In-depth study of *Mollivirus sibericum*, a new 30,000-y-old giant virus infecting *Acanthamoeba*

Matthieu Legendre[a,1], Audrey Lartigue[a,1], Lionel Bertaux[a], Sandra Jeudy[a], Julia Bartoli[a,2], Magali Lescot[a], Jean-Marie Alempic[a], Claire Ramus[b,c,d], Christophe Bruley[b,c,d], Karine Labadie[e], Lyubov Shmakova[f], Elizaveta Rivkina[f], Yohann Couté[b,c,d], Chantal Abergel[a,3], and Jean-Michel Claverie[a,g,3]

[a]Information Génomique and Structurale, Unité Mixte de Recherche 7256 (Institut de Microbiologie de la Méditerranée, FR3479) Centre National de la Recherche Scientifique, Aix-Marseille Université, 13288 Marseille Cedex 9, France; [b]Université Grenoble Alpes, Institut de Recherches en Technologies et Sciences pour le Vivant–Laboratoire Biologie à Grande Echelle, F-38000 Grenoble, France; [c]Commissariat à l'Energie Atomique, Centre National de la Recherche Scientifique, Institut de Recherches en Technologies et Sciences pour le Vivant–Laboratoire Biologie à Grande Echelle, F-38000 Grenoble, France; [d]INSERM, Laboratoire Biologie à Grande Echelle, F-38000 Grenoble, France; [e]Commissariat à l'Energie Atomique, Institut de Génomique, Centre National de Séquençage, 91057 Evry Cedex, France; [f]Institute of Physicochemical and Biological Problems in Soil Science, Russian Academy of Sciences, Pushchino 142290, Russia; and [g]Assistance Publique–Hopitaux de Marseille, 13385 Marseille, France

Edited by James L. Van Etten, University of Nebraska, Lincoln, NE, and approved August 12, 2015 (received for review June 2, 2015)

Acanthamoeba **species are infected by the largest known DNA viruses. These include icosahedral Mimiviruses, amphora-shaped Pandoraviruses, and *Pithovirus sibericum*, the latter one isolated from 30,000-y-old permafrost. *Mollivirus sibericum*, a fourth type of giant virus, was isolated from the same permafrost sample. Its approximately spherical virion (0.6-µm diameter) encloses a 651-kb GC-rich genome encoding 523 proteins of which 64% are ORFans; 16% have their closest homolog in Pandoraviruses and 10% in *Acanthamoeba castellanii* probably through horizontal gene transfer. The Mollivirus nucleocytoplasmic replication cycle was analyzed using a combination of "omic" approaches that revealed how the virus highjacks its host machinery to actively replicate. Surprisingly, the host's ribosomal proteins are packaged in the virion. Metagenomic analysis of the permafrost sample uncovered the presence of both viruses, yet in very low amount. The fact that two different viruses retain their infectivity in prehistorical permafrost layers should be of concern in a context of global warming. Giant viruses' diversity remains to be fully explored.**

giant virus | permafrost | Pleistocene

Following the serendipitous discovery of Mimivirus, the first giant virus with particles large enough to be easily visible under a light microscope (1, 2), systematic surveys were launched to assess the diversity of these spectacular *Acanthamoeba*-infecting viruses in a planet-wide variety of environments. This led to the discovery and characterization of additional Mimivirus-like viruses now gathered into their own distinct family of DNA viruses, the *Mimiviridae*. They share a unique external fiber layer enclosing a pseudoicosahedral protein capsid of about 0.5 µm in diameter, itself containing lipid membranes surrounding an electron-dense nucleoid. Their genomes are made of an adenine-thymine A+T-rich linear dsDNA molecule up to 1.26 Mb in length predicted to encode up to 1,120 proteins (3), including a transcription apparatus allowing them to replicate in the host's cytoplasm (4, 5). The *Mimiviridae* family is still expanding (6) and diversifying with more distant and smaller representatives (both in terms of particle and genome size) that infect nonamoebal unicellular protists (7–9).

The search for additional *Acanthamoeba*-infecting viruses led to the discovery of the *Marseilleviridae*, now a rapidly growing family of large dsDNA viruses with icosahedral particles 0.2 µm in diameter and genome sizes in the 346- to 380-kb range (10–13). The next discovery was that of the spectacular Pandoraviruses isolated from two remote locations, central Chile (*Pandoravirus salinus*), and Melbourne, Australia (*Pandoravirus dulcis*) (14). Their amphora-shaped virions are 1.0–1.2 µm in length and 0.5 µm in diameter and exhibit a membrane-bound empty-looking compartment encased into a ~70-nm-thick tegument-like envelope. Their particles carry a linear G+C-rich dsDNA genome of 2.77 Mb for *P. salinus*, and 1.93 Mb for *P. dulcis*. The 2.24-Mb sequence of a third Pandoravirus genome was recently made available [*Pandoravirus inopinatum* (15)]. These genomes encode a number of predicted proteins comparable to that of the most reduced parasitic unicellular eukaryotes, such as encephalitozoon species (14). In contrast with *Mimiviridae*, Pandoraviruses' replication cycle involves (and disrupts) the host nucleus.

Searching for *Acanthamoeba*-infecting virus in increasingly exotic environments allowed the discovery of *Pithovirus sibericum* infectious particles, which were recovered from a sample of Late Pleistocene Siberian permafrost (16). Although Pithovirus's virions looked similar to those of Pandoraviruses both in terms of size and overall shape, further analyses indicated that the two

Significance

The saga of giant viruses (i.e. visible by light microscopy) started in 2003 with the discovery of Mimivirus. Two additional types of giant viruses infecting *Acanthamoeba* have been discovered since: the Pandoraviruses (2013) and *Pithovirus sibericum* (2014), the latter one revived from 30,000-y-old Siberian permafrost. We now describe *Mollivirus sibericum*, a fourth type of giant virus isolated from the same permafrost sample. These four types of giant virus exhibit different virion structures, sizes (0.6–1.5 µm), genome length (0.6–2.8 Mb), and replication cycles. Their origin and mode of evolution are the subject of conflicting hypotheses. The fact that two different viruses could be easily revived from prehistoric permafrost should be of concern in a context of global warming.

Author contributions: Y.C., C.A., and J.-M.C. designed research; M. Legendre, A.L., L.B., S.J., J.B., M. Lescot, J.-M.A., C.R., Y.C., C.A., and J.-M.C. performed research; K.L. performed sequencing; A.L., C.R., C.B., K.L., L.S., E.R., and Y.C. contributed new reagents/analytic tools; M. Legendre, A.L., M. Lescot, Y.C., C.A., and J.-M.C. analyzed data; M. Legendre, 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 Mollivirus genome sequence reported in this paper has been deposited in the GenBank database (accession no. KR921745). The transcriptomic data have been deposited in the Sequence Read Archive, www.ncbi.nlm.nih.gov/Traces/sra/ [accession no. SRX1078581 (SRR2084123 for the early class of expression, SRR2103267 for the intermediate one, and SRR2103268 for the late class of expression)]. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium, proteomecentral.proteomexchange.org, via the Proteomics Identifications partner repository [dataset identifiers <PXD002375> (Particule and Surfome) and <PXD002374> (Time Course)]. All data can be visualized on an interactive genome browser at the following link: www.igs.cnrs-mrs.fr/cgi-bin/gb2/gbrowse/Mollivirus/.

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

[2]Present address: Laboratoire d'Ingénierie des Systèmes Macromoléculaires, Unité Mixte de Recherche 7255 (Institut de Microbiologie de la Méditerranée, FR3479) Centre National de la Recherche Scientifique, Aix-Marseille Université, 13009 Marseille, France.

[3]To whom correspondence may be addressed. Email: Chantal.Abergel@igs.cnrs-mrs.fr, jean-michel.claverie@univ-amu.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1510795112/-/DCSupplemental.

types of viruses were unrelated (16). Pithovirus genome is a much smaller 600-kb circular A+T-rich dsDNA molecule predicted to encode only 467 proteins. In contrast with Pandoravirus, Pithovirus replicates in the host cytoplasm.

From the same permafrost sample, we isolated *Mollivirus sibericum*, the first representative (to our knowledge) of a fourth type of giant viruses infecting *Acanthamoeba*. Both transcriptomic and a detailed proteomic time course were used to analyze the infectious cycle of Mollivirus, which appeared markedly different from the previously described viruses infecting the same host. A metagenomic survey was performed to validate the presence and to quantify Pithovirus and Mollivirus in the original permafrost sample. Our results suggest that giant viruses are much more diverse than initially assumed and demonstrate that infectious viral particles with different replications schemes are present in old Siberian permafrost layers.

