Supporting Information

Legendre et al. 10.1073/pnas.1320670111

SI Materials and Methods

Sample Recovery and Radiocarbon Dating. The sampling site is located in a tundra ecosystem, adjacent to the East Siberian Sea. This area is characterized by an arctic climate with a mean annual air temperature of −13.4 °C. The seasonal thawing of the soil surface starts in June and reaches a maximum of 0.5 m in mid-September. Sample P-1084 T was collected in 2000 and never thawed after collection. In the field, the sample was kept in a cave dug into the permafrost under natural temperatures (−10 °C) before transport to the Pushchino storage facility in a freezer (−20 °C). The late Pleistocene ice complex found throughout the coastal lowlands of northeastern Siberia is of a syncriogenic origin, i.e., sediment accumulation and freezing occurring simultaneously (1). Samples of buried soils were taken from the frozen outcrop walls in Chukotka, on the Stanchikovsky Yar exposure located on the North right-hand side bank of the Anuy River [68.370155 (68°22′13″ N), 161.415553 (161°24′56″ E)]. The height of the bluff is about 50 m above the river level and represents the ice complex thickness. The main unit consists of syngenetic bluish-gray or gray-brown silt with large ice wedges up to 6–8 m wide and up to 30–40 m high. These ice wedges divide the outcrop into separate mineral blocks. Sample P-1084 T was taken from a profile of buried soil 23 m above the river level (thus about 30 m below the surface). After carefully trimming the outcrops of unfrozen sediment, a horizontal 40-cm long borehole was made by slow rotary drilling without the use of drilling mud to minimize possible contamination. After treating the core surface with 95% ethanol, a central sample was taken and placed in a sterile plastic bag and stored frozen. The procedure developed for sterile sampling and contamination testing was described previously (2–4). Paleosoil age was determined by radiocarbon method and is 34,000–37,000 y old (5).

Virus Isolation and Production. Pithovirus was isolated from a piece of the buried soil sample P1084-T. Briefly, 400 mg of P1084-T were resuspended in 6 mL Prescott and James medium (6). The infection trials were performed twice and produced identical results. Each 3 mL were supplemented with 300 μL of Amphotericin B (Fungizone) at 250 μg/mL. Of this 10% fungizone solution, 1.65 mL was left overnight at 4 °C under stirring. After decantation, the supernatant was recovered and centrifuged at 800 × *g* for 5 min. *Acanthamoeba castellanii* (Douglas) Neff (American Type Culture Collection 30010TM) cells adapted to resist Fungizone 2.5 μg/mL were inoculated with 100 μL supernatant and with the pellet resuspended in 50 μL Tris 20 mM/$CaCl_2$ 1 mM (pH 7.4). The cells were cultured at 32 °C in microplates with 1 mL 2% (wt/vol) proteose peptone, 0.1% yeast extract, 100μM glucose, 4mM $MgSO_4$, 0.4mM $CaCl_2$, 50 μM $Fe(NH_4)_2(SO_4)_2$, 2.5 mM Na_2HPO_4, 2.5 mM KH_2PO_4, pH 6.5 (home made PPYG) medium supplemented with antibiotics [ampicilline 100 μg/mL and penicillin–streptomycin (Gibco) 100 μg/mL, Fungizone 2.5 μg/mL] and monitored for cell lysis.

To verify that the permafrost sample was not cross-contaminated with another environmental sample handled in the laboratory, we reproduced the initial coculture experiment from another piece of the P1084-T sample in a microbiological hood. We then repeated the infection protocol. A lytic infection of Acanthamoeba was again observed. Viruses were amplified, purified, and the genomic DNA was extracted as before. PCR amplifications were performed on the purified DNA using four sets of primers corresponding to four Pithovirus genes. All sequenced PCR products confirmed that we had again isolated the same virus. This result was confirmed by observation of inclusion of the purified particles by electron microscopy.

Virus Purification. The wells presenting an infected phenotype were recovered, centrifuged 5 min at 500 × *g* to remove the cellular debris, and used to infect four T-75 tissue-culture flasks plated with fresh Acanthamoeba cells. After lysis completion, the cultures were recovered, centrifuged 5 min at 500 × *g* to remove the cellular debris, and the virus was pelleted by a 30 min centrifugation at 3,000 × *g* prior purification. The viral pellet was then resuspended and washed twice in PBS and layered on a discontinuous sucrose gradient (30/40/50/60% wt/vol), and centrifuged at 5,000 × *g* for 15 min. The virus produced a white disk which was recovered and washed twice in PBS and stored at 4 °C or −80 °C with 7.5% DMSO.

Electron Microscopy. *A. castellanii*-infected cell cultures were fixed by adding an equal volume of PBS with 2% glutaraldehyde and 20 min incubation at room temperature. Cells were recovered and pelleted 20 min at 5,000 × *g*. The pellet was resuspended in 1 mL PBS with 1% glutaraldehyde, incubated at least 1 h at 4 °C, and washed twice in PBS prior coating in agarose and embedding in epon resin. Each pellet was mixed with 2% low melting agarose and centrifuged to obtain small flanges of approximately 1 mm^3 containing the sample coated with agarose. These samples were then embedded in epon resin using a standard method: 1 h fixation in 1% osmium tetroxyde, dehydration in increasing ethanol concentrations (50%; 70% including 2% uranyl acetate; 90%; and 100% ethanol), and then embedded in Epon-812. Ultrathin sections of 70 nm were poststained with 4% uranyl acetate and lead citrate and observed using a Zeiss EM 912 operating at 100 kV.

