Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology

Matthieu Legendre1,2, Julia Bartoli1, Lyubov Shmakova3, Sandra Jeudy4, Karine Labadie4, Annie Adrait4, Magali Lescot2, Olivier Poirot5, Lionel Bertaux6, Christophe Bruley7, Yohann Coutdé7, Elizaveta Rivkina8, Chantal Abergel6,2, and Jean-Michel Claverie6,2

1Structural and Genomic Information Laboratory, Unité Mixte de Recherche 7256 (Institut de Microbiologie de la Méditerranée) Centre National de la Recherche Scientifique, Aix-Marseille Université, 13288 Marseille Cedex 9, France; 2Institute of Physicochemical and Biological Problems in Soil Science, Russian Academy of Sciences, Pushchino 142290, Russia; 3Commissariat à l’Energie Atomique, Institut de Génomique, Centre National de Séquençage, 91057 Evy Cedex, France; 4Commissariat à l’Energie Atomique, Institut de Recherches en Technologies et Sciences pour le Vivant, Biologie à Grande Échelle, Institut National de la Santé et de la Recherche Médicale, Unité 1038, Université Joseph Fourier Grenoble 1, 38054 Grenoble, France; 5Assistance Publique - Hôpitaux de Marseille, 13385 Marseille, France

Edited by James L. Van Etten, University of Nebraska–Lincoln, Lincoln, NE, and approved January 30, 2014 (received for review November 7, 2013)

The largest known DNA viruses infect Acanthamoeba and belong to two markedly different families. The Megaviridae exhibit pseudo-icosahedral virions up to 0.7 μm in diameter and adenine–thymine (AT)-rich genomes of up to 1.25 Mb encoding a thousand proteins. Like their Mimivirus prototype discovered 10 y ago, they entirely replicate within cytoplasmic virion factories. In contrast, the recently recovered Pandoraviruses exhibit larger amphora-shaped virions 1 μm in length and guanine–cytosine-rich genomes up to 2.8 Mb long encoding up to 2,500 proteins. Their replication involves the host nucleus. Whereas the Megaviridae share some general features with the previously described icosahedral large DNA viruses, the Pandoraviruses appear unrelated to them. Here we report the discovery of a third type of giant virus combining an even larger pandoravirus-like particle 1.5 μm in length with a surprisingly smaller 600 kb AT-rich genome, a gene content more similar to irdiviruses and Marsileivirus, and a fully cytoplasmic replication reminiscent of the Megaviridae. This suggests that pandoravirus-like particles may be associated with a variety of virus families more diverse than previously envisioned. This giant virus, named Pithovirus sibericum, was isolated from a >30,000-y-old radiocarbon-dated sample when we initiated a survey of the virome of Siberian permafrost. The revival of such an ancestral amoeba-infecting virus used as a safe indicator of the presence of unconventional DNA viruses, suggests that the thawing of permafrost either from global warming or industrial exploitation of circumpolar regions might not be exempt from future threats to human or animal health.

giant DNA virus | pale Pleistocene | icosahedral capsid

Ten years ago, the discovery of Acanthamoeba polyphage Mimivirus revealed the existence of giant DNA viruses with particles large enough to be visible under a light microscope (1, 2). Further sampling of various environments and geographical locations led to the isolation of Mimivirus variants (3, 4) and more distant relatives, two of which have been fully sequenced: Moumouivirus (5) and Megavirus chilensis (6). All of these viruses share the same distinctive particle structure: 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 nucleocapsid. They share an adenine–thymine (AT)-rich (>70%) linear DNA genome with sizes up to 1.25 Mb encoding up to 1,120 proteins (6). They all encode a full transcription apparatus allowing them to replicate in the host’s cytoplasm. These common features suggested that all giant viruses were to belong to a single family (Megaviridae) sharing the same particle morphology, genomic features, and replication strategy (7, 8).

