with electron microscopy, Valorie D. Bowman for collecting data for the cryo-EM reconstruction of Marseillevirus, Claude Nappes for anti-Marseillevirus monoclonal antibody production, Bernadette Giumelli and Thi Tien N'Guyen for technical assistance in genome sequencing, Philippe de Cloquement for protein identification, Angélique Campocasso for help in detection of viral RNA, Kira Makarova and Yuri Wolf for help with sequence analysis, Michèle Merchat for providing water samples, and Christelle Desnues for reading the manuscript. This work was funded by the Centre National de la Recherche Scientifique (CNRS, credits récurrents), a Conventions Industrielles de Formation par la Recherche fellowship (I.P.), Intramural Research Program of the National Institutes of Health National Library of Medicine (N.Y. and E.K), and National Institutes of Health Grant AI11219 (to S.S. and M.G.R.).

- Lwoff A (1957) The concept of virus. *J Gen Microbiol* 17:239–253.
- Raoult D, Forrester P (2008) Redefining viruses: Lessons from Mimivirus. *Nat Rev Microbiol* 6:315–319.
- Van Etten JL, Meints RH (1999) Giant viruses infecting algae. *Annu Rev Microbiol* 53:447–494.
- Audic S, et al. (2007) Genome analysis of *Minibacterium massiliensis* highlights the convergent evolution of water-living bacteria. *PLoS Genet* 3:e138.
- La Scola B, et al. (2003) A giant virus in amoebae. *Science* 299:2033.
- Raoult D, et al. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306:1344–1350.
- La Scola B, et al. (2008) The virophage as a unique parasite of the giant mimivirus. *Nature* 455:100–104.
- Iyer LM, Aravind L, Koonin EV (2001) Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol* 75:11720–11734.
- Iyer LM, Balaji S, Koonin EV, Aravind L (2006) Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res* 117:156–184.
- Wilson WH, et al. (2005) Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. *Science* 309:1090–1092.
- Fitzgerald LA, et al. (2007) Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A. *Virology* 358:472–484.
- Gubbels MJ, Vaishnav S, Boot N, Dubremetz JF, Striepen B (2006) A MORN-repeat protein is a dynamic component of the *Toxoplasma gondii* cell division apparatus. *J Cell Sci* 119:2236–2245.
- Bai C, et al. (1996) SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* 86:263–274.
- Kobayashi I (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res* 29:3742–3756.
- Cheng CH, et al. (2002) Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the *Baculoviridae*. *J Virol* 76:9024–9034.
- Ibrahim AMA, Choi JY, Je YH, Kim Y (2005) Structure and expression profile of two putative *Cotesia plutellae* bracovirus genes (CpBV-H4 and CpBV-E94[alpha]) in parasitized *Plutella xylostella*. *J Asia Pacific Entomol* 8:359–366.
- Moreira D, Brochier-Armanet C (2008) Giant viruses, giant chimeras: The multiple evolutionary histories of Mimivirus genes. *BMC Evol Biol* 8:12.
- Filee J, Pouget N, Chandler M (2008) Phylogenetic evidence for extensive lateral acquisition of cellular genes by nucleocytoplasmic large DNA viruses. *BMC Evol Biol* 8:320.
- Wolf YI, Rogozin IB, Grishin NV, Koonin EV (2002) Genome trees and the tree of life. *Trends Genet* 18:472–479.
- Claverie JM, et al. (2006) Mimivirus and the emerging concept of “giant” virus. *Virus Res* 117:133–144.
- Claverie JM, et al. (2009) Mimivirus and Mimiviridae: Giant viruses with an increasing number of potential hosts, including corals and sponges. *J Invertebr Pathol* 101:172–180.
- Greub G, Raoult D (2004) Microorganisms resistant to free-living amoebae. *Clin Microbiol Rev* 17:413–433.
- Okada M, et al. (2005) Proteomic analysis of phagocytosis in the enteric protozoan parasite *Entamoeba histolytica*. *Eukaryot Cell* 4:827–831.
- Jacobs ME, et al. (2006) The *Tetrahymena thermophila* phagosome proteome. *Eukaryot Cell* 5:1990–2000.
- Suzan-Monti M, La Scola B, Barrassi L, Espinosa L, Raoult D (2007) Ultrastructural characterization of the giant volcano-like virus factory of *Acanthamoeba polyphaga* Mimivirus. *PLoS One* 2:e328.
- Margulies M, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
- Lukashin AV, Borodovsky M (1998) GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res* 26:1107–1115.
- Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
- Marchler-Bauer A, et al. (2009) CDD: Specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 37:D205–D210.
- Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
- Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18.