Environmental Microbiology (2017) **19**(10), 4022–4034 doi:10.1111/1462-2920.13813

Cedratvirus lausannensis – digging into *Pithoviridae* diversity

Claire Bertelli,[1,2†] Linda Mueller,[1†] Vincent Thomas,[3] Trestan Pillonel,[1] Nicolas Jacquier[1] and Gilbert Greub[1*]

[1]*Institute of Microbiology, University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland.*

[2]*SIB Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland.*

[3]*Technology Research Institute Bioaster, 28, Rue du Docteur Roux, Paris, 75015, France.*

Summary

Amoeba-infecting viruses have raised scientists' interest due to their novel particle morphologies, their large genome size and their genomic content challenging previously established dogma. We report here the discovery and the characterization of *Cedratvirus lausannensis*, a novel member of the *Megavirales*, with a 0.75–1 μm long amphora-shaped particle closed by two striped plugs. Among numerous host cell types tested, the virus replicates only in *Acanthamoeba castellanii* leading to host cell lysis within 24 h. *C. lausannensis* was resistant to ethanol, hydrogen peroxide and heating treatments. Like 30 000-year-old *Pithovirus sibericum*, *C. lausannensis* enters by phagocytosis, releases its genetic content by fusion of the internal membrane with the inclusion membrane and replicates in intracytoplasmic viral factories. The genome encodes 643 proteins that confirmed the grouping of *C. lausannensis* with Cedratvirus A11 as phylogenetically distant members of the family *Pithoviridae*. The 575,161 bp AT-rich genome is essentially devoid of the numerous repeats harbored by Pithovirus, suggesting that these non-coding repetitions might be due to a selfish element rather than particular characteristics of the *Pithoviridae* family. The discovery of *C. lausannensis* confirms the contemporary worldwide distribution of *Pithoviridae* members and the characterization of its genome paves the way to better understand their evolution.

Received 2 March, 2017; revised 24 May, 2017; accepted 30 May, 2017. *For correspondence. E-mail gilbert.greub@chuv.ch; Tel. +41 21 314 49 79; Fax +41 21 314 40 60. †These authors contributed equally to the manuscript.

Introduction

Viruses infecting free-living and parasitic unicellular eukaryotic species have been reported for years (Wang and Wang, 1991). In 1969, virus-like particles were observed in the nucleus of free-living amoebal species *Naegleria gruberi* (Schuster, 1969). Intriguingly, these 100 nm large viruses preferentially appeared when amoebae were grown in the presence of living bacteria rather than in axenic conditions. More recently, a number of 'giant' viruses infecting amoebae or algae have been reported and fully described since the seminal paper of Raoult *et al.* (Raoult *et al.*, 2004) describing *Acanthamoeba polyphaga* Mimivirus (APMV). These giant viruses all belong to the group of Nucleo-Cytoplasmic Large DNA Viruses (NCLDV) – recently proposed to be renamed as the order '*Megavirales'* (Colson *et al.*, 2013) – in which six major families of novel giant viruses emerge: *Megaviridae, Marseilleviridae, Pandoraviridae, Pithoviridae, Molliviridae* and *Faustoviridae* (Thomas *et al.*, 2011; Abergel *et al.*, 2015; Reteno *et al.*, 2015). These large viruses encode functions not commonly found in viruses, including proteins involved in translation (*Mimiviridae*) (Raoult *et al.*, 2004), inteins and histone doublets (*Marseilleviridae*) (Thomas *et al.*, 2011). These rather unique features generated a vast amount of research papers as well as passionate debates on the definition of viruses (Forterre, 2010; Claverie and Abergel, 2016), the classification of giant viruses as a fourth domain of life and the potential role of these viruses in the origin of eukaryotes (López-García and Moreira, 2009; Boyer *et al.*, 2010; Williams *et al.*, 2011; Legendre *et al.*, 2012; Yutin *et al.*, 2014; Erives, 2015).