This assumption was radically challenged by the discovery of Pandoraviruses (9), which show different and somewhat opposite characteristics. These Acanthamoeba-infecting viruses exhibit much larger amphora-shaped virions 1–1.2 μm in length. Their guanine–cytosine (GC)-rich (>61%) genomes are up to 2.8 Mb long and encode up to 2,500 proteins sharing no resemblance with those of Megaviridae (9). Finally, Pandoravirus particles do not incorporate the transcription machinery that would allow them to entirely replicate in the host’s cytoplasm. Known giant viruses infecting Acanthamoeba were thus thought to belong to two very dissimilar types in terms of particle structure, genome characteristics, and replication strategies. Here we describe a third type of giant virus named “Pithovirus” (from the Greek word pithos designating the kind of large amphora handed over by the gods to the legendary Pandora) propagating in an even larger pandoravirus-like particle, but exhibiting a replication cycle and genomic features reminiscent of those of large icosahedral nucleocyttoplasmic DNA viruses. Giant viruses are thus much more diverse than initially assumed.

Significance

Giant DNA viruses are visible under a light microscope and their genomes encode more proteins than some bacteria or intracellular parasitic eukaryotes. There are two very distinct types and infect unicellular protists such as Acanthamoeba. On one hand, Megaviridae possess large pseudoicosahedral capsids enclosing a megabase-sized adenine–thymine-rich genome, and on the other, the recently discovered Pandoraviruses exhibit micron-sized amphora-shaped particles and guanine–cytosine-rich genomes of up to 2.8 Mb. While initiating a survey of the Siberian permafrost, we isolated a third type of giant virus combining the Pandoravirus morphology with a gene content more similar to that of icosahedral DNA viruses. This suggests that pandoravirus-like particles may correspond to an unexplored diversity of unconventional DNA virus families.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The Pithovirus genome sequence reported in this paper has been deposited in GenBank database (accession no. KF746464). The MS proteomics data reported in this paper have been deposited in the ProteomeXchange Consortium database, http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000460, via the Proteomics Identifications database (www.ebi.ac.uk/pridearchive) partner repository [identifier no. PXD000460 (10.6019/PXD000460)].

M. Legendre and J.B. contributed equally to this work.

2To whom correspondence may be addressed. E-mail: claverie@igs.cnrs-mrs.fr or Chantal.Abergel@igs.cnrs-mrs.fr.

This article contains supporting information online at www.pnas.org/cgi/suppl/doi:10.1073/pnas.13206707111/DCSupplemental.
Results
Particle Morphology. If the notion that DNA may persist over geological timespans (>1 My) is gaining a progressive acceptance, the fact that cellular organisms might survive that long remains a contentious issue (10, 11). Because of its neutral pH and reducing and anaerobic properties, northeast Siberian permafrost is among the most suitable environments to look for long-term surviving microorganisms (12) or even plants (13). Except for few studies targeting the Influenza virus and the Variola virus over historical timespans (14, 15), the possibility that DNA viruses might remain infectious over a much longer time scale has not yet been investigated. We recently initiated such a survey using Acanthamoeba as bait and its giant DNA viruses as safe surrogates for pathogenic viruses. In this context, Pithovirus was initially spotted using light microscopy as ovoid particles (Fig. S1) multiplying in a culture of Acanthamoeba castellanii inoculated with a sample of Siberian permafrost from the Kolyma lowland region (SI Materials and Methods). This sample was aseptically collected from a permafrost layer corresponding to late Pleistocene sediments older than 30,000 y (13). Similar paleosols are known to contain cysts of Acanthamoeba cells (16). After amplification, the particles were analyzed by transmission electron microscopy confirming that Pithovirus shared the overall morphology of the Pandoraviruses (9) with slightly larger dimensions (~1.5 μm in length, 500 nm in diameter). The virions exhibit a 60 nm-thick structured envelope made of one layer of parallel stripes (Fig. 1). An internal membrane (Fig. 1A) encases a compartment without discernible substructures except for an electron-dense sphere (50 nm in diameter) seen episodically but in a reproducible fashion (Fig. S1) and a tubular structure parallel to the long axis of the particle (Fig. 1B). At variance with the Pandoraviruses, the apex aperture of the Pithovirus particle appears sealed by a protruding cork (80 nm thick and 160 nm wide) with a hexagonal grid structure (Fig. 1C) reminiscent of the organization of capsomers in icosahedral virions. A coil of rolled-up membrane seems to be con-
Pandoravirus, the envelope and the interior of the Pithovirus particles being assembled or “knitted” simultaneously (9) (Fig. 1H and Movie S1). First, rectangular-shaped closed particles with their characteristic cork appear at the periphery of the virophage factory (Fig. 1I). Later on, their outer wall thickens and the particles take their final ovoid shape, still lacking their thick striated tegument (Fig. 1J). This layer is built subsequently and piecewise, as evidenced in cross section images of immature particles (Fig. 1K) as well as by the fuzzy appearance of the tegument-like envelope in the latest stages of particle maturation (Fig. 1G and Fig. S1). After 6–8 h, particles at various stages of maturation may coexist in the same virion factory.