DNA Extraction. The genomic DNA was recovered from 48 × 10^9 purified particles by combining a phenol/chloroform extraction and the Purelink genomic DNA extraction mini kit (Life Technologies) protocols. First the viruses were lysed using the lysis buffer from the kit supplemented with 2 mg/mL proteinase K and 6 mM DTT. After a standard phenol/chloroform and ethanol precipitation protocol, an extra step of purification was performed using silica columns from the Purelink kit to improve the quality of the genomic DNA.

Genome Sequencing and Assembly. ***Nextera Mate Pair library preparation and sequencing.*** The mate pair library was prepared following Nextera protocol (Nextera Mate Pair sample preparation kit; Illumina). Briefly, genomic DNA was simultaneously enzymatically fragmented and tagged with a biotinylated adaptor. Fragments were size-selected (5–8 kb) through regular gel electrophoresis, and circularized overnight with a ligase. Linear, noncircularized fragments were digested and circularized DNA was fragmented to 300- to 1,000-bp size range using Covaris E210. Biotinylated DNA was immobilized on streptavidin beads, end-repaired, 3′-adenylated, and Illumina adapters were added. DNA fragments were PCR-amplified using Illumina adapter-specific primers and then purified. Finally, the mate pair library was quantified by quantitative PCR (qPCR) and library profile evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies). Each library was sequenced using 250 bp-length read chemistry on a paired-end flow cell on the Illumina MiSeq (Illumina).

Genome assembly. An initial assembly of the Pithovirus genome was performed using Illumina 100-bp paired-end reads (24,884,112 sequences) with the Velvet assembler (7) and a k parameter of 97.

This resulted in an assembly of 602,026 bp composed of 80 nonoverlapping scaffolds. We next used a combination of sequencing technologies: the large insert (5–8 kb) Illumina MiSeq mate pair reads (1,066,320 sequences), and long single-end reads sequenced by the PacBio RS technology (77,241 sequences) to perform a second independent assembly. The Illumina MiSeq dataset was first assembled using Velvet with an optimal k parameter (k = 199) estimated using VelvetOptimizer script. This resulted in the assembly of a single scaffold of 614,434 bp composed of 30 contigs. The relative orientation of the contigs was manually checked using the mate-pair orientation and insert size information. We next closed the sequence gaps between contigs using the Illumina MiSeq dataset with GapCloser software (8). In addition we used the PacBio long reads with the PBjelly tool (9) to fill the remaining gaps. All of the sequence gaps were closed, resulting in a single contig assembly. Finally sequencing errors were corrected by mapping the Illumina MiSeq reads on the reference assembly using Bowtie2 (10), and consensus calling was performed using Gap5 (11). The same procedure was recursively done seven times, resulting in a high-quality genomic sequence of 610,033 nt. The Pithovirus protein coding regions were identified using the GeneMarkS algorithm (12) with the following parameters: combine, gcode = 1 and motif = 0. No tRNA was identified using the tRNAscan-SE (13) algorithm.

Gene Annotation. The homology search was performed using blastp against the nonredundant GenBank database (14) (Version April 4, 2014) with an E-value threshold $<10^{-5}$. The functional annotation of Pithovirus-predicted proteins was complemented by Pfam motif searches (15), the FUGUE program (16) and Reverse Position Specific-BLAST (17) against Clusters of Orthologous Groups (18) (E-value threshold $<10^{-3}$). The 3D-fold recognition server returned the pv460 encoded protein as a probable major capsid protein with a typical jelly-roll fold.

Identification of the Transcription Termination/Polyadenylation Palindromic Signals. The RNAmotif (19) tool was used to extract hairpins composed of a stem of at least 10 Watson–Crick bp, allowing for two mispairs, and a loop of 0–10 nt. Hairpins exhibiting mispairs at two consecutive positions were excluded. For each pattern we also computed a score corresponding to the length of the stem minus the number of mismatches. All patterns with a score below 11 were also excluded (20).

Transcriptome Library Preparation. ***A. castellanii infection by Pithovirus.*** A total of 1.75×10^8 adherent cells in 250 mL culture medium were infected by Pithovirus with a multiplicity of infection of 30 and distributed in 12 flasks (1.4×10^7 cells per 175 cm^2-flask) containing 20 mL PPYG. Cells were harvested by centrifugation at $1{,}000 \times g$ for 5 min and for each infection time (4, 11, and 19 h postinfection), the cells from three of the four flasks were kept aside for RNA extraction, and the remaining cells were pooled to make a fourth point corresponding to the mix of T4-11-19 h pi. An extra 5.6×10^7 of uninfected cells were kept as a control, with 10% of each cell culture set aside for electron microscopy observation.

RNA extraction. RNA was extracted using the RNeasy Midi kit (catalog no. 75144; QIAGEN) using the manufacturer's protocol. Briefly, the cells were resuspended in 4 mL RLT buffer supplemented with 0.1% β-mercaptoethanol and disrupted by subsequent −80 °C freezing and thawing at 37 °C for 10 min. Total RNA was eluted with two successive additions of ∼200 μL RNase free water.

RNA quantification and quality control. RNA was quantified by measuring the absorbance at 260 nm using a NanoDrop Spectrophotometer. The integrity of the RNA sample was assessed using the Experion Automated Electrophoresis System with RNA StdSens chips and reagents (Bio-Rad).

cDNA production. Full-length cDNA synthesis. Full-length cDNA synthesis was performed using switching mechanism at 5′ end of RNA transcript (SMART) cDNA library construction kit (catalog no. 634901; Clontech). First-strand cDNA synthesis was performed with the Moloney murine leukemia virus Reverse Transcriptase using SMART PCR technology (Clontech Laboratories) but following a modified protocol as described previously (18).