Representatives of the *Pandoraviridae* and *Pithoviridae* families harbor striking amphora-shaped morphologies (Scheid *et al.*, 2010; Philippe *et al.*, 2013; Legendre *et al.*, 2014; Levasseur *et al.*, 2016) but were shown phylogenetically to be only distantly-related based on reconstruction from both core protein alignment and their gene content (Legendre *et al.*, 2014, 2015). On the one hand, *Pandoravirus dulcis* and *Pandoravirus salinus* are able to grow in

Acanthamoeba castellanii and harbour very large (1μm in length, 0.5 μm in diameter) particles that contain a 1.9–2.5 Mb-long GC-rich genomes (Philippe *et al*., 2013). Another amoebal parasite named KLaHel isolated several years ago (Scheid *et al*., 2008) was also recently shown to represent another *Pandoravirus* strain (Philippe *et al*., 2013; Antwerpen *et al*., 2015; Scheid, 2015). The *Pandoravirus* genomes are highly divergent and all lack many of the core genes usually found in dsDNA viruses, including the gene encoding a major capsid protein (Philippe *et al*., 2013). They only possess several translation-related genes, including two genes encoding amino acid tRNA ligases probably acquired from their amoebal host (Philippe *et al*., 2013). Conversely, they encode many DNA-processing and transcription enzymes, including enzymes that have never been found in viruses before. On the other hand, *Pithovirus sibericum*, isolated from a 30 000 year old layer of Siberian permafrost (Legendre *et al*., 2014), and *Pithovirus massiliensis*, a closely-related contemporary isolate (Levasseur *et al*., 2016), possess an even larger amphora-shaped particle closed by a striped plug and an AT-rich genome with a genetic content that is more similar to *Iridoviruses* and *Marseilleviridae*. Cedratvirus A11 was recently reported as a distant relative of Pithovirus that harbours two corks (one at each extremity of the virion) and a distinct genetic content (Andreani *et al*., 2016). KC5/2 an amoebal parasite discovered more than 10 years ago, might represent a third strain of *Pithoviridae* based on its morphological characteristics (Michel *et al*., 2003; Legendre *et al*., 2014).

In this contribution, we describe the discovery and the characterization of *Cedratvirus lausannensis*. We perform comparative genomics analyses to investigate its gene content and confirm its phylogenetic relatedness to Pithovirus, grouping this novel virus as a new member of the *Pithoviridae* family. The existence of new *Pithoviridae* in the water from the Seine river (France) suggests the widespread distribution of these fascinating viruses. Besides, the low number of repeats identified in these new genomes questions their significance, their role and their rapid expansion in some members of the *Pithoviridae* family.

Results

Cedratvirus lausannensis *discovery*

Over years, amoebal co-culture has become a method of choice to recover new fascinating microbes, whether bacteria or giant viruses (Jacquier *et al*., 2013; Kebbi-Beghdadi and Greub, 2014). While looking for amoeba-resisting bacteria in the drinking water plant of Morsang-sur-Seine (France), we retrieved a new microorganism possessing a thick envelope that was named *Cedratvirus lausannensis* CRIB-75. It replicated in *A. castellanii* ATCC 30010 used as a cell background, finally leading to host cell lysis.

C. lausannensis *has a narrow host range and resists selected inactivation treatments*

The host range of *C. lausannensis* was investigated using different amoebal strains (*A. castellanii*, *A. comandoni* and *D. discoideum*) as well as some arthropod (Sf9 and *Aedes albopictus*), mammalian (Vero, A549 and THP1) and fish (RTG-2) cell lines. The permissivity of *C. lausannensis* assessed by quantitative real-time PCR showed a 37-fold increase in virus genome copy number in *A. castellanii* from time 0 to 24 h, whereas no growth was observed in other tested cell lines (Supporting Information Fig. S1). *A. castellanii* cells were highly susceptible to *C. lausannensis* growth and lysis was observed 24 h post infection. Conversely, none of the mammalian, arthropods or fish cell lines tested were susceptible to *C. lausannensis* infection and no cell lysis was observed, even one week after viral infection.