Besides dense vesicles of unknown composition accumulating inside the infected cells, shapeless blobs made of pieces of striated envelope, pieces of the corks, and diffuse material reminiscent of the mature particles interior (Fig. 1F and Fig. S1) are seen at the periphery of the virion factories. They might be reservoirs of partially organized virophage building blocks. Mature particles are found in equal amounts in the cytoplasm or in vacuoles suggesting that they could exit the cell by exocytosis (Movie S1). The replicative cycle ends with the cells releasing hundreds of particles upon lysis. Despite their bacterial-like dimensions no image of particles undergoing binary fission was encountered throughout our comprehensive electron microscopy study, hinting at the viral nature of Pithoviruses before the analysis of their gene content.

**Genome Sequencing.** DNA from purified particles was sequenced using the standard Illumina protocol (2× 100 nt paired end reads). The resulting 600 kb of unique sequence data readily allowed the identification of most genes, but could not be assembled in less than 80 nonoverlapping scaffolds. The finished genome was obtained by sequencing a Nextera mate pair library (5–8 kb inserts) using the Illumina MiSeq platform (1,066,320 reads, 2× 250 nt) combined with 77,241 PacBio RS long single-end reads. Unexpectedly, given the similar morphology of their particles, the Pithovirus genome was found to be completely different from the Pandoravirus’s in terms of size, nucleotide composition, topology, and gene content. The Pithovirus genome consists of an AT-rich (64%) dsDNA molecule of size, nucleotide composition, topology, and gene content. The Pithovirus genome appears globally more similar to Marseilleviridae (19 best hits), then Megaviridae (15 best hits), and then Iridoviridae (10 best hits), all of which are well-established families of icosahedral large DNA viruses. Remarkably, there were only five Pithovirus proteins with their closest homologs in Pandoraviruses. None of these proteins are clearly associated with a functional attribute, except for a remote phosphoglycerate mutase homolog (Pithovirus protein #15 (pv15), 27% identity).

An unusually large fraction of the Pithovirus genome (129 kb, 21.2%) corresponds to multiple regularly interspersed copies of a noncoding repeat, the intricate structure of which produces a unique fractal-like dot-plot pattern (Fig. 3A and B). These repeats exhibit GC content (23%) much lower than the coding regions (41%), similar to the one of intergenic regions (32.5%). For comparison, it was 60.6% for Mimivirus (2), 41% for Marseillevirus (18), and only 15.7% for Pandoravirus (9). The best matches are distributed almost equally among DNA viruses, bacteria, and eukaryotes, suggesting the absence of close relatives among previously sequenced organisms (Fig. 2A). A very similar distribution ($\chi^2 = 0.683$, $P > 0.95$) was computed for the 159 predicted protein validated by their detection in the particle proteome (Table S1), confirming that the small fraction of database matches was not due to bioinformatic overpredictions (Fig. 2C).