Results

Particle Morphology. Mollivirus was initially spotted using light microscopy as rounded particles multiplying in a culture of *Acanthamoeba castellanii* inoculated with a sample of Siberian permafrost from the Kolyma lowland region (*SI Methods*). After amplification, the particles were analyzed by transmission electron microscopy (TEM) and scanning electron microscopy. Mollivirus's roughly spherical particles are 500–600 nm in diameter and appear surrounded by a hairy tegument (Fig. 1*A*). By thin-section TEM, the particles appear to be surrounded by two to four 25-nm-spaced rings corresponding to fibers of different lengths (Fig. 2*D*). The tegument is made of at least two layers of different densities and structures. The external layer (10 nm thick) appears to form 30- to 40-nm-interspaced strips tangent to the surface of the particle (Figs. 1*B* and 2*B*). The internal layer is 12–14 nm thick and is made of a mesh of fibrils resembling those constituting the central layer of Pandoravirus's tegument (Fig. 1*C*) (14). On the surface of the Mollivirus particle, the genome-delivery portal coincides with a circular depression 160–200 nm in diameter (Fig. 1), which could be the consequence of the lack of fiber at the virion apex. At least one internal lipid membrane is delimiting the spacious inner compartment of the virion that is devoid of discernible substructures (Fig. 1 *B* and *C*).

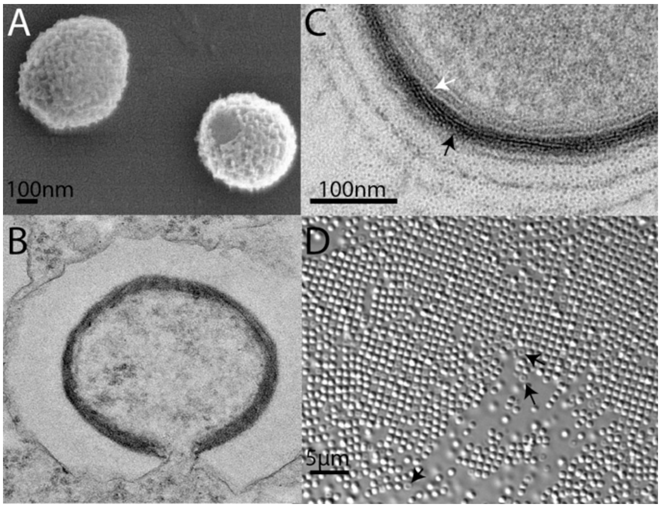

Fig. 1. Imaging of Mollivirus particles. (*A*) Scanning electron microscopy of two isolated particles showing the apex structure. (*B*) Transmission electron microscopy (TEM) imaging of an ultrathin section of an open particle after fusion of its internal lipid membrane with that of a phagosome. (*C*) Enlarged view of the viral tegument of a Mollivirus particle highlighting the layer made of a mesh of fibrils (black arrow), resembling Pandoraviruses' intermediate layer, and the underneath internal membrane (white arrow). Three ~25-nm interspaced rings are visible around the mature particle. (*D*) Light microscopy (Nomarski optics 63×) imaging of a lawn of Mollivirus particles, some of them (black arrow) exhibiting a depression at the apex.

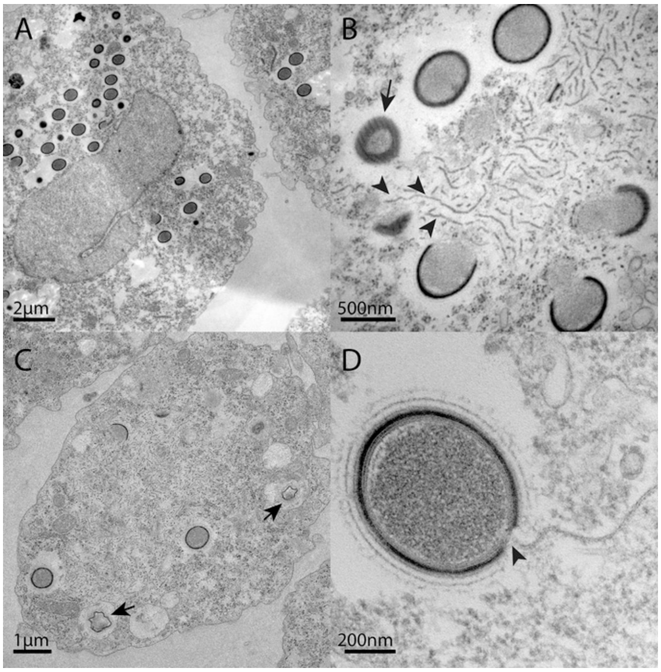

Fig. 2. Ultrathin-section TEM imaging of Mollivirus-infected *Acanthamoeba* cells. (*A*) Appearance of the nucleus 5 h PI. The nucleolus has almost vanished, filled with fibrillary structures of unknown composition, and the nuclear membrane presents invaginations. The nucleus is surrounded by Mollivirus particles at various stages of maturation. (*B*) Details of a virus particle assembly. Arrowheads point to fibrillary structures. A black arrow points to a section tangent to the virion surface revealing the tegument organization. (*C*) Overall view of the cell at a late stage of infection. Black arrows point to deformed mature virions that are reproducibly seen in vacuoles. A mesh of fibers fills the VF. (*D*) Mollivirus particle at a late assembly stage. The particle is crowned with several fuzzy rings, and different tegument layers are visible. At least one lipid membrane is lining the internal face of the virion tegument. One of the numerous fibers filling the VF is reproducibly seen associated with the apex of the maturing particle.

Replication Cycle. Mollivirus replication strategy was documented by following its propagation in axenic *A. castellanii* cultures over an entire multiplication cycle, starting from purified particles at a very high multiplicity of infection (MOI of 50) to warrant the synchronization of the infection. As for all previously described giant viruses infecting *Acanthamoeba*, the replication cycle begins with the phagocytosis of Mollivirus particles with up to 10 virions per cell, either distributed in individual vacuoles or gathered in the same vacuole. The opening of the particle was never clearly visualized due to the thickness of the ultrathin sections, larger than the dimension of the genome-delivery funnel. However, the fusion between the virion internal lipid membrane and the vacuole membrane was clearly observed (Fig. 1*B*). The release of the 5-ethynyl-2′-deoxyuridine (EdU)-labeled Mollivirus viral DNA into the cell cytoplasm and its migration to the nucleus was visualized using fluorescence microscopy (Fig. 3). The *Acanthamoeba* cells maintained their trophozoite shape and remained adherent throughout the whole cycle. The number of visible vacuoles started to decrease 4–5 h postinfection (PI), and neosynthesized virions appeared in the extracellular medium 6 h PI without exhibiting the cell lysis characterizing previously described giant viruses (2, 14, 16). As for Pandoraviruses, the cell

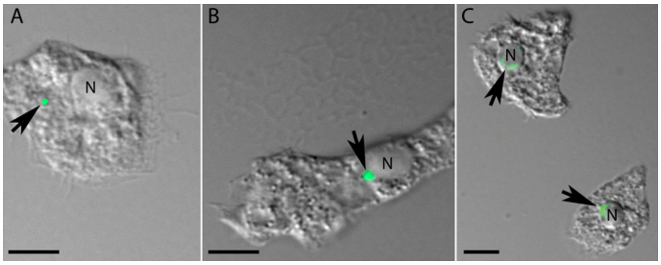

Fig. 3. Mollivirus EdU-labeled DNA visualized in infected *Acanthamoeba castellanii*. (*A*) Early transfer of the labeled viral DNA in the cell cytoplasm. (*B*) Viral DNA after migration next to the cell nucleus. (*C*) The labeled DNA seems to diffuse in the cell nucleus. Images were taken 30–90 min PI. The viral DNA is signaled by a black arrow, and the cell nucleus is identified by "N." (Scale bar: 5 μm.)

nucleus becomes disorganized with numerous invaginations of the nuclear membrane, but although the nucleus vanishes to be replaced by the Pandoravirus virion factory (VF), the synthesis of new Mollivirus virions occurs at the periphery of the persisting, albeit deformed, nucleus (Fig. 2*A*). Organelles are excluded from this area, which becomes filled with a mesh of fibrillary structures (Fig. 2 *B* and *C*), presumably corresponding to viral proteins composing the particles (Fig. 2*D*). The nucleus appears also filled with these fibrillary structures, making it an integral part of the Mollivirus VF (Fig. 2*A*).