Full-length long-distance PCR. For the long-distance PCR (LD PCR), we used the Advantage 2 PCR Kit (Clontech Laboratories). Only sscDNAs that have the 5′-SMART anchor can be used as a template for LD PCR, thus ruling out eventual genomic DNA contamination. For each time course (uninfected; 4, 11, and 19 h pi; and the T4-11-19 h pi mix), we performed 2× 100 μL PCR reactions as previously described (18). The optimal number of cycles was determined as 26 by comparing results obtained with 22, 24, 26, and 28 cycles. To assess the quality of the dscDNA sample, we loaded 5 μL of each sample onto a 1.2% agarose gel, resulting in a smear from 100 bp to 3.5 kb. The dscDNA was cleaned and concentrated using the PureLink PCR Purification Kit (Invitrogen Life Technologies). Samples were eluted with 50 μL H_2O. cDNA purity was measured using the 260:280 nM and 260:230 nm ratio against H_2O.

Transcriptomic Illumina Library Preparation and Sequencing. The cDNAs were sonicated to a 150- to 1,000-bp size range using the E210 Covaris instrument (Covaris, Inc.). Fragments were end-repaired, then 3′-adenylated, and Illumina adapters were added by using NEBNext Sample Reagent Set (New England BioLabs). Ligation products were purified and DNA fragments (>200 pb) were PCR-amplified using Illumina adapter-specific primers. After library profile analysis by Agilent 2100 Bioanalyzer (Agilent Technologies) and qPCR quantification (MxPro; Agilent Technologies), each library was sequenced using 250 bp-length read chemistry in a paired-end flow cell on the Illumina MiSeq (Illumina).

Transcriptomic Data Analysis. The 5,909,660 paired-end reads sequenced from the five samples (uninfected; 4, 11, 19 h pi; and the T4-11-19 h pi mix) were mapped to a reference database containing the genome sequence of *A. castellanii* (GenBank assembly accession no. GCA_000193105.1), *A. castellanii* mitochondrion (Refseq accession no. NC_001637.1) and Pithovirus (GenBank accession no. KF740664), using GSNAP software (21) (Version September 11, 2013) with parameter w = 3,000. Overall 71% of the reads were successfully mapped to the reference, among which 142,724 corresponded to Pithovirus transcripts. We then extracted 23,224 reads that contained a SMART primer at the 5′ extremity and reciprocally 31,234 reads with a 3′-end primer, to identify Pithovirus transcripts boundaries. The stacking of at least three reads at an identical genomic position was used to delineate transcriptional start site (TSS) and transcriptional end site (TES). We were finally able to assign the TSS of 216 Pithovirus genes and the TES of 130 genes. The whole transcriptome shotgun sequencing experiment was primarily designed to validate Pithovirus gene predictions and cannot be exploited to analysis in depth the kinetic of gene expression during the replicative cycle.

Protein Extraction. Purified particles (10^8) were resuspended in 100 μL lysis buffer/Tris 20 mM/$CaCl_2$ 1 mM (pH 7.4) and incubated 5 min at 95 °C.

Protein Electrophoresis. Extracted proteins (10 μg and 15 μg) from Pithovirus solubilized in Laemmli buffer were separated on a 4–12% gradient polyacrylamide gel (NuPAGE; Invitrogen) before staining using respectively colloidal Coomassie blue (GelCode Blue Stain Reagent; Pierce, Thermo Scientific) and periodic acid-Schiff method (Glycoprotein detection kit; Sigma-Aldrich). Horseradish peroxidase (5 μg) was used as a positive control for glycoprotein detection.

Protein Digestion. Proteins were stacked in the top of a 4–12% NuPAGE gel (Invitrogen) before R-250 Coomassie blue staining. The gel band was manually excised and cut in pieces before being washed by six successive incubations of 15 min in 25 mM NH_4HCO_3 and in 25 mM NH_4HCO_3 containing 50% (vol/vol) acetonitrile. Gel pieces were then dehydrated with 100% acetonitrile and incubated for 45 min at 53 °C with 10 mM DTT in 25 mM NH_4HCO_3 and for 35 min in the dark with 55 mM iodoacetamide in 25 mM NH_4HCO_3. Alkylation was stopped by adding 10 mM DTT in 25 mM NH_4HCO_3 and mixing for 10 min. Gel pieces were then washed again by incubation in 25 mM NH_4HCO_3 before dehydration with 100% acetonitrile. Modified trypsin (sequencing grade; Promega) in 25 mM NH_4HCO_3 was added to the dehydrated gel pieces for an overnight incubation at 37 °C. Peptides were then extracted from gel pieces in three 15-min sequential extraction steps in 30 μL 50% acetonitrile, 30 μL 5% formic acid, and finally 30 μL 100% acetonitrile. The pooled supernatants were then dried under vacuum.

Nano–Liquid Chromatography-Tandem MS Analyses. The dried extracted peptides were resuspended in 5% acetonitrile and 0.1% trifluoroacetic acid and analyzed by online nano–liquid chromatography-tandem MS (LC-MS/MS) (Ultimate 3000 from Dionex and LTQ-Orbitrap Velos pro from Thermo Fischer Scientific). Peptides were sampled on a 300 μm × 5 mm PepMap C18 precolumn and separated on a 75 μm × 250 mm C18 column (PepMap, Dionex). The nano-LC method consisted of a 120-min gradient at a flow rate of 300 nL/min, ranging from 5% to 37% acetronitrile in 0.1% formic acid for 114 min before reaching 72% acetronitrile in 0.1% formic acid for the last 6 min. MS and MS/MS data were acquired using Xcalibur (Thermo Fischer Scientific). Spray voltage and the heated capillary were respectively set at 1.4 kV and 200 °C. Survey full-scan MS spectra (m/z = 400–1,600) were acquired in the Orbitrap with a resolution of 60,000 full width at halftime, at 400 m/z after accumulation of 10^6 ions (maximum filling time: 500 ms). The 20 most intense ions from the preview survey scan delivered by the Orbitrap were fragmented by collision-induced dissociation (collision energy 35%) in the LTQ after accumulation of 10^4 ions (maximum filling time: 100 ms).