To test *C. lausannensis* resistance, virus particles were exposed to selected disinfectants and inactivation treatments. Ethanol, hydrogen peroxide and heating at 55°C for 10 min had no significant effect on *C. lausannensis* infectivity (Table 1). Conversely, bleach, glutaraldehyde and orthophtalaldehyde completely abolished infectivity whereas peracetic acid and heating at 65°C for 10 min presented only limited activity with approximatively 2-$\log_{10}$ reduction of infectivity (Table 1).

Morphology and life cycle of C. lausannensis

Scanning and transmission electron microscopy showed that *C. lausannensis* amphora-shaped particles were similar to Pandoravirus (Philippe *et al*., 2013) and Pithoviruses (Legendre *et al*., 2014; Andreani *et al*., 2016) (Fig. 1A). The apex of the particle exhibited the typical stripes of the protruding cork (Fig. 1B) that differentiate Pithovirus from Pandoravirus (Legendre *et al*., 2014). Although rarely observed by transmission electron microscopy due to the difficulty of obtaining perfect transverse sections, the two

Table 1. Reduction in *C. lausannensis* infectivity after selected inactivation treatments.

Inactivation treatment	Log reduction
Bleach 2.5%	> 4.3
Ethanol 70%	0.1 ± 0.1
Glutaraldehyde 2%	> 4.3
Hydrogen peroxide 7.5%	0.2 ± 0.2
Orthophthalaldehyde 0.55%	> 4.3
Peracetic Acid 0.2%	2.0 ± 0.7
55°C	0.1 ± 0.2
65°C	1.8 ± 0.3

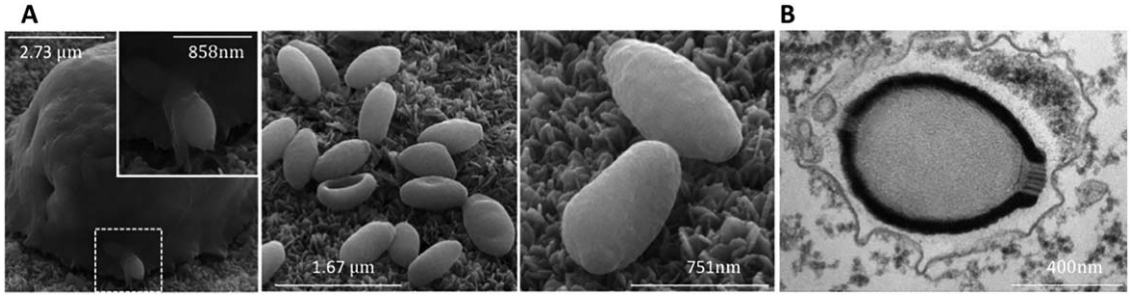

Fig. 1. *C. lausannensis* morphology.
A. Scanning electron microscopy pictures of *C. lausannensis* particles on the grid and *C. lausannensis* entry into the amoeba *A. castellanii* (left panel).
B. Electron microscopy picture of *C. lausannensis* particle exhibiting the characteristic striped protruding cork of Pithoviruses on both sides of the virus particle like Cedratvirus A11, a striated envelope, and an internal membrane that delimits a compartment without substructures.

apexes of the particle harbor the cork (Figs 1B and 3C2), similar to Cedratvirus A11 (Andreani *et al.*, 2016). Moreover, the clear symmetrical shape of the virus particle by scanning electron microscopy also indicates the presence of two plugs, of which one is always clearly visible. The particle harbors a striated envelope with an internal membrane that delimits a compartment without substructures (Fig. 1B). *C. lausannensis* particles were approximately 0.75–1 μm long and 0.4–0.6 μm broad (Fig. 1) – slightly smaller dimensions than Pithovirus (Legendre *et al.*, 2014).