The process by which virions are formed is reminiscent of Pandoraviruses, with the envelope and the interior of the Mollivirus particles being synthesized simultaneously (14) (Fig. 2*B*), but the genome delivery portal apex of Mollivirus particles appears to be formed last instead of first. After 6–8 h, particles at various stages of maturation may coexist in the same VF while mature virions are seen in vacuoles, suggesting that most, if not all, of them are released via exocytosis. Each cell seems to release few hundreds (200–300) new viral particles.

Mollivirus Genomic Features. Mollivirus genome is a linear dsDNA G+C-rich (60%) molecule of 651,523 bp, including a ∼10-kb-long inverted repeat at each extremity. In contrast to the 610-kb A+T-rich genome of Pithovirus, it is remarkably devoid of internal repeats (Fig. S1), hence making sequence assembly comparatively easier (16).

Protein-coding regions were predicted using Genemark (17), and the limits of the corresponding genes precisely mapped using transcriptome sequencing. Poly-A$^+$–enriched RNA were extracted from Mollivirus-infected *Acanthamoeba* cells 30 min to 9 h PI, and then used to build three different sequencing libraries by pooling three consecutive times roughly corresponding to the "early" (30 min and 1 and 2 h), "intermediate" (3, 4, and 5 h), and "late" transcripts (6, 7, and 9 h). Most reads (96–98%) could be mapped onto the Mollivirus or *Acanthamoeba castellanii* (18) genome sequences.

The above analyses identified 523 protein-coding genes (noted ml_#ORF number; ORF, open reading frame) and three tRNAs (Leu_{TTG}, Met_{ATG}, Tyr_{TAC}). There was no clear signal for the presence of non–protein-coding poly-A$^+$ transcripts and more than 90% of the predicted genes were associated to RNA-seq coverage values higher than that of the intergenic regions (Fig. S2). The total protein-coding moiety corresponds to 82.2% of the genome, 8.2% to the short 5′- and 3′-untranslated regions and 9.6% to the intergenic regions (120 nt long in average). The analysis of these intergenic regions did not reveal enriched sequence motifs that might indicate a conserved promoter signals. This negative finding is consistent with the lack of stringent transcriptional regulation exhibited by most Mollivirus genes, the transcripts of which were detected, albeit at various levels, in the early, intermediate, and late mRNA pools.

The mapping of the RNA-seq reads on the Mollivirus genome sequence pointed out the presence of short (87.6 ± 8.6 nt; min, 59; max, 159; median, 84) spliceosomal introns delimited by the canonical 5′-GT–3′-AG rule in 21 (4%) of the 523 protein-coding genes, evenly distributed along the genome. With the exception of ml_476 (with four introns of 87, 59, 70, and 83 nt) and ml_89 (with three introns of 67, 75, and 67 nt), there is a single intron per gene. All correspond to transcripts remaining detectable in the intermediate and late mRNA pools (Table S1), and two introns-containing genes (ml_476, ml_320) correspond to proteins found in the virion, suggesting that the host's spliceosome remains functional throughout the entire replication cycle, despite the morphological changes exhibited by the nucleus. Alternatively, the viral mRNAs may be stable enough to remain present all along the replicative cycle.

The sequences of the Mollivirus predicted proteins were analyzed using BLAST against the nonredundant protein sequence database (National Center for Biotechnology Information) (19) and a combination of motif search and protein-fold recognition methods [as previously described (16)]. As it is customary upon discovery of the first member of a previously unknown virus group, the proportion of Mollivirus protein without a recognizable homolog was high (337/523 = 64.4%). Among the Mollivirus proteins with homolog in the databases, 93 (17.8%) were most similar to a virus protein, 50 (9.6%) to an *A. castellanii* protein, 22 (4.2%) to proteins of other eukaryotes, and 18 (3.4%) to prokaryotic proteins (Fig. 4).

In contrast to Pithovirus (also the unique representative of its kind), a highly dominant proportion (83/93 = 89.2%) of the viral best matches to Mollivirus proteins correspond to a single known virus group, the Pandoraviruses (14). However, the Pandoravirus homologs exhibit low sequence similarities (identity, 40.4 ± 2.8%; median, 37.7%). Fifty-one (61.4%) of them display a recognizable

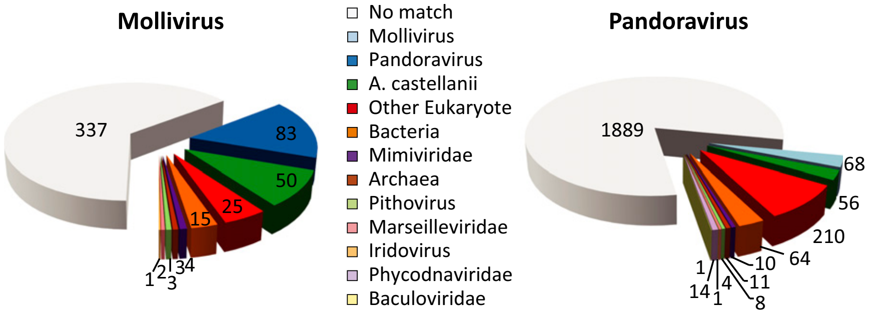

Fig. 4. Distribution of the best-matching homologs to Mollivirus and *Pandoravirus salinus* proteins. Best-matching homologous proteins were determined using BLASTP (E value < 10^{-5}) against the nonredundant (NR) database at the National Center for Biotechnology Information (19).

functional attribute, including 11 proteins containing generic conserved domains such as ankyrin repeats, BTB/POZ domains, or zinc fingers. Among the proteins with more specific functional predictions, we noted seven enzymes related to DNA metabolism: a B-type DNA polymerase (only 27.2% identical to that of *Pandoravirus dulcis*), a divergent primase (24% identical), two helicases, one exonuclease, one methylase, and one recombinase. Other predicted functions included four transcription components (RNA polymerase subunits RPB1, RPB2, and RPB10, and transcription elongation factor S-II), two serine/threonine protein kinases, one protein phosphatase, and a homolog to the cap-binding translation initiation factor 4E. The predicted virion packaging ATPase of Mollivirus is also remotely similar to that of Pandoraviruses (31.5% identical residues). The 50 Mollivirus proteins most similar to *Acanthamoeba castellanii* homologs are also quite divergent (47.7 ± 4.8% identity; median, 45.7%) and exhibit a large proportion of anonymous proteins (38/50 = 76%). Among the 12 proteins with a predicted function we noted an mRNA capping enzyme (35.3% identical), a dCMP deaminase (74.8% identical), a haem peroxidase (ml_395; 59.3% identical), a putative autophagy protein (49% identical), as well as two endonucleases. Curiously, the sole Mollivirus protein (ml_347) remotely similar to a major capsid protein was closest (62.6% identical) to a homolog encoded by *A. castellanii* (18). Table S2 lists the 110 Mollivirus predicted proteins (110/523 = 21%) that could be associated to a putative function. The dominant categories include proteins containing nonspecific protein–protein interaction motifs (e.g., Ankyrin repeats) followed by DNA-processing enzymes (e.g., DNA polymerase, primase, nuclease, helicase, etc.) as well as nucleotide biosynthesis such as a deoxycytidylate deaminase, a deoxyuridine 5′-triphosphate nucleotidohydrolase, a guanylate kinase, and two nucleotide diphosphate kinase homologs. However, several key DNA biosynthesis enzymes such as thymidylate synthase and thymidylate kinase usually found in large DNA viruses are not encoded by Mollivirus. Even more remarkably, Mollivirus appears to be the sole large DNA virus without its own ribonucleotide reductase, a key enzyme required for the synthesis of all deoxyribonucleotides.