MS Bioinformatics Analyses. Data were processed automatically using Mascot Daemon software (Version 2.3.2; Matrix Science). Concomitant searches against Pithovirus protein sequence databank (471 entries), *A. castellanii* protein sequence databank (14,974 entries) classical contaminants database (67 sequences, homemade), and the corresponding reversed databases were performed using Mascot (Version 2.4). ESI-TRAP was chosen as the instrument, trypsin/P as the enzyme, and two missed cleavage allowed. Precursor and fragment mass error tolerances were set respectively at 10 ppm and 0.6 Da. Peptide modifications allowed during the search were as follows: carbamidomethyl (C, fixes) acetyl (N-ter, variable), oxidation (M, variable), and deamidation (NQ, variable). The IRMa software (22) (Version 1.31.1) was used to filter the results: conservation of rank 1 peptides, peptide identification false discovery rate < 1% (as calculated on peptide scores by using the reverse database strategy), and minimum of one specific peptide per identified protein group.

Phylogenetic Analyses. The multiple sequence alignments were performed using the default options of the MAFFT server (23). For Fig. 4, the neighbor-joining tree was computed on the MAFFT server, using the JTT substitution model. The parameter of heterogeneity among sites was estimated at $\alpha = 1.05$. One hundred bootstrap resamplings were performed. The tree was rooted at the basis of the eukaryotes and collapsed for bootstrap values <50 before drawing using MEGA5 (24). For Fig. S3, the trees were computed using PhyML (25), according to the optimal protein evolution model (LG + G + I + F) suggested by ProtTest (26). Branch support is estimated using the aLRT parameter (27).

1. Schirrmeister L, et al. (2011) Sedimentary characteristics and origin of the Late Pleistocene Ice Complex on north-east Siberian Arctic coastal lowlands and islands—A review. *Quat Int* 241(1-2):3–25.
2. Shi T, Reeves RH, Gilichinsky DA, Friedmann EI (1997) Characterization of viable bacteria from Siberian permafrost by 16S rDNA sequencing. *Microb Ecol* 33(3): 169–179.
3. Juck DF, et al. (2005) Utilization of fluorescent microspheres and a green fluorescent protein-marked strain for assessment of microbiological contamination of permafrost and ground ice core samples from the Canadian High Arctic. *Appl Environ Microbiol* 71(2):1035–1041.
4. Gilichinsky DA, et al. (2007) Microbial populations in Antarctic permafrost: Biodiversity, state, age, and implication for astrobiology. *Astrobiology* 7(2):275–311.
5. Gubin SV, Zanina OG (2013) Variation of soil cover during the ice complex deposit formation, Kolyma lowland (Part 1). *Earth Cryosphere* XVII(1):48–56.
6. Page FC (1988) *A New Key to Freshwater and Soil Gymnamoebae* (Freshwater Biological Association, Ambleside, UK).
7. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829.
8. Luo R, et al. (2012) SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1(1):18.
9. English AC, et al. (2012) Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE* 7(11):e47768.
10. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
11. Bonfield JK, Whitwham A (2010) Gap5—editing the billion fragment sequence assembly. *Bioinformatics* 26(14):1699–1703.
12. 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.
13. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33(Web Server issue):W686-9.
14. Sayers EW, et al. (2012) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 40(Database issue):D13–D25.
15. Punta M, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40(Database issue):D290–D301.
16. 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.
17. 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.
18. Tatusov RL, et al. (2003) The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* 4:41.
19. Macke TJ, et al. (2001) RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res* 29(22):4724–4735.
20. Byrne D, et al. (2009) The polyadenylation site of Mimivirus transcripts obeys a stringent 'hairpin rule'. *Genome Res* 19(7):1233–1242.
21. Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26(7):873–881.
22. 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.
23. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9(4):286–298.
24. Tamura K, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739.
25. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3): 307–321.
26. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104–2105.
27. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55(4):539–552.