To assess viral growth on *A. castellanii* ATCC 30010 over time, quantitative real-time PCRs were performed at different viral dilutions (1/10, 1/100, 1/1000) (Fig. 2A). Viral growth was lower for small dilutions (1/10 and 1/100) that reached a plateau in only 6–8 h, and higher for the 1/1000 dilution that yielded more viruses in 24 h, probably owing to the occurrence of multiple cycles of infection (Fig. 2A). Electron microscopy (Fig. 2B) confirms the presence of viral particles in amoebal vacuoles after virus internalization, at time 0. Then, 2 and 4 h post infection, *C. lausannensis* underwent an eclipse phase, during which no viral particle was observed. A viral factory containing viral material and unfinished virions was observed 8 h post infection. Finally, at 12 h post infection new virions were assembled, while at 16 and 24 h post infection amoebal lysis occurred, allowing the release of new mature virions ready to infect other amoebae (Fig. 2B).

The life cycle of *C. lausannensis* may thus be represented in six steps (Fig. 3A) observed in more details by scanning, electron and confocal microscopy (Fig. 3B and 3C). Firstly, the virus enters in amoebal cells by phagocytosis (Fig. 3A1 and C1). Once internalized in amoebal vacuoles, the viral content is released in the host cytosol, probably after expulsion of the apical cork (Legendre *et al.*, 2014), and fusion of the virus internal membrane with the vacuole membrane (Fig. 3A2 and C2). During the third step – the eclipse phase – no virion can be observed (Fig. 3A3). Then, a viral factory forms near the intact cell nucleus as seen in the strong DAPI-stained area (blue) by confocal microscopy (Fig. 3B4), suggesting the occurrence of intensive viral DNA replication. Electron microscopy showed electron dense material representing different stages of viral assembly within and at the periphery of virion factories (Fig. 3C4 and C5). Entire virus particles are finally found in the cytoplasm and in vacuoles, as previously described for Pithovirus suggesting that the virus could exit the cell by exocytosis (Legendre *et al.*, 2014). Finally, newly assembled viral particles are released by amoebal lysis (Fig. 3A6 and C6).

C. lausannensis *genome lacks non-coding repeats found in Pithoviruses*

Although the morphology of *C. lausannensis* was largely reminiscent of Pithovirus, *C. lausannensis* harbored an AT-rich dsDNA genome of 575 161 bp that presented only limited similarity at the nucleotide level with other Pithoviruses. Indeed, only 8% of the *C. lausannensis* genome could be aligned to *P. sibericum*, whereas *P. sibericum* and *P. massiliensis* genomes aligned over 70% of their length. Moreover, the genome was surprisingly easily assembled in three main contigs from MiSeq 150 bp paired-end reads, before being closed by additional PCRs and Sanger sequencing to solve two internal repetitive regions as well as terminal repetitions. These first results already suggested the presence of few repeats within the genome of *C. lausannensis*. Similar to Cedratvirus A11 (Andreani *et al.*, 2016), coverage as well as insert size exhibited some statistical variations but were globally consistent across the genome, thereby confirming the absence of large repeats within the sequence, which differs from the abundant repeats discovered in the genome of *Pithovirus sibericum* (Legendre *et al.*, 2014). Only one region between 383 708 bp and 383 885 bp was shown to present repeated palindromic sequences of 53 bp. Similar