Proteome of Purified Mollivirus Particles. The particles of the previously described giant viruses (14, 16, 20, 21) are all associated with a large number of proteins. The *Mollivirus sibericum* virion is no exception, with up to 230 proteins, each of them reliably detected by the identification of at least two different peptides using tandem mass spectrometry (*SI Methods*). Most of them (187) were detected in at least two out of three independent virion preparations (biological replicates) (Table S3*A*). Out of the 230 virion proteins, 136 (59%) are from Mollivirus and 94 (41%) from *Acanthamoeba*, a proportion twice as large as that found in *Pithovirus sibericum* virions (37/196= 18.9%) (16) or *Pandoravirus salinus* (56/266 = 21%) (14) using proteomic approaches of similar sensitivity. Among the Mollivirus-encoded proteins detected in the virions, 74 (54.4%) are ORFans and only 35 (25.7%) could be associated to functional or domain-only predictions (Table S3*A*), a proportion similar to that in the whole viral gene content. The seven components of virus-encoded transcription apparatus are conspicuously absent in the virion proteome (Table S2) confirming that the early stage of Mollivirus replication requires nuclear functions, as already suggested by the rapid migration of the Mollivirus genome into the nucleus (Fig. 3) and the morphological changes undergone by the host nucleus during the infectious cycle (Fig. 2*A*). The three most abundant virion proteins correspond to ORFans, followed by the ml_347 gene product, homolog to the major capsid protein found in all large icosahedral DNA viruses, although Mollivirus particles do not exhibit such symmetry. This protein was not among the seven proteins found to lie at the virion surface as probed by limited trypsin proteolysis of intact purified particles (highlighted in blue, Table S3*A*). Only 3 of the 22 Ankyrin-repeat containing proteins identified in the Mollivirus genome are part of the particle proteome, indicating that most of them are not structural proteins but might participate in intracellular interactions. The same is true for the three BTB/POZ domain-containing proteins. In contrast, six of the eight predicted oxidoreductases are detected in the particle (Table S2), most likely to counteract the oxidative stress encountered in the *Acanthamoeba* phagosome (21). A YjgF-like domain, putative translation inhibitor homolog (ml_79, ranking 11th) and a lipocalin (ml_287, ranking 12th) are among the proteins of functional interest in the particle, together with two enzymes that might participate to the glycosylation of the Mollivirus virion proteins (GlcNAc transferase ml_336, glycosyltransferase ml_353) labeled by standard in-gel glycoprotein detection kits.

In contrast to the Mollivirus-encoded proteins, a putative function could be predicted for 84 (89.4%) of the 94 *Acanthamoeba* gene products detected in the Mollivirus virions (Table S3*A*). The first host-derived protein ranks 55th in the particle proteome abundance list, a difference further reflected by the comparison of the whole abundance index distributions of the 136 Mollivirus proteins vs. the 94 *Acanthamoeba* proteins (65 of which were detected in 2/3 replicates) (Table S3*A*). The two distributions are significantly different (Kolmogorov–Smirnov, $P < 0.001$) with respective average values of $6.45 \pm 2.36 \times 10^{-3}$ vs. $1.07 \pm 0.16 \times 10^{-3}$. Assessing to which level of abundance low ranking *Acanthamoeba* gene products retain a functional significance will require further experimental studies. Pending these validations, three categories of functions appear to dominate the *Acanthamoeba*-derived moiety of the Mollivirus particle. The first one—and most unexpected—are 11 ribosomal proteins both from the small (S4, S7, S8, S9, S15, S23) and large subunits (L5, L6, L18, L30, L35) detected in two or all of our biological replicates (Table S3*B*). A total of 23 different ribosomal proteins (together with a ribosomal RNA assembly protein and the ribosome anti-association factor IF6) are detected to various extents of reproducibility, most likely due to a combination of their small size and low abundance (Table S3*B*). The second largest category is constituted of other mRNA binding/processing enzymes such as five helicases including a homolog of the cap-binding translation initiation factor 4A, and three other proteins involved in nuclear RNA processing and transport. Adding three histone homologs, HMG-like chromatin-associated proteins, and a homolog of the nuclear Es2 proteins, the Mollivirus particles incorporate a total of 13 different proteins normally confined to the host nucleus. The third most important category consists of *Acanthamoeba* gene products with similarity to actin, acting-binding or acting cross-linking proteins (profilins, actophorin, fascin, talin, etc.). This is a total of 11 proteins that might participate in the transport of the virion content to the nucleus through the reorganization of the host cytoskeleton (22).

We further investigated the location of the host-contributed proteins in the virion by comparing the peptides identified after limited proteolysis of intact particles to the fully digested particles. Our results (Table S3*A*) suggest that the detected *Acanthamoeba* proteins are most likely not simply associated to the particle surface and are thus presumably incorporated within the virions. These host-derived proteins are thus in a position to be involved in the early stage of the next infectious process.

Host–Virus Proteome Dynamics Throughout a Full Replication Cycle. For convenience, mRNA abundances measured by deep sequencing are widely used as proxies for protein abundances, even though weak correlations have often been demonstrated between the two measures (23). Indeed, what happens within a cell at a given time is a direct consequence of protein abundances, not of the levels of their cognate transcripts. In this study, we directly examined the variation of expressed functions and proteins occurring in *A. castellanii* infected by Mollivirus by performing a series of proteome analyses at regular intervals throughout the whole virus replication cycle. First, the quantified abundances of proteins were analyzed globally, mixing Mollivirus-encoded, mitochondrion-encoded, and host's nucleus-encoded proteins (Fig. S3).

As expected, a small peak of viral proteins was detected at 30 min PI, corresponding to the most abundant proteins of the Mollivirus virions detectable in the host after internalization. The relative abundance of neosynthesized viral proteins then increased steadily over time, as to represent about 16% of the total protein content of the virus–host system 6 h PI when the first virions are synthesized, and 23% 9 h PI. Symmetrically, the relative abundance of the host-encoded protein linearly decreases over time, a pattern consistent with the release of neoformed particles through exocytosis (i.e., preserving the host cell integrity) rather than through cell lysis. Interestingly, the relative abundance of mitochondrion-encoded proteins followed a parallel pattern, suggesting that these ATP-producing organelles are neither activated nor specifically degraded during the Mollivirus infectious cycle.

Given the above finding that no component of the Mollivirus-encoded transcription apparatus was detected in the virion proteome, we expected that its replication cycle should exhibit two markedly different phases during which viral genes are initially transcribed in the nucleus by the host apparatus before being taken over by the virus-encoded apparatus. Such a shift is well illustrated in Fig. 5, showing the respective abundances of the DNA-dependent RNA polymerase main subunits (RPB1, RPB2) of the host and virus, during the first 6 h PI. Before 4 h PI, the Mollivirus-encoded RPB1 and RPB2 proteins remain undetected, and then appear and maintain the same abundance throughout the rest of the replication cycle. This indicates that all viral proteins newly produced before 3–4 h PI are the products of genes transcribed by the host transcription machinery, presumably within the intact amoeba nucleus. Once established, the abundances of the Mollivirus RPB1 and RPB2 proteins remain mostly unchanged during the rest of the replication cycle, whereas those of their cellular counterparts, initially present at the same level, begin to decrease after 4 h simultaneously to the morphological change exhibited by the nucleus (Fig. 2*A*). Noticeably, the viral DNA polymerase is readily detected 3 h PI, whereas the level of its cellular homolog remains very low (Fig. 5).

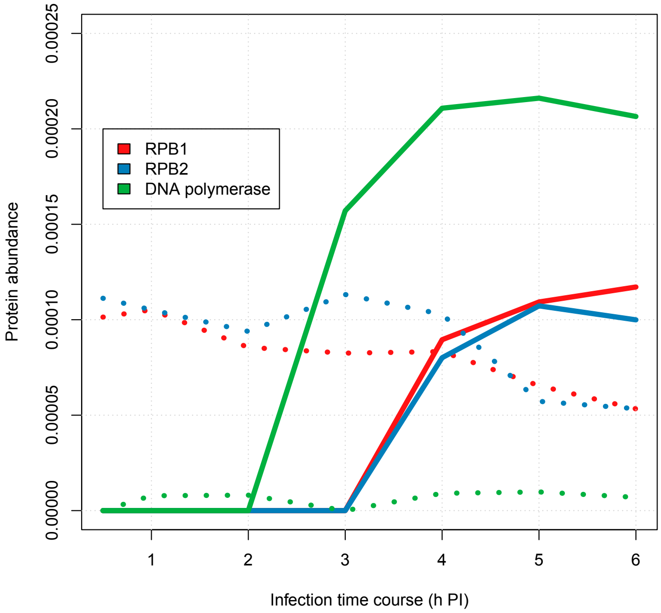

Fig. 5. Abundance of key DNA transcription and replication enzymes at various times PI. The variation in protein abundances [label-free quantification (LFQ)] (30) are plotted with solid lines for the Mollivirus proteins and in dashed lines for their cellular homologs. The quantification method used is described in *SI Methods*.