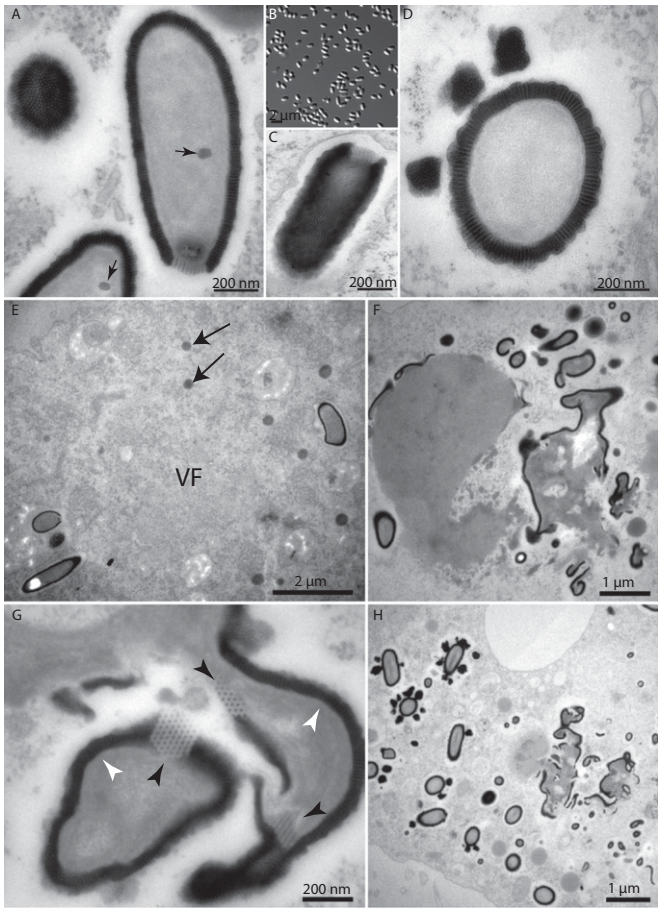


Fig. S1. Electron microscope imaging of Pithovirus particles and infected Acanthamoeba cells. (*A*) Ultrathin sections of Pithovitus particles showing a similar electron-dense sphere 50 nm in diameter (black arrows). (*B*) Light microscopy image of purified Pithovirus particles. (*C*) Section through the Pithovirus envelope showing its striated organization. The apex cork is also visible. (*D*) Globular striated structures participating to the envelope maturation at the periphery of a virion. (*E*) The virion factory (VF) appears as a clear area depleted of cytoplasmic subcellular structures. Dense vesicles accumulate in the periphery of the VF (black arrows). (*F*) Electron-dense unstructured intracytoplasmic areas. (*G*) Enlarged view showing the complex nature of these patches composed of an assortment of striated envelopes (white arrowheads), pieces of corks (black arrowheads), and diffuse material reminiscent of the mature particles' interior. (*H*) Later stage of the virion maturation inside a VF. Mature particles coexist with maturing virions surrounded by globular striped structures (enlarged view in Fig. 1*F*).

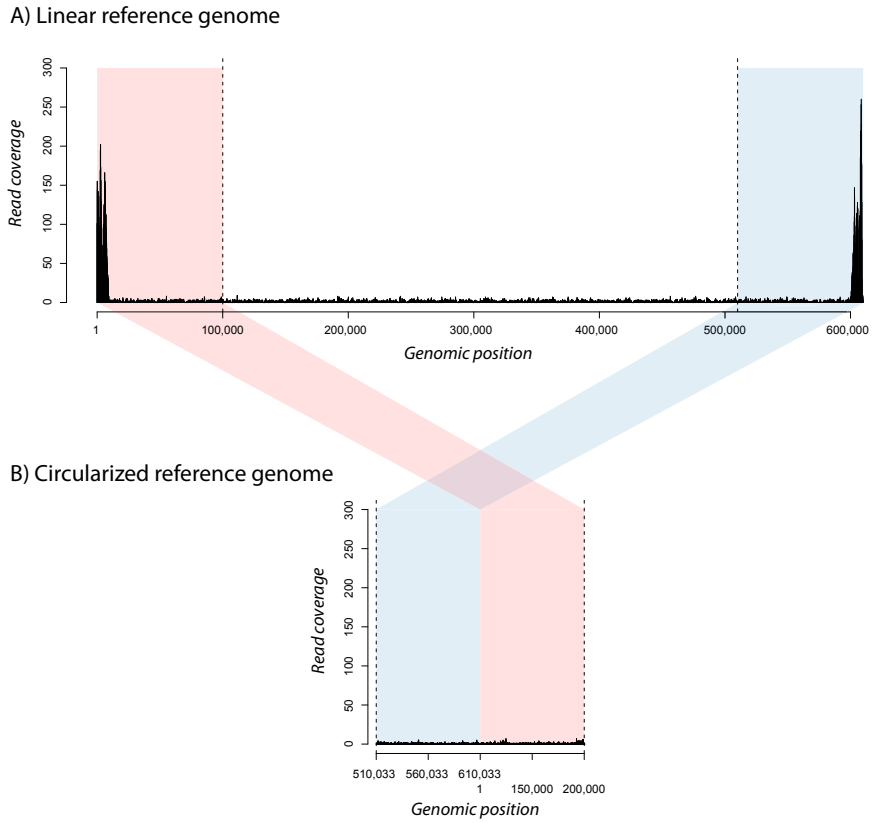


Fig. S2. Linear/circular Pithovirus genome. (*A*) The distribution of mate-paired reads exhibiting a reverse orientation (i.e., inward facing) and an abnormally large insert length (median +3σ), is biased toward the extremities of the linearized Pithovirus genome. (*B*) Conversely, this excess vanishes when the reads are mapped to a circularized version of the Pithovirus genome, made of the concatenation of the 3′ (blue region) and 5′ (red region) genomic extremities.