Although only one-third of the best database matches corresponded to viral proteins (representing 11% of the total gene content), the absence of genes coding for translation components, DNA replication enzymes, or related to cell-division confirmed the viral nature of Pithovirus (20). The low level of sequence similarity of these best matches (44% identical residues in average across the highest BLAST scoring segment pairs) (Tables S1 and S2) as well as their dispersion among different DNA virus families argues against Pithovirus being a member of any one of them. On the basis of this distribution (Fig. 2B), the Pithovirus appears globally more similar to Marseilleviridae (19 best hits), then Megaviridae (15 best hits), and then Iridoviridae (10 best hits), all of which are well-established families of icosahedral large DNA viruses. Remarkably, there were only five Pithovirus proteins with their closest homologs in Pandoraviruses. None of these proteins are clearly associated with a functional attribute, except for a remote phosphoglycerate mutase homolog (Pithovirus protein #15 (pv15), 27% identity).

**Fig. 2.** Distributions of the Pithovirus protein closest homologs. (A) All predicted protein sequences against the NCBI NR (non-redundant) database. (B) Distribution of the 51 best-matching viral proteins. (C) Subset of the 159 proteins detected in the particle proteome. (D) Distribution of the predicted protein functions.

Legendre et al.
Other categories include carbohydrate processing, RNA processing, and various hydrolases and oxidoreductases (Fig. 2D). At variance with other giant DNA viruses, no component of the protein translation machinery, including tRNA, is encoded in the Pithovirus genome. The presence of a complete virus-encoded transcription machinery (most of which found in the particle; Table S1) is consistent with the cytoplasmic location of the Pithovirus replication (Fig. 1). In contrast with the extremely high-repeat content of the genome, a single intervening sequence, a group I self-splicing intron, was detected and validated in the DNA-dependent RNA polymerase large subunit gene (RPB1, pv366–368). As borderline database matches were further scrutinized, we detected a remote similarity (E value = 0.002, 21% identical residues) between the pv460 gene product and the divergent major capsid protein (MCP) characteristic of Iridoviruses (e.g., in Megalocytivirus; GenBank accession no. AFE85881.1). The FUGUE server (25) predicted that the pv460 protein would adopt the structural “jelly-roll” fold common to large DNA virus MCPs. However, although the pv460 gene is transcribed, its product was not found in the mature virion. Like the Poxvirus D13 protein (26, 27), the pv460 protein might only play a transient role in the particle morphogenesis.

**Phylogenetic Analysis.** Despite its sizable Pandoravirus-like particle, Pithovirus exhibits a replication cycle and a gene content (Fig. 2 and Table S2) more similar to those of previously described large icoshedral eukaryotic DNA viruses, such as Marseilleviridae and Iridoviridae. This global picture is consistent with the neighboring clustering pattern of the Pithovirus DNA polymerase (Fig. 4). It is further confirmed by the phylogenetic positioning of three other viral core proteins shared with Pandoraviruses using maximum likelihood-based analyses and a cladogram based on the presence/absence of 205 conserved viral genes (Fig. S3). In all cases, Pithovirus was positioned within well-supported clades including Marseilleviruses, Iridoviruses, and both. The tree topologies confirmed the unexpected absence of a close evolutionary relationship between Pithovirus and the look-alike Pandoraviruses.