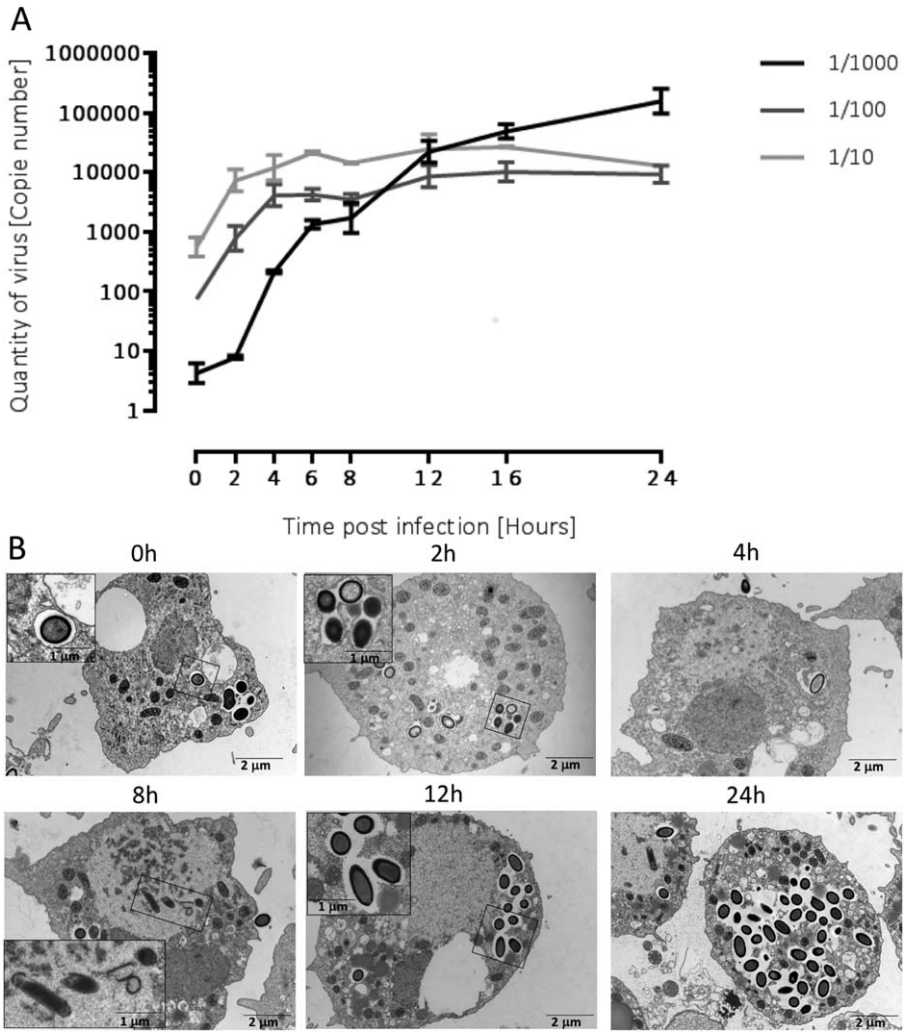

Fig. 2. *C. lausannensis* kinetics in *A. castellanii*.
A. *C. lausannensis* growth was measured at different time points and at different virus dilutions (1/1000, 1/100 and 1/10) using quantitative real-time PCR.
B. Electron microscopy pictures with magnified images of interesting regions in the internal upper and bottom left rectangles. Amoebal vacuoles (electron lucent) containing one or more viral particles (electron dense) were visible at 0 and 2 h post infection. A medium electron dense area, defined as viral factory, filled with genetic viral material (electron dense) and unfinished virions were observed at 8 h post infection. Magnified images at 0 h and 2 h post infection showed different sections of internalized *C. lausannensis* particles. The magnified image at 8 h post infection showed the presence of viral genetic components in the viral factory. The magnified image at 12 h post infection showed different sections of finished virions.

to other viruses, *C. lausannensis* has retained a high coding density (83%) congruent with the lack of noncoding repeats between its predicted 643 proteins. Genome extremities harbor terminal repeats that have been only partially resolved. Together with positive PCR results, they suggest the existence of a circularly permutated linear DNA molecule such as for *Pithovirus sibericum* (Legendre *et al.*, 2014).

Phylogenetic positioning

To confirm the phylogenetic positioning of *C. lausannensis*, the coding sequences of 83 viruses belonging to the *Megavirales* (Supporting Information Table S1) were used to reconstruct groups of orthologous proteins. The proportion of groups of orthologs shared divided by the number of proteins of the smallest genome in the pair was taken as a measure of similarity, as suggested by Snel *et al.* (Snel *et al.*, 1999). The Neighbor-joining tree inferred on the corresponding matrix of distance showed that Cedratvirus A11 and *C. lausannensis* formed a close monophyletic group that branched deeply with *P. sibericum* and *P. massiliensis*, forming an extended family of *Pithoviridae* (Fig. 4A). A maximum-likelihood tree reconstructed from a concatenated alignment of the 8 proteins belonging to the most conserved groups of orthologous with no in-genome paralogs, results in a similar grouping of the four *Pithoviridae* together with *Marseilleviridae, Phycodnaviridae* and *Ascoviridae* (Fig. 4B).