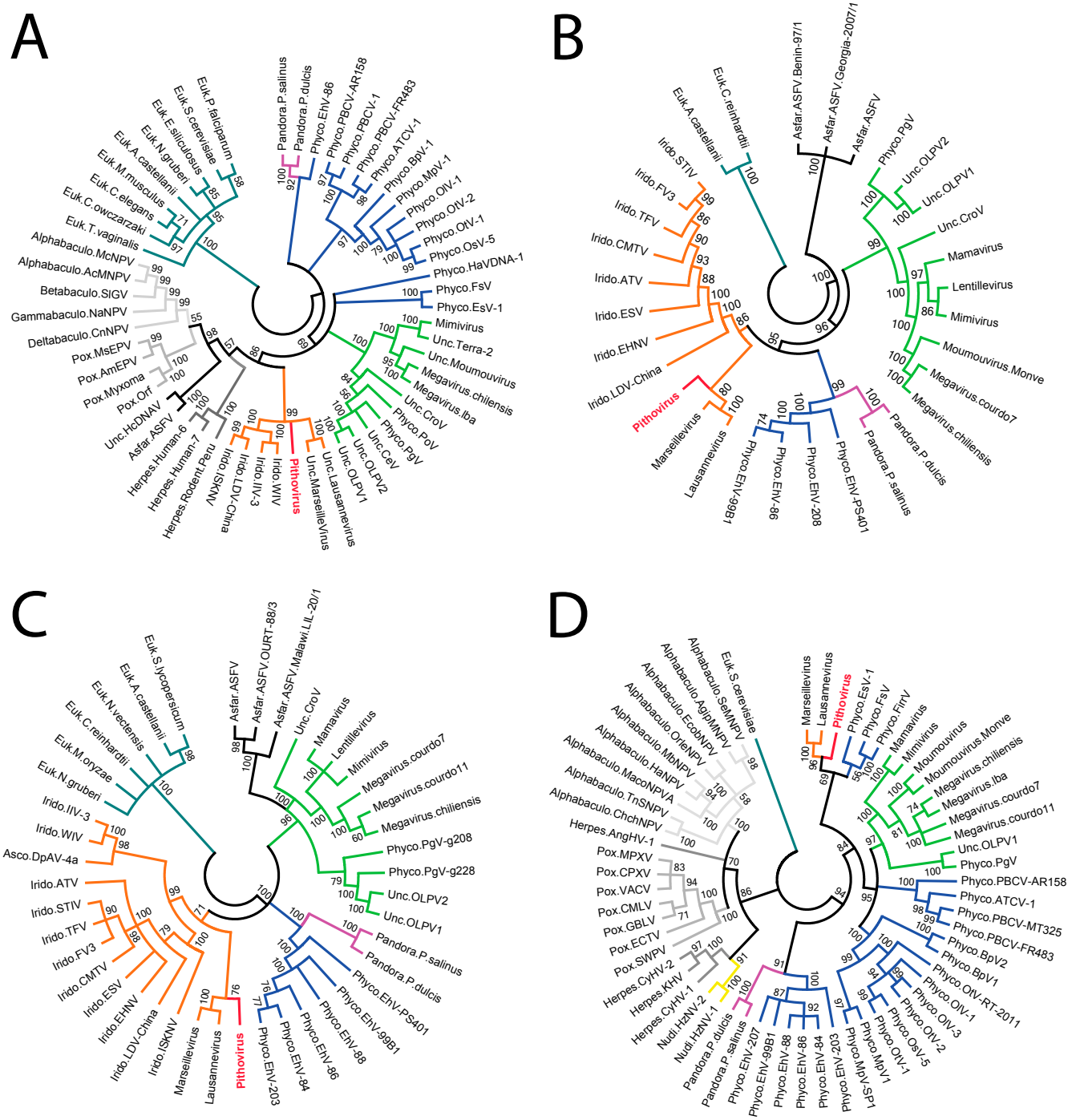


Fig. S3. (Continued)