**Particle Proteome.** The proteome analysis of purified Pithovirus virions identified 159 different gene products, two-thirds of them corresponding to unknown functions (Fig. 2C and Table S1). The number of proteins making the Pithovirus and Pandoravirus particles (i.e., 210) is thus similar, despite their fivefold difference in gene number (9). This finding refutes the simple idea that large and complex particles should always correlates with large genomes, as observed until now (2, 3, 5, 6, 9). However, despite their overall similar morphology and complexity, the Pithovirus and Pandoravirus particles are made of entirely different sets of proteins, only sharing one pair of homologous proteins (pv384 and pv500) with highly discrepant abundances in their respective proteomes (ranked 8th and 162nd, respectively) (Table S1). In further contrast with the Pandoraviruses, the Pithovirus particles incorporate the complete transcriptional machinery encoded by its genome (Table S1). The availability of a preloaded functional transcription apparatus in the particle is the sine qua non for a fully cytoplasmic replication cycle also characterizing the Megaviridae (8) and the Poxviridae (26, 28). The presence of glycosylated proteins in the Pithovirus particles (also found in Marseillevirus (18) and Mimivirus (29)) is another noticeable difference from Pandoravirus, in which none were found (Fig. S4). In the absence of detectable MCP or core protein homologs, the major structural components of the Pithovirus particle remain to be identified. Potential candidates include the ankyrin domain-containing proteins and collagen triple helix repeat-containing proteins encoded in multiple copies in the genome. None of the ankyrin domain-containing proteins are detected in the proteome whereas 7 of the 12 collagen repeat-containing proteins were readily detected and may participate in the structural scaffold, corresponding to the striated tegument patterns (Fig. 1). Upon semiquantitative analysis (30) (Table S1), the four far most abundant proteins are encoded by genes pv449, pv461, pv93, and pv106. Unfortunately neither their amino acid sequences nor predicted fold provide clues about their functions or evolutionary origin.

Besides most of the enzymes related to nucleotide synthesis or nucleic acid processing, a few others are associated with the particle, such as an adenylsulfinate synthetase (pv90), an ADP ribosyl glycohydrolase (pv118), GTP-binding proteins (pv116, pv127, and pv137) and a GTPase (pv213), a protease (pv133), protein kinases (pv156, pv405, and pv438), hydrolases (pv212...
Our discovery of the giant Pandoraviruses, exhibiting astonishing gene contents packaged into uniquely amphora-like shaped particles, hinted that an entire world of viruses remained to be unraveled. The simultaneous isolation of two pandoravirus family members (P. salinus and Pandoravirus dulcis) from two remote locations suggested that they were not rare. The lack of gene content similarity between these Pandoraviruses and the newly isolated Pithovirus is now adding another dimension in the exploration of the world of giant viruses. We now have evidence that viruses from entirely different families (from a genomic point of view) may be associated with the same particle morphotype defined by an amphora-like shape, an apex aperture, and a thick electron-dense tegument lined by a lipid membrane enclosing an empty-looking internal compartment. At this point, one may wonder about the diversity of viruses using Pandora box-like particles that remains to be discovered. If this diversity is as large as that of viruses associated with iccosahedral capsids, Pandoravirus-like particles might be used to propagate genomes with vastly different sizes and/or molecular structures (e.g., dsDNA, ssDNA, or RNA). The hunt for these viruses is only beginning and speculations about their evolutionary origin should be postponed before many more are found. Unfortunately, there are only three representatives known today (Pithovirus sibiricum, P. salinus, and P. dulcis) and they do not share a distinctive set of genes from which to design molecular probes or mine metagenomic data. In the absence of such trademark genes, the identification of additional Pandora box-associated viruses will require a challenging prospective screening effort.