This coarse analysis of Mollivirus infectious cycle was refined by analyzing the abundance of all of the virus-encoded proteins at various times PI (from 30 min to 6 h). A clustering algorithm (*SI Methods*) was used to help visualize discriminating patterns in the resulting heat map (Fig. S4) built with all of the proteins reliably detected at two or more time points.

The largest set of proteins corresponds to those exhibiting a peak of production between 3 and 4 h PI. They correspond to "intermediate" genes presumably transcribed by the neosynthesized virus-encoded machinery (although the cellular one may still be functional). Proteins belonging to this group include the virus mRNA capping enzyme (ml_416) and its cap-binding translation initiation factor 4E (ml_363) amid a large number of proteins of unknown functions. Before this phase, a group of "early" proteins involved in the replication of DNA (e.g., DNA polymerase: ml_318; various helicases: ml_266, ml_359, ml_385) becomes detectable 1 h before (i.e., 3 h PI). Consistently, the few Mollivirus-encoded enzymes predicted to participate in the synthesis (or salvage) of nucleotides becomes detectable at the same time, or 1 h before (e.g., nucleotide diphosphate kinase: ml_233; dihydropholate reductase: ml_37; deoxyuridine 5′-triphosphate nucleotidohydrolase: ml_29; guanylate kinase: ml_103). A fourth well-delineated pattern (Fig. S4, *Bottom*) characterizes a group of 48 proteins brought in by the infecting virus particles. Their abundance decreases steadily, as they are presumably degraded in the host cell phagosome, until 4 h PI. Their increase signs the synthesis of new viral particles and the end of the replication cycle. The 11 proteins detected 1 h PI, but not found in the particle proteome, must be transcribed before that time. This can only be achieved if the Mollivirus genome can quickly reach the nucleus in the 30 min following its translocation from the phagosome to the cytoplasm (Fig. 3). Only three of these early proteins have a predicted function (ml_25 is a nuclease, ml_29 is a dUTPase, and ml_114 is a serine/threonine protein kinase).

We then investigated the influence of Mollivirus infection on the expression of *Acanthamoeba* proteins. We first noticed that, among 2,474 different *Acanthamoeba* proteins reliably detected in uninfected cells and at least two time points PI, 2,406 exhibited less than a twofold change in their relative abundance at any point in time during the first 6 h PI (Fig. 6). Thirty of them appear "up-regulated" and 38 "down-regulated." This first result indicates that Mollivirus infection preserves the global integrity of the host cell even after the apparent disruption of its nucleus. This is consistent with our observation (Fig. 2) that the infected cells can support the production of Mollivirus virions for several hours, shedding virions in the surrounding medium without drastic impairment of their viability. Focusing on proteins exhibiting more than a twofold change in abundance (Fig. 6 and Table S4), the most dramatic (about 16× fold) increase corresponds to the largest subunit of the diphosphate ribonucleotide reductase, an enzyme absolutely required for DNA synthesis, conspicuously absent from the viral genome. The abundance profile of this cellular enzyme steadily increases from 2 to 5 h PI (Fig. 6), in phase with that of the viral DNA polymerase (Fig. 5). We also noticed that three of the host proteins exhibiting more than a twofold increase in abundance ended up associated with the Mollivirus particle: an autophagy protein that might be involved in intracellular membrane reorganization along with an H2A core histone paralog and a high mobility group box domain-containing protein that might be involved in DNA packaging. On the other end, the list of proteins exhibiting a significant decrease in abundance contains a variety of enzymes without clear functional relationship. The most important decreases concern a haem peroxidase and a monoamine oxidase corresponding to two adjacent genes (Table S4). Interestingly, the virus encoded haem peroxidase (ml_395) starts accumulating when the host enzyme reaches a minimal level (3 h PI). Intriguingly, the *Acanthamoeba* proteins associated to

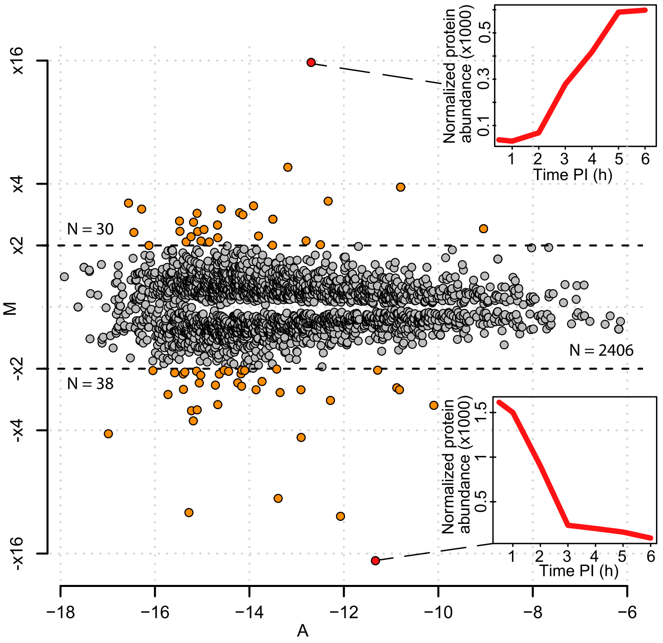

Fig. 6. Maximal variation in abundance of host proteins in Mollivirus-infected *Acanthamoeba* cells. Protein abundances (LFQ; *SI Methods*) were normalized by the total abundance of all identified *A. castellanii* proteins (including the mitochondrial ones) at each time point. These normalized abundances were used to construct an *MA* plot (31). The *M* axis corresponds to the binary logarithm of the ratio of the abundance measured 30 min PI and at the time point corresponding to the largest variation (relative to 30 min PI). The *A* axis corresponds to the average between those two values. Gray points correspond to proteins showing a maximal variation of less than twofold, whereas orange points have a maximal variation of more than twofold (Table S4). The abundance profiles (*Top* and *Bottom Insets*) throughout the entire infectious cycle are shown for the protein exhibiting the largest increment (ribonucleoside diphosphate reductase) and the largest decrease (a peroxidase) (red points).

the Mollivirus particles exhibit a variety of abundance profiles, many of which are not progressively increasing over time.

Direct Detection of Pithovirus and Mollivirus DNA in the Original Permafrost Sample. *Mollivirus sibericum* and the previously described *Pithovirus sibericum* (16) have been isolated from the same 30,000-y-old permafrost sample using a similar *Acanthamoeba* co-cultivation protocol. To demonstrate their presence and measure their relative abundance, we sequenced DNA directly extracted from the original permafrost sample, in search for cognate sequences. Out of 368,474,026 100-nt pair-ended reads generated on an Illumina platform, 336 and 125 could be mapped (>92.5% identity) on the Mollivirus and Pithovirus genome sequence, respectively (Table 1). As a control for an eventual cross-contamination, we looked for the presence of reads matching the genome of the modern giant viruses routinely cultivated in the laboratory. A total of only seven (most likely spurious) reads were found to match: Mimivirus (four reads, <92% identical), *Pandoravirus salinus* (one read, 96% identical), *Pandoravirus dulcis* (one read, 85% identical), and *Megavirus chilensis* (one read, 66% identical). Although the mapped reads only covered 4.8% and 2% of the Mollivirus and Pithovirus genomes, respectively, their distributions are quite uniform (Fig. 7). Interestingly a coverage of 3.6% was found for the 46.7 million-bp haploid genome of *A. castellanii* (27,894 mapped reads), indicating the presence at a very low—but similar—abundance level of the viruses host in the permafrost sample. More precisely, these respective coverages values correspond to a ratio of 1.1 Pithovirus and 2.7 Mollivirus virions per *A. castellanii* cell in the original sample.