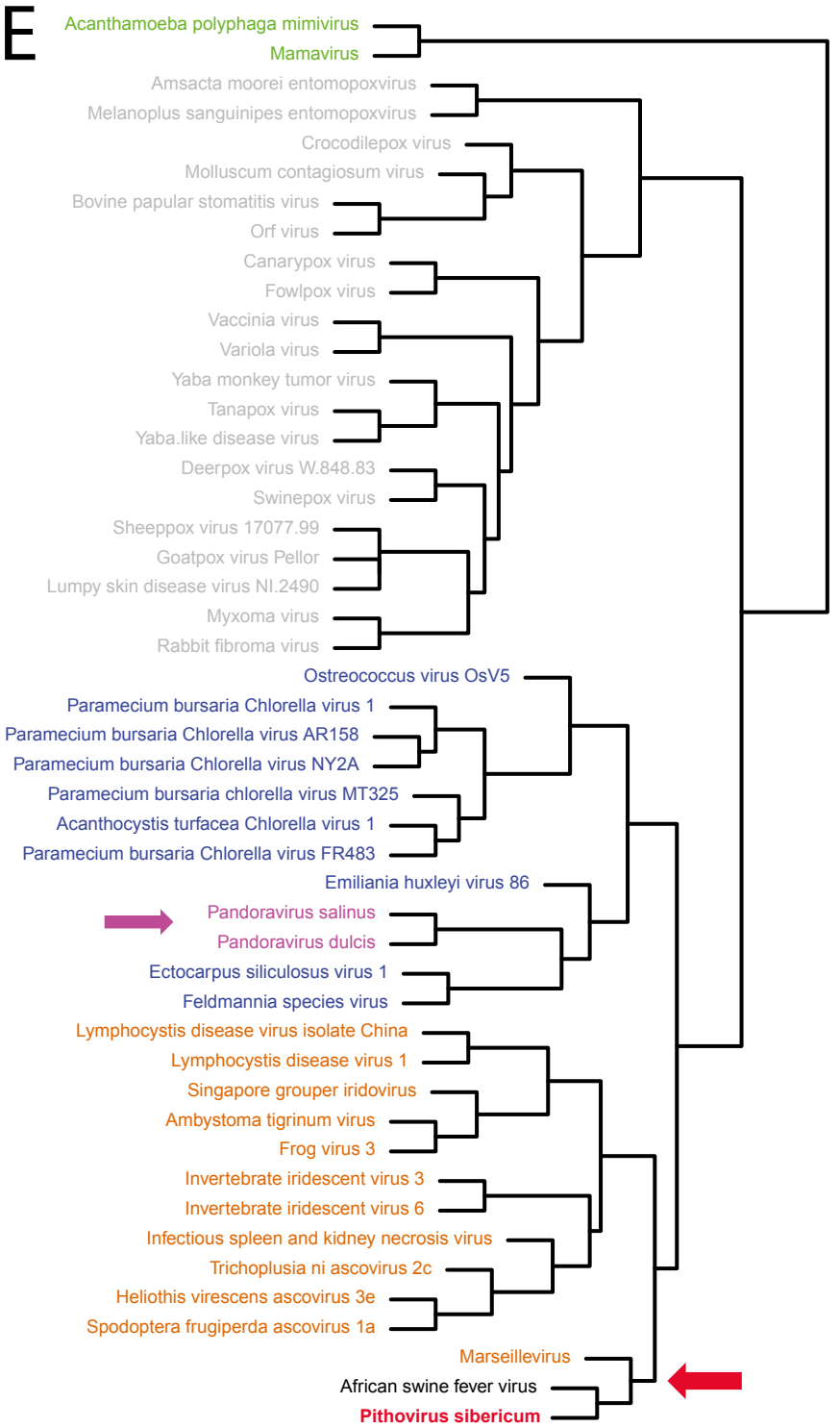


Fig. S3. Maximum likelihood and cladistics phylogenetic analyses. Pithovirus sequences are shown in red, Pandoraviruses in pink, Iridoviruses and Marseilleviridae in orange, Megaviridae in green, Baculoviridae in light gray, Poxviruses in gray, Herpesviridae in dark gray, Asfarviridae in black, Phycodnaviruses in blue, Nudiviruses in yellow, and eukaryotes in turquoise. Trees are rooted on the node leading to the eukaryotes when applicable. (*A*) B-family DNA polymerases. (*B*) DNA-directed RNA polymerase RPB1 subunits. Type I introns detected in Marseillevirus and Lausannevirus genes were manually removed. (*C*) DNA-directed RNA polymerase RPB2 subunits. (*D*) Ribonucleoside-diphosphate reductase large subunit. (*E*) Cladistic tree computed from the presence/absence of 205 viral genes from the Nucleo-Cytoplasmic Virus Orthologous Groups (NCVOGs) database (ftp.ncbi.nih.gov/pub/wolf/COGs/NCVOG). The analysis was restricted to NCVOGs belonging to well-defined functional categories (i.e., DNA replication, recombination and repair, transcription and RNA processing, translation, nucleotide metabolism, virion structure and morphogenesis, host–virus interactions, and signal transduction regulation). In all cases Pithovirus is positioned within a clade including the Marseilleviruses, the Iridoviruses, or both.

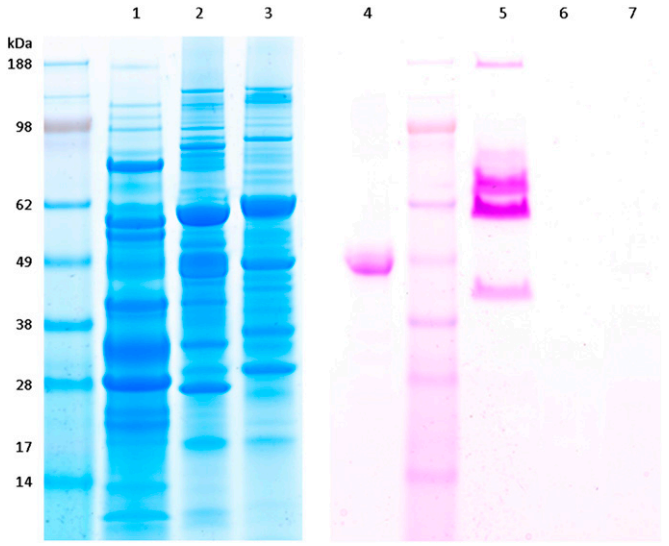


Fig. S4. Proteomic profiles of (1) *Pithovirus sibericum*-, (2) *Pandoravirus dulcis*-, and (3) *Pandoravirus salinus*-lysed particles (*Left*). Horseradish peroxidase positive control (4). Glycoproteins detected in lysed particles of (5) *P. sibericum*, (6) *P. dulcis*, and (7) *P. salinus* (*Right*).

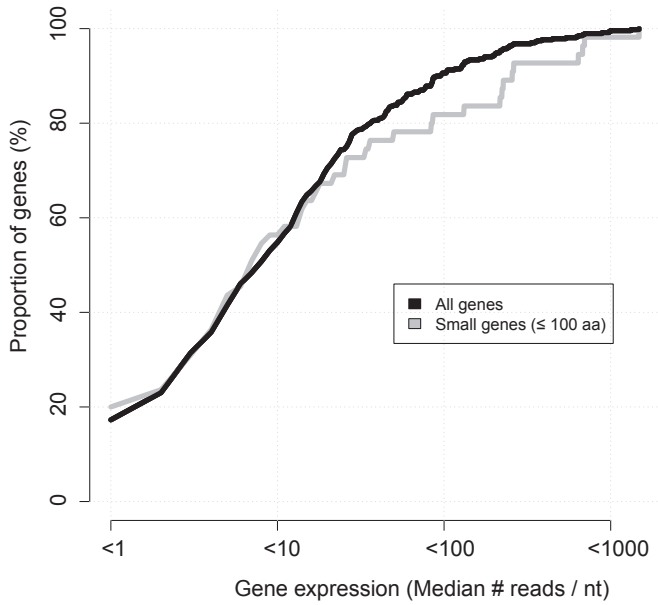


Fig. S5. Cumulative relative frequency distribution of Pithovirus gene expression (black curve). The expression of each gene is measured by the median value of the RNA-seq read coverage per nucleotide. The distribution computed for small ORFs only (<100 aa, gray curve) is very similar, indicating that they correspond to bona fide genes and proteins.