### Conclusion

Using Acanthamoeba as bait, we isolated a large DNA virus from a 30,000-y-old permafrost layer. To our knowledge this is the most ancient eukaryote-infecting DNA virus revived to date. Previous studies reported the detection of genomic signatures of Emiliania huxleyi virus in 7,000-y-old sediments underlying the Black Sea (33) and of tomato mosaic tobamovirus in 140,000-y-old glacial ice in Greenland (34), but did not address their viability. Although Pithovirus is the prototype of a new family, its genome structure and replication cycle are similar to that of other large eukaryotic DNA viruses, many of which are human or animal pathogens. Our results thus further substantiate the possibility that infectious viral pathogens might be released from ancient permafrost layers exposed by thawing, mining, or drilling. Climate change in the Russian Arctic is more evident than in many other regions of the world. Whereas the average global temperature has increased by 0.7 °C during the last 100 y, the average temperatures of the surface layer of Arctic permafrost have increased by 3 °C during the same period. In the 20th century, the permafrost in the Northern Hemisphere has diminished by 7% (35). This no doubt corresponded to a large release of microorganisms from previously frozen soils, an unknown fraction of which was revived upon thawing. Indeed, pathogenic bacteria can survive under low temperatures recurrently causing diseases in circumpolar regions (35). The rich mineral resources and oil reserves of the arctic regions are under increasing pressure for their industrial exploitation (involving mining and drilling). It thus becomes urgent to examine which viruses are expected to
be encountered not only near the surface, but in increasingly deeper and more ancient permafrost layers (down to 3 My) (36, 37). Attempting to isolate and revive increasingly ancient amoeba-infecting viruses from permafrost is an inexpensive and safe way to realistically assess the threat posed by pathogenic animal DNA viruses eventually released from soils and sediments frozen in a distant past. Metagenomic analyses will also be used to safely document the eventual presence of known viral pathogens in permafrost samples containing viable amoeba-infecting viruses.

**Materials and Methods**

**Sample Recovery and Radiocarbon Dating.** The P-1084 T permafrost sample was collected in 2000 from the frozen outcrop walls in Chukotka, on the Stanichkivsky Yar, a steep bank 23 m above the level of the Anui River. It had never thawed. See SI Materials and Methods for details.

**P. sibiricum Isolation and Production.** The buried soil sample was resuspended in Prescott and James medium (38) and used to infect A. castellanii (Douglas Neff, American Type Culture Collection 30010TM) cells. The wells presenting amoeba-infecting viruses from permafrost did not produce the dscDNA. Illumina 250-bp paired-end libraries were prepared for sequencing. SI Materials and Methods contains experimental details and the bioinformatic procedures used for data analyses.

**Proteome Analysis.** Purified particles (10^7) were resuspended in 100 μL lysis buffer Tris20 mM CaCl2/1 mM pH 7.4 and incubated 5 min at 95 °C. Proteins were stained in the top of a 4–12% (wt/vol) NuPAGE gel (Invitrogen) for 19 h pi. RNA was extracted and quantified and full-length cDNA synthesis was performed using the switching mechanism at 1′ end of RNA transcript cDNA library construction kit. The Advantage 2 PCR Kit (Clontech) was used to produce the dscDNA. Illumina 250-bp paired-end libraries were prepared for sequencing. SI Materials and Methods contains experimental details and the bioinformatic procedures used for data analyses. The buried soil sample was resuspended in Prescott and James medium (38) and used to infect A. castellanii (Douglas Neff, American Type Culture Collection 30010TM) cells. The wells presenting amoeba-infecting viruses from permafrost did not produce the dscDNA. Illumina 250-bp paired-end libraries were prepared for sequencing. SI Materials and Methods contains experimental details and the bioinformatic procedures used for data analyses.

**Acknowledgments.** We thank Dr. J-P. Chauvin, F. Richard, and A. Aouane for their expert assistance on the electronic microscopy platform (Institut de Biologie du Développement de Marseille). Technical assistance was provided by Dr. N. Philippe, V. Schmidt, and M.-A. Galeron. We also thank F. Lopez (Jeol), D. Bertrand, and S. Searle for their expert assistance on the 3D reconstruction shown in Movie 5. This work was partially supported by France Génomique Grant ANR-10-INSB1-01-01, by the Provence-Alpes-Côte-d’Azur région (2010 12125), and by the Program of the Presidium of the Russian Academy of Sciences 28 and the Russian Foundation of Basic Research (Projects 12-04-31827 and 12-05-01085).