Orthologous proteins shared by Pithoviridae and other Megavirales

As commonly observed with newly discovered families of viruses, as many as 45% of the 643 *C. lausannensis* proteins did not show homologs by BLASTP search in the NCBI non-redundant database (With the exclusion of proteins from Cedratvirus A11: Fig. 5), a rate similar to that observed for Mimivirus (Raoult *et al.*, 2004). The taxonomic affiliation of the best BLAST hits is largely skewed

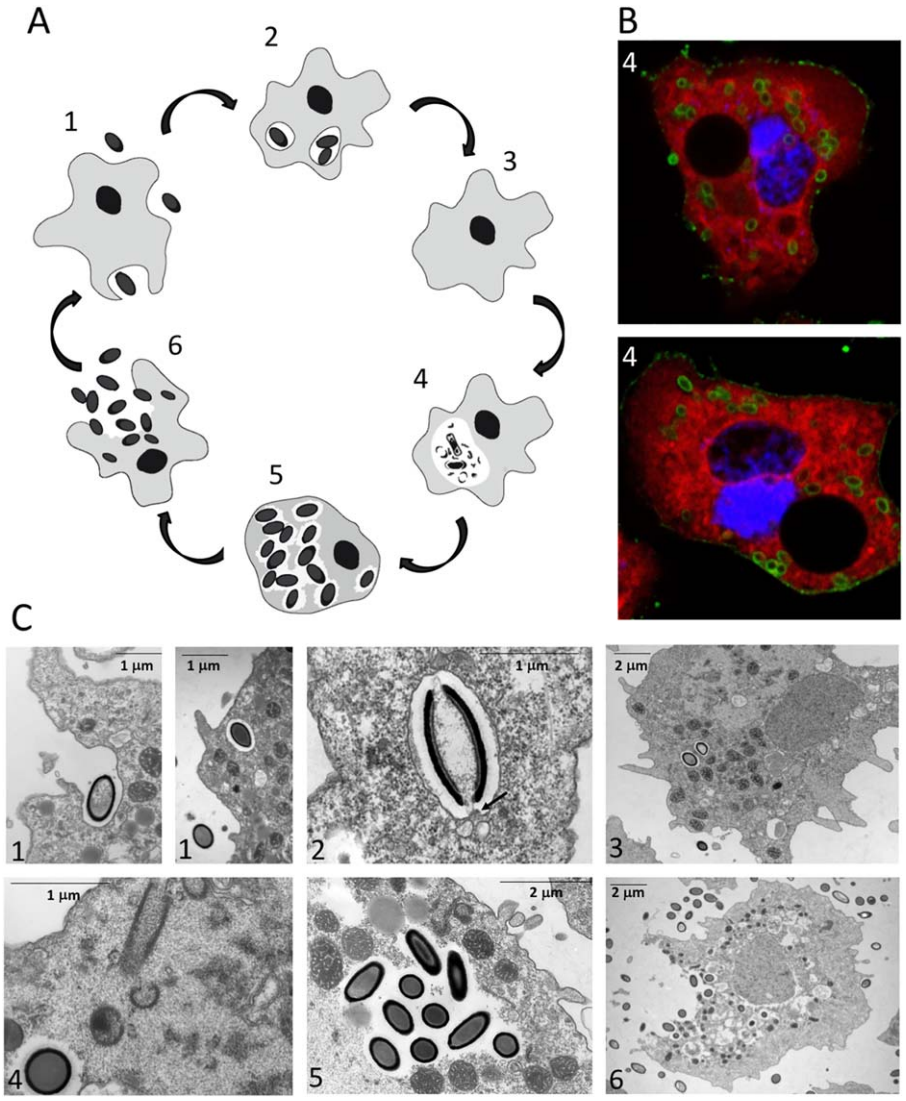

Fig. 3. *C. lausannensis* life cycle. Schematic representation (A) and electron microscopy pictures (C) of *C. lausannensis* life cycle that includes 6 main steps: (1) viral entry; (2) expulsion of the protruding cork (black arrow) and membrane fusion; (3) viral integration, or eclipse phase; (4) DNA replication in the viral factory and virus assembly; (5) accumulation of assembled virions; (6) exit by amoebal lysis. (B) Confocal microscopy pictures 12 h post infection displaying the presence of a viral factory. Viruses were stained with mouse anti- *C. lausannensis* polyclonal antibodies (green), *A. castellanii* were stained with concavalin A (red) and DNA was marked with DAPI (blue).