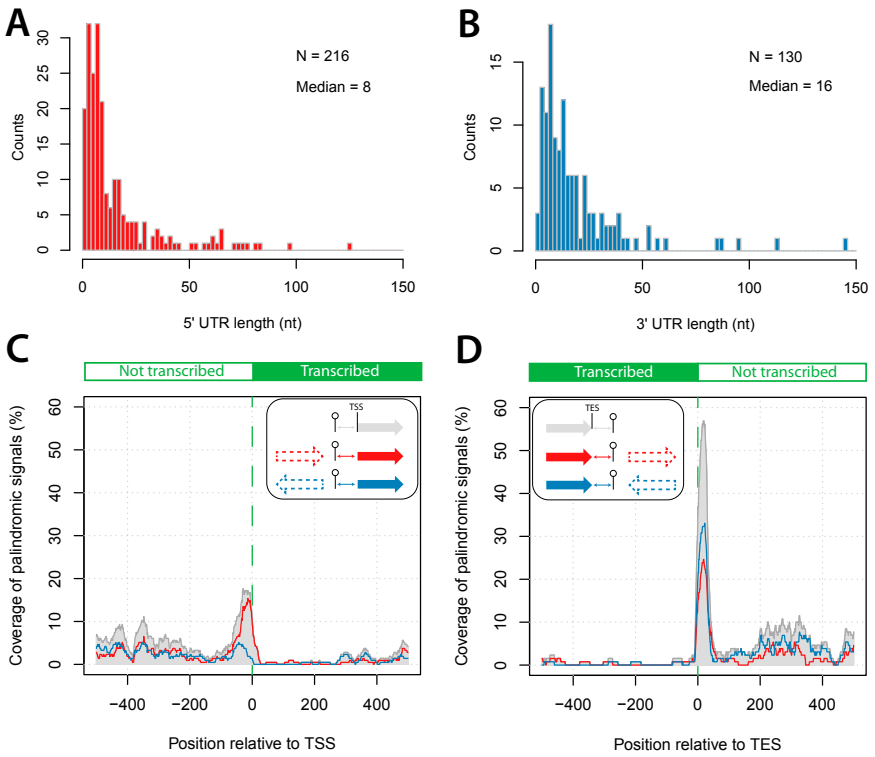
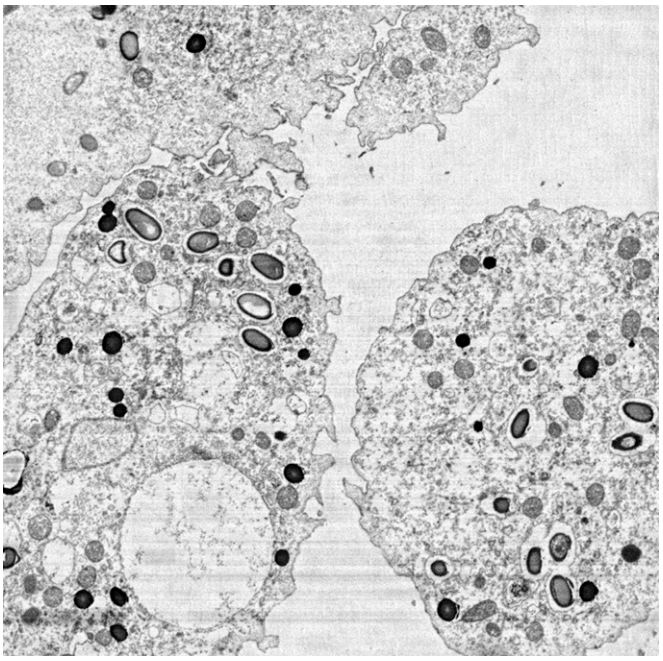


Fig. S6. Structure of the Pithovirus gene boundaries. Distribution of the 5′ (*A*) and 3′ (*B*) UTR lengths. (*C*) Coverage of palindromic motifs relative to the TSS of (gray curve and *Inset* arrow). The signal was deconvoluted in two parts depending on whether the upstream gene of each TSS is in the same orientation (red curve and *Inset* arrows) or in the reverse orientation (blue curve and *Inset* arrows). The enrichment seen at the TSS appears mostly due to palindromic signals associated with the upstream gene TES rather than to the TSS. (*D*) Coverage of palindromic motifs relative to the TES (gray curve and *Inset* arrow). The signal can be deconvoluted in two parts depending on whether the downstream gene of each TES is in the same orientation (red curve and *Inset* arrows) or in the reverse orientation (blue curve and *Inset* arrows). Regardless of the downstream gene, the observed enrichment of palindromic signals is correlated to the presence of a TES. Note that palindromic motifs are enriched in untranscribed (intergenic) regions compared with transcribed regions.



Movie S1. Movie from subsequent sections of *A. castellanii*-infected cells (9 h pi). The dataset was acquired on the Jeol7100F equipped with a Gatan 3View2XP to perform serial block-face scanning electron microscopy. The image dataset consists of 354 layers, each 50-nm thick, giving a total depth of 18 μm. The field of view spans 16 μm and the images were acquired at a resolution of 2,048 × 2,048 pixels with a pixel size of 7.9 nm, using an accelerating voltage of 1.1 kV. The Jeol 7100F uses a field emission gun for high-resolution acquisition. The images were postprocessed with ImageJ (http://imagej.nih.gov/ij/) to normalize intensity changes due to sample charging. Remarkable features have been indicated on the movie using Adobe Premiere software.

Movie S1

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)