Discussion

Using *Acanthamoeba castellanii* as bait, we isolated a new type of giant DNA virus from the same sample of 30,000-y-old permafrost from which we recently characterized *Pithovirus sibericum* (16). Although this virus, named *Mollivirus sibericum*, again exhibits a nonicosahedral ovoid particle, its nucleus-dependent mode of replication, genome organization, and gene content definitely indicate that it does not belong to the Pithovirus (proposed) family nor to the Iridovirus and Marseillevirus families with which *Pithovirus sibericum* exhibit a weak phylogenetic affinity (Fig. 8). Instead, and quite unexpectedly given their differences in morphologies and virion and genome sizes, *Mollivirus sibericum* phylogenetically clusters as a distant relative of the giant Pandoraviruses (14, 15). However, it is not yet clear if this phylogenetic position is due to a truly ancestral relationship of *Mollivirus sibericum* with the Pandoraviruses, or to the insertion of 83 Pandoravirus-derived genes into an otherwise unrelated Mollivirus genomic framework, unusually prone to horizontal gene transfers as also suggested by the presence of 50 putatively *Acanthamoeba*-derived genes (Fig. 4). The latter hypothesis is consistent with the lack of detectable colinearity between Mollivirus and the Pandoraviruses genomes (Fig. S5). We also noticed that, out of the 136 viral proteins in Mollivirus virions, only 28 have a homolog in Pandoravirus particles. Moreover, these pairs of homologous proteins exhibit highly discrepant abundances in their respective virion proteomes and the sets of the 20 most abundant virion proteins in Mollivirus and *P. salinus* (representing more than 60% of the total protein abundance) do not overlap. This suggests that the basic structures of the two particles are very different despite their common ovoid shape. The virion proteins shared by Mollivirus and Pandoravirus might thus be involved in host-specific interactions rather than bona fide structural features. The isolation and characterization of additional independent *Mollivirus sibericum* relatives will permit to assess the diversity of this new candidate family and the robustness of its phylogenetic relationship with the Pandoraviruses. It is worthwhile to notice that, to the best of our knowledge, no previous sighting of a Mollivirus-like endoparasite have been reported in the past literature, in contrast with Pithovirus (16, 24). Until *Mollivirus sibericum* relatives are isolated

Table 1. Metagenomic data statistics

Reference	Genome length, bp	No. mapped reads	Total mapped read length, bp	Genome coverage, %
A. castellanii	46,714,639	27,894	1,693,714	3.6
Mollivirus sibericum	651,523	336	31,081	4.8
Pithovirus sibericum	610,033	125	12,123	2
Mimivirus	1,181,549	4	369	0
Megavirus chilensis	1,259,197	1	66	0
Pandoravirus salinus	1,908,524	1	96	0
Pandoravirus dulcis	2,473,870	1	85	0

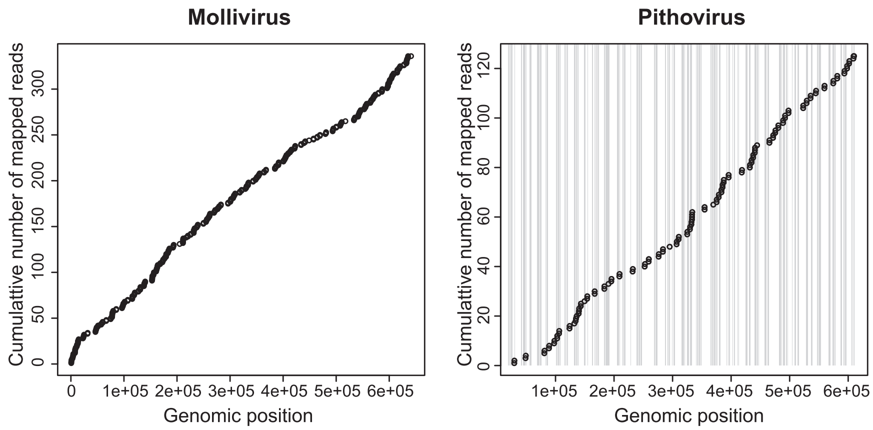

Fig. 7. Cumulative mapping of the metagenomic reads on the Mollivirus and Pithovirus genomes. Cumulative distribution of the 336 and 125 100-nt metagenomics reads that could be mapped (>92.5% identity) on the Mollivirus and Pithovirus genome sequence, respectively. Although only 4.8% of the Mollivirus genome and 2% of the Pithovirus genome are covered, the mapped read distributions are quite uniform, consistent with the presence of the whole viral genomes in the DNA mixture extracted from the permafrost sample. The vertical bars correspond to the multiple regularly interspersed copies of the noncoding repeat scattered along the Pithovirus genome (16). These repeats do not coincide with a local increase in genome coverage by metagenomics reads.

in contemporary environments, we cannot rule out that the permafrost was the only reservoir left for this viral family.

Our characterization of *Mollivirus sibericum* extensively relied on detailed proteomic analyses of both the particle and its replication cycle. The proteome of the particle revealed two main features: the absence of an embarked transcription apparatus and the unusual presence of many ribosomal (and ribosome-related) proteins. The first feature is consistent with the early migration of the viral genome to the nucleus (Fig. 3), the host nucleus morphological changes, the perinuclear location of the virion factory (Fig. 2), and the presence of introns in some viral genes (Table S1), all suggesting that the early replication stages of Mollivirus requires nuclear functions.

The large number of ribosomal proteins detected within Mollivirus particles is most unusual, and to our knowledge unique among

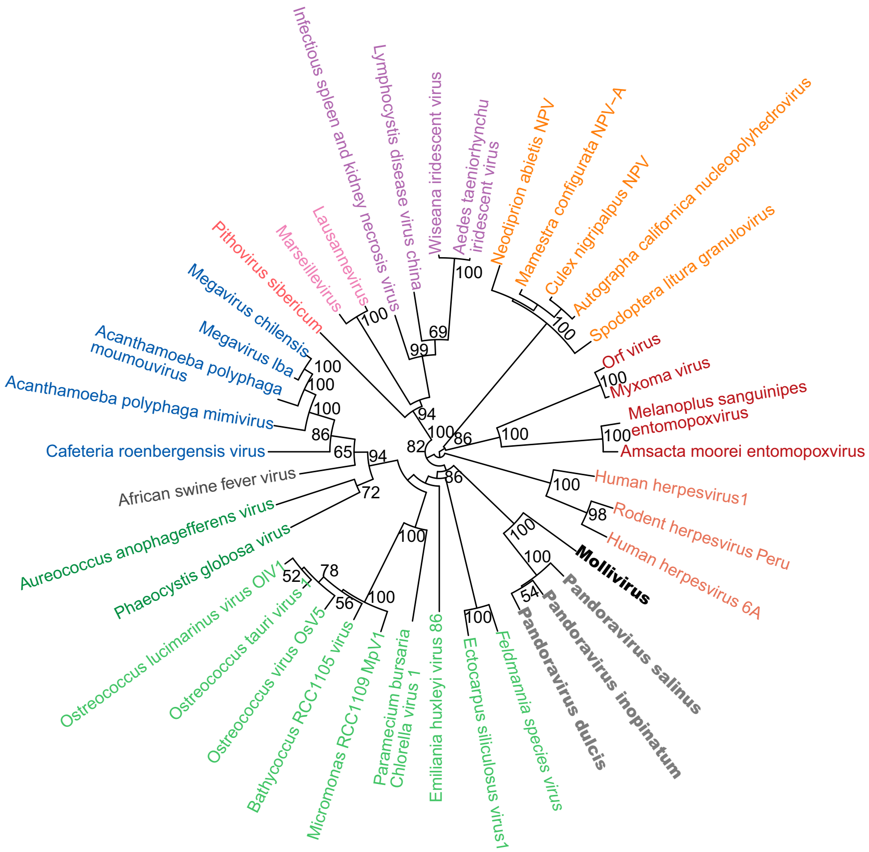
Infectious spleen and kidney necrosis virus
Lymphocystis disease virus china
Wiseana iridescent virus
Aedes taeniorhynchu iridescent virus
Neodiprion abietis NPV
Mamestra configurata NPV–A
Culex nigripalpus NPV
Autographa californica nucleopolyhedrovirus
Spodoptera litura granulovirus
Orf virus
Myxoma virus
Melanoplus sanguinipes entomopoxvirus
Amsacta moorei entomopoxvirus
Human herpesvirus1
Rodent herpesvirus Peru
Human herpesvirus 6A
Mollivirus
Pandoravirus salinus
Pandoravirus inopinatum
Pandoravirus dulcis
Feldmannia species virus
Ectocarpus siliculosus virus1
Emiliania huxleyi virus 86
Paramecium bursaria Chlorella virus 1
Micromonas RCC1109 MpV1
Bathycoccus RCC1105 virus
Ostreococcus virus OsV5
Ostreococcus tauri virus 1
Ostreococcus lucimarinus virus OlV1
Phaeocystis globosa virus
Aureococcus anophagefferens virus
African swine fever virus
Cafeteria roenbergensis virus
Acanthamoeba polyphaga mimivirus
Acanthamoeba polyphaga moumouvirus
Megavirus Iba
Megavirus chilensis
Pithovirus sibericum
Marseillevirus
Lausannevirus
100
69
99
100
100
94
100
100
100
86
65
94
82
100
86
100
100
100
100
100
100
98
86
100
100
54
100
72
100
78
52
56
100
100