towards viruses (37%), mainly owing to the relatedness to Pithovirus ($n = 194$), but also *Mimiviridae* ($n = 20$), *Marseilleviridae* ($n = 11$) and Pandoraviruses ($n = 6$). Eukaryota also represent a significant fraction of best BLAST hits (14%), with the notable presence of 13 hits against the amoeba *A. castellanii*.

The groups of orthologous proteins were used to further investigate the conservation of gene content in Megavirales. As shown in Fig. 6A, viruses belonging to the same family typically show small to large groups of signature proteins unique to the family. *Pithoviridae* share 70 groups of orthologs that are not found in other *Megavirales*. Among these proteins of special interest, all except the divergent major capsid protein (Legendre *et al*., 2014), a glycosyltransferase, a polyA polymerase reg subunit and an EGF-like domain-containing protein, have unknown functions. Interestingly, among the few genes common only to *Marseilleviridae* and *Pandoraviridae* is the gene encoding histone doublets (H2A-H2B type) (Thomas *et al*., 2011; Erives, 2015). Finally, only one gene is shared solely by *Pandoraviridae* and *Pithoviridae*: a phosphoglycerate mutase (OG0000492, Supporting Information Table S2), that was previously reported as one of five proteins having closest homologs in Pandoraviruses (Legendre *et al*., 2014). Figure 6B represents a subset of recently-described amoebal viruses that only share 20 groups of orthologous proteins despite their large genome size. As expected, *C. lausannensis* shares most groups of orthologous proteins with *P. sibericum* than other *Megavirales* (Fig. 6B). Among the proposed family *Pithoviridae*, 151 groups of orthologous proteins form a conserved core proteome (Fig. 6C). Both Cedratviruses share most groups of orthologous proteins ($n = 224$) but clearly harbors a distinct genome content (Fig. 6C). The median amino acid identity of orthologs (excluding groups that contain paralogs) is 95% among Cedratviruses and 90% among

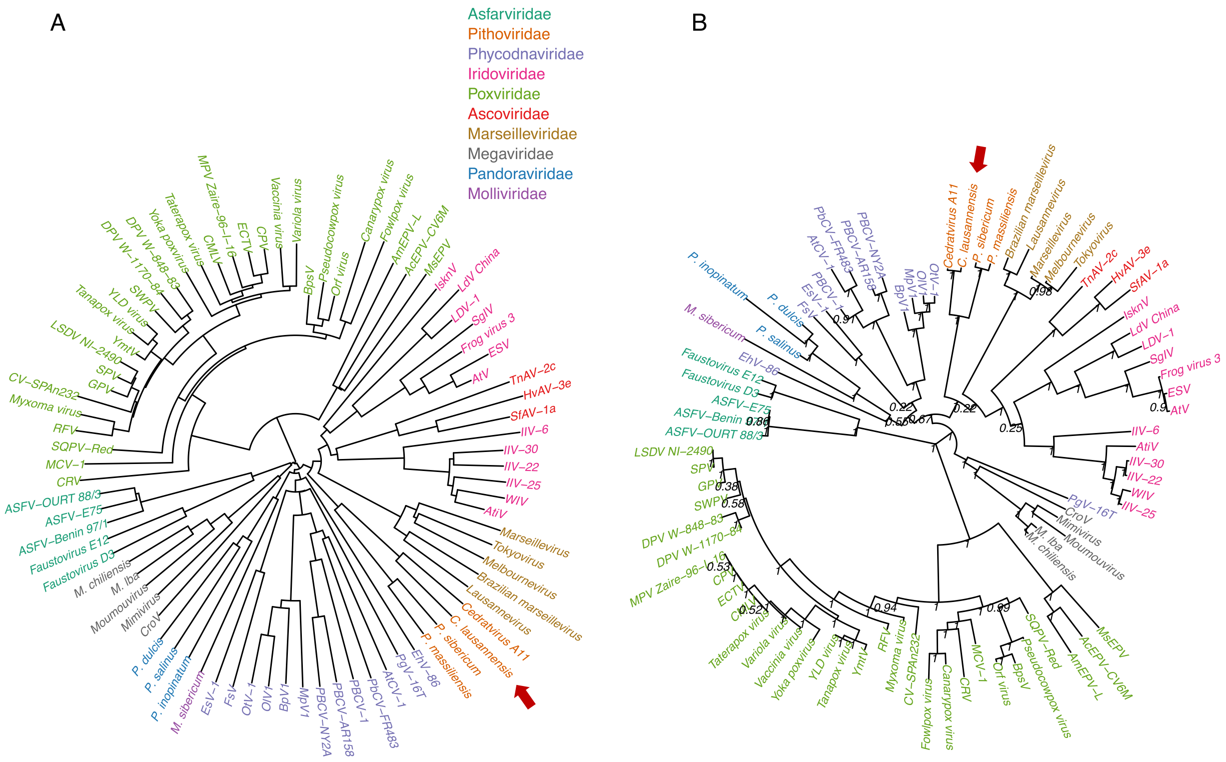

Fig. 4. Phylogenetic clustering of *Pithoviridae*.
A. Neighbor-joining trees based on a gene content distance matrix.
B. Maximum likelihood tree based on a concatenated alignment of the 8 proteins that were found in the highest number of *Megavirales* genomes by the reconstruction of orthologous groups: D6-like helicase, DNA-directed RNA polymerase RPB1 and RPB2, DNA polymerase delta subunit, VV A18-like helicase, VV D5-type primase, transcription elongation factor TFIIS, ribonucleoside diphosphate reductase small chain. Branch support was estimated using the aLRT non-parametric Shimodaira-Hasegawa-like procedure. In both cases, *C. lausannensis* clusters closely with Cedratvirus A11 and more distantly with *P. sibericum* and *P. massiliensis*.

Pithoviruses (Supporting Information Fig. S2). Conversely, proteins from Cedratviruses and Pithoviruses only share approximately 44% amino acid identity (Supporting Information Fig. S2). Both phylogenetic reconstruction and similarity in gene content show that *C. lausannensis* and Cedratvirus A11 are similarly distantly-related as *P. sibericum* and *P. massiliensis*, validating the use of similar higher-level lineages names for both pairs of viruses, but considering them as different species.

Discussion

The recent discovery of Pithoviruses, that share striking morphological similarities with Pandoraviruses, but have very different genomic characteristics, has raised new questions on the evolution of similar traits in giant viruses. The novel giant virus described here shares morphological characteristics such as the protruding cork and an amphora-shaped particle as well as a similar life cycle as Pithovirus (Table 2). It shares the symmetrical double plug on each side of the particle with the recently described Cedratvirus A11 (Andreani *et al.*, 2016), a low number of genomic repeats as well as a large part of its gene content. It has, therefore, been named *Cedratvirus lausannensis*, as it clearly represents a new species within the *Cedratvirus* lineage, which encodes a largely distinct set of genes that are mostly of unknown function.

The extension of the *Pithoviridae* family by the discovery of *C. lausannensis*, presents interesting possibilities to investigate the evolution of this recently described branch of *Megavirales*. Indeed, the description of *P. massiliensis*, a contemporary virus, as opposed to the 30 000 year old *P. sibericum*, has shown a striking genome collinearity despite the presence of numerous repeats (Levasseur *et al.*, 2016). The extreme similarity in gene sequence enabled the authors to estimate the mutation rate of Pithovirus to 3 10^{-6} mutations/site/year, a result similar to other dsDNA viruses (Levasseur *et al.*, 2016). The *C. lausannensis* genome harbored different characteristics, with almost no repetitive sequences and a very distinct gene set. These genomic features