Fig. 8. Cladistic clustering of representatives of the main families of large and giant DNA viruses infecting eukaryotes. The phylogenetic tree was constructed using neighbor joining on distances computed from a presence/absence matrix defined on 3,001 distinct genes clusters (*SI Methods*). Support values were estimated using bootstrap resampling (*n* = 10,000) and indicated when >50%.

giant (14, 16, 20, 21) or more conventional DNA viruses for which proteomic studies are available (25). The sole other documented association of virus particles and ribosomes acquired from their host concerns the Arenaviruses, a single-stranded RNA virus family, that do not have any phylogenetic relationship with Mollivirus. In contrast with Arenavirus, TEM images of Mollivirus virion do not exhibit the grainy appearance of ribosomal particles. In addition, our attempts to reveal the presence of ribosomal RNA in purified Mollivirus particles have remained unsuccessful. This suggests that the ribosomal (and ribosome-related) proteins we detected are not from intact ribosomes, but may originate from the nucleolus in the vicinity of which part of the virion assembly takes place. The fact that ribosomal proteins were never identified in other viruses such as Pandoraviruses or Chlorella viruses installing their virion factory in the host nucleus is puzzling and raises the question whether these ribosomal proteins are mere bystanders or play a role in the Mollivirus infectious process.

We then explored the course of an entire infectious cycle using quantitative proteomics to examine the nature and dynamics of virus–host interactions. At the most global level, the relative proportions of Mollivirus-, mitochondrion-, and *Acanthamoeba*-encoded proteins were found to vary rather smoothly, consistently with an infectious pattern preserving the cellular host integrity as long as possible and with the release of neoformed particles through exocytosis. The relative abundance of mitochondrion-encoded proteins followed a pattern parallel to that of other proteins, suggesting that these ATP-producing organelles are neither activated nor specifically degraded, during the first 6 h PI. Only 30 and 38 host proteins (1.45% and 1.53%, respectively, of 2,474 total protein seen at least in two time points) were found to be significantly up-regulated or down-regulated. None of them are homologous to innate immunity response related proteins seen to be up-regulated in mammalian cells undergoing a cytomegalovirus infection (26). An unexpected finding of the proteomic time course was that the synthesis of the virus-encoded DNA polymerase appeared to precede that of the viral transcription machinery, also shown to coexist with that of the host (Fig. 5). This suggests that the transcription of Mollivirus intermediate and late genes (i.e., 3 h PI) might proceed from both the original and neosynthesized DNA molecules and be simultaneously performed by the viral and host-derived transcription apparatus.

Mollivirus sibericum is now the third type of nonicosahedral giant virus discovered in less than 3 y using *A. castellanii* as a model host (14–16). This suggests that this morphotype is probably not rare and predicts that many more are to be found, including some that might have been misidentified as uncultivable bacteria in the context of human and animal diseases. These three types of nonicosahedral viruses have several key characteristics in common, despite exhibiting no or little phylogenetic affinities:

i) Their large, empty-looking virions enclose large dsDNA genomes not compacted in electron-dense nucleoids. With its 650-kb genome occupying a very large virion core of about 86 10^6 nm^3, Mollivirus exhibits a low DNA packing density (0.0075 bp/nm^3), a characteristic feature of the previously described Pandoraviruses and Pithovirus. The DNA-packaging mechanisms used by these giant viruses remain to be elucidated, as well as the molecular 3D structure of the viral genome inside the particle.

ii) All three types of virus particles are lined by an internal lipid membrane and use the same mechanism to infect their *Acanthamoeba* host: phagocytosis followed by the opening of a delivery portal, fusion of the internal virion membrane with the phagosome membrane, and delivery of the particle content in the cytoplasm.

iii) All of these giant viruses exhibit large proportions (>2/3) of encoded proteins without homologs, even between each other, raising the question on the origin of the corresponding genes, or of the mechanisms by which such diverse gene repertoire could be generated. With the exception of Pithovirus, the genomes of Mollivirus and Pandoraviruses do not exhibit mobile elements or repeated structures known to promote genomic instability.

On the other hand, these three types of giant nonicosahedral viruses exhibit marked differences: (*i*) G+C-rich (Mollivirus, Pandoraviruses) or A+T-rich (Pithovirus) genomes; (*ii*) linear (Mollivirus, Pandoraviruses) or circular (Pithovirus) genomes; (*iii*) high variability in genome sizes (600 kb for Mollivirus and Pithovirus, up to 2.8 Mb for Pandoraviruses); and (*iv*) their replication mode is either nucleocytoplasmic (Mollivirus, Pandoraviruses) or entirely cytoplasmic (Pithovirus).

Such different features in giant viruses infecting the same *Acanthamoeba* host support our previous suggestion that Pandoravirus-like particles might be associated to a diversity of viruses as large as that associated with icosahedral capsids in terms of evolutionary origins, genome size, or molecular nature (DNA or RNA) (14, 16).

Finally, our finding that two different viruses infecting the same host could be revived from a single permafrost sample, definitely suggests that prehistory “live” viruses are not a rare occurrence. Furthermore, the roughly equal representation of the two viruses in the metagenomics data suggests that there is no difference in the survival capacity of particles of either cytoplasmic (Pithovirus) or nucleus-dependent viruses (Mollivirus). Such modes of replication also correspond to the Poxvirus and Herpesvirus families, respectively. Although no read sequences were close enough to detect known Poxvirus and Herpesvirus isolates in the metagenome of our permafrost sample, we cannot rule out that distant viruses of ancient Siberian human (or animal) populations could reemerge as arctic permafrost layers melt and/or are disrupted by industrial activities.

Methods

***Mollivirus sibericum* Isolation and Production.** Mollivirus was isolated from a piece of the buried soil sample P1084-T as previously reported (16). Four hundred milligrams of P1084-T were resuspended in 6 mL of Prescott and James medium (27), and then used for infection trials of *A. castellanii* (Douglas) Neff (ATCC 30010TM) cells adapted to resist Fungizone. Cultures presenting an infected phenotype were recovered, centrifuged for 5 min at 500 × *g* to remove the cellular debris, and used to infect T-75 tissue culture flasks plated with fresh *Acanthamoeba* cells. After a succession of passages, viral particles produced in sufficient quantity were recovered and purified. See *SI Methods* for details.

Genome Sequencing, Assembly, and Annotation. Five hundred nanograms of purified genomic DNA were sheared to a 150- to 700-bp range using the Covaris E210 instrument (Covaris) prior Illumina library preparation using a semiautomatized protocol. Briefly, end repair, A-tailing, and ligation of Illumina compatible adaptors (Bioo Scientific) were performed using the SPRIWorks Library Preparation System and SPRI TE instrument (Beckman Coulter), according to the manufacturer’s protocol. A 300- to 600-bp size selection was applied to recover most of the fragments. DNA fragments were amplified by 12 cycles of PCR using Platinum Pfx Taq Polymerase Kit (Life Technologies) and Illumina adapter-specific primers. Libraries were purified with 0.8× AMPure XP beads (Beckman Coulter). After library profile analysis by Agilent 2100 Bioanalyzer (Agilent Technologies) and quantitative PCR quantification, the libraries were sequenced using 151 base-length read chemistry in paired-end flow cell on the Illumina MiSeq (Illumina). About 3 M useful pair-end reads were obtained. *SI Methods* contains further details on the bioinformatic genome assembly and annotation procedures.

Metagenomic Study. DNA was extracted from 0.52 and 0.242 g of the 1084T permafrost sample using the PowerSoil DNA isolation kit (Mo Bio) following the manufacturer’s protocol except that we added 83 mM DTT to the second sample to permit a more effective lysis of the viral particles. We recovered respectively 744 ng and 1.12 μg of pure DNA (Qubit). *SI Methods* contains further details on the library preparation and bioinformatic analysis of the data.

Transcriptomic and Proteomic Samples Preparation. Adherent cells were infected by Mollivirus at a MOI of 50 and distributed in 30 flasks (1.4 × 10^7 cells/flasks of 175 cm^2) containing 20 mL of protease peptone–yeast extract–glucose and left at 32 °C for 30 min, before removing excess viruses. For each time point, 12 mL were recovered to make three pools (1: 30 min, 1, 2 h; 2: 3, 4, 5 h; 3: 6, 7, 9 h) for transcriptomic analysis, 3 mL for the quantitative temporal proteomic study and 3 mL for inclusions and TEM observations. Each mRNA pool was sequenced on the Illumina MiSeq platform leading to 61 million, 71 million, and 71 million paired-ended 100-nt reads, respectively, of which 96–97.8% could be mapped on the Mollivirus or *A. castellanii* (18) genome sequences. For proteomic study of the infection, all time points were analyzed independently. *SI Methods* contains further details on sample preparations.

Proteome Analyses. For particle proteome and infectious cycle analyses, proteins were extracted in gel loading buffer and heated for 10 min at 95 °C. Proteins were stacked in the top of a 4–12% (wt/vol) polyacrylamide gel and in-gel digested before nano–liquid chromatography (LC)-MS/MS analyses of resulting peptides. For surfome analyses, purified virions were incubated for 30 min in digestion buffer (50 mM Tris·HCl, pH 7.5, 150 mM NaCl, and 5 mM $CaCl_2$) with or without trypsin (control). After centrifugation, supernatants were digested overnight with trypsin and resulting peptides analyzed by nano–LC-MS/MS. Peptides and proteins were identified using Mascot (Matrix Science) and IRMa (28) (version 1.31.1) for particle and surface proteomes and identified and quantified using MaxQuant (29) for infectious cycle analysis. Detailed procedures are presented in *SI Methods*.

ACKNOWLEDGMENTS. We thank Dr. J.-P. Chauvin, Dr. A. Kosta, F. Richard, and A. Aouane and Serge Nitsche for their expert assistance on the imaging platforms, Dr. Dorothée Murat for providing some transmission electron microscopy (TEM) images, and Miguel Ortiz Lombardia for his thorough reading of the manuscript. This work was partially supported by France Génomique Grant ANR-10-INSB-01-01, French National Research Agency Grant ANR-14-CE14-0023-01, the Provence-Alpes-Côte-d'Azur Région (2010 12125), ProFi Grant ANR-10-INBS-08-01, and Russian Scientific Fund 14-14-01115.

1. La Scola B, et al. (2003) A giant virus in amoebae. *Science* 299(5615):2033.
2. Raoult D, et al. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306(5700):1344–1350.
3. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie J-M (2011) Distant Mimivirus relative with a larger genome highlights the fundamental features of *Megaviridae*. *Proc Natl Acad Sci USA* 108(42):17486–17491.
4. Claverie J-M, Abergel C (2009) Mimivirus and its virophage. *Annu Rev Genet* 43:49–66.
5. Mutsafi Y, Zauberman N, Sabanay I, Minsky A (2010) Vaccinia-like cytoplasmic replication of the giant Mimivirus. *Proc Natl Acad Sci USA* 107(13):5978–5982.
6. Yoosuf N, et al. (2012) Related giant viruses in distant locations and different habitats: *Acanthamoeba polyphaga moumouvirus* represents a third lineage of the *Mimiviridae* that is close to the Megavirus lineage. *Genome Biol Evol* 4(12):1324–1330.
7. 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(45): 19508–19513.
8. Santini S, et al. (2013) Genome of *Phaeocystis globosa* virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proc Natl Acad Sci USA* 110(26):10800–10805.
9. Moniruzzaman M, et al. (2014) Genome of brown tide virus (AaV), the little giant of the *Megaviridae*, elucidates NCLDV genome expansion and host-virus coevolution. *Virology* 466-467:60–70.
10. Boyer M, et al. (2009) Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc Natl Acad Sci USA* 106(51): 21848–21853.
11. Thomas V, et al. (2011) Lausannevirus, a giant amoebal virus encoding histone doublets. *Environ Microbiol* 13(6):1454–1466.
12. Colson P, et al. (2013) "*Marseilleviridae*," a new family of giant viruses infecting amoebae. *Arch Virol* 158(4):915–920.
13. Doutre G, Philippe N, Abergel C, Claverie J-M (2014) Genome analysis of the first *Marseilleviridae* representative from Australia indicates that most of its genes contribute to virus fitness. *J Virol* 88(24):14340–14349.
14. Philippe N, et al. (2013) Pandoraviruses: Amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341(6143):281–286.
15. Antwerpen MH, et al. (2015) Whole-genome sequencing of a pandoravirus isolated from keratitis-inducing acanthamoeba. *Genome Announc* 3(2):e00136-15.
16. Legendre M, et al. (2014) Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci USA* 111(11): 4274–4279.
17. Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29(12):2607–2618.
18. Clarke M, et al. (2013) Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol* 14(2):R11.
19. NCBI Resource Coordinators (2015) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 43(Database issue):D6–D17.
20. Renesto P, et al. (2006) Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. *J Virol* 80(23):11678–11685.
21. Claverie JM, Abergel C, Ogata H (2009) Mimivirus. *Curr Top Microbiol Immunol* 328: 89–121.
22. Greber UF, Fassati A (2003) Nuclear import of viral DNA genomes. *Traffic* 4(3): 136–143.
23. Wu G, Nie L, Zhang W (2008) Integrative analyses of posttranscriptional regulation in the yeast *Saccharomyces cerevisiae* using transcriptomic and proteomic data. *Curr Microbiol* 57(1):18–22.
24. Michel R, Schmid EN, Hoffmann R, Müller KD (2003) Endoparasite KC5/2 encloses large areas of sol-like cytoplasm within *Acanthamoebae*. Normal behavior or aberration? *Parasitol Res* 91(4):265–266.
25. Maxwell KL, Frappier L (2007) Viral proteomics. *Microbiol Mol Biol Rev* 71(2):398–411.
26. Weekes MP, et al. (2014) Quantitative temporal viromics: An approach to investigate host-pathogen interaction. *Cell* 157(6):1460–1472.
27. Page FC (1988) *A New Key to Freshwater and Soil Gymnamoebae* (Freshwater Biological Association, Ambleside, UK).
28. Dupierris V, Masselon C, Court M, Kieffer-Jaquinod S, Bruley C (2009) A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa. *Bioinformatics* 25(15):1980–1981.
29. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367–1372.
30. Cox J, et al. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13(9):2513–2526.
31. Yang YH, et al. (2002) Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30(4):e15.
32. Shatilovich A, Shmakova L, Mylnikov A, Gilichinsky D (2009) Ancient protozoa isolated from permafrost. *Permafrost Soils*. Soil Biology, ed Margesin R (Springer, Berlin), pp 97–115.
33. Shatilovich AV, Shmakova LA, Gubin SV, Gudkov AV, Gilichinskiĭ DA (2005) Viable protozoa in late Pleistocene and Holocene permafrost sediments. *Dokl Biol Sci* 401: 136–138.
34. Simpson JT, Durbin R (2012) Efficient de novo assembly of large genomes using compressed data structures. *Genome Res* 22(3):549–556.
35. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
36. Luo R, et al. (2012) SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1(1):18.
37. Bonfield JK, Whitwham A (2010) Gap5—editing the billion fragment sequence assembly. *Bioinformatics* 26(14):1699–1703.
38. Marchler-Bauer A, Bryant SH (2004) CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res* 32(Web Server issue):W327–W331.
39. Shi J, Blundell TL, Mizuguchi K (2001) FUGUE: Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* 310(1):243–257.
40. Marchler-Bauer A, et al. (2002) CDD: A database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* 30(1):281–283.
41. Tatusov RL, et al. (2003) The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* 4:41.
42. Li L, Stoeckert CJ, Jr, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178–2189.
43. Snel B, Bork P, Huynen MA (1999) Genome phylogeny based on gene content. *Nat Genet* 21(1):108–110.
44. Cantarel BL, et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18(1):188–196.
45. Stanke M, Tzvetkova A, Morgenstern B (2006) AUGUSTUS at EGASP: Using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol* 7(Suppl 1):S11.1–8.
46. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res* 33(20):6494–6506.
47. Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* 5(1):59.
48. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18):3551–3567.
49. Conesa A, et al. (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674–3676.
50. Krumsiek J, Arnold R, Rattei T (2007) Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23(8):1026–1028.
51. Kim D, et al. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
52. Arike L, et al. (2012) Comparison and applications of label-free absolute proteome quantification methods on *Escherichia coli*. *J Proteomics* 75(17):5437–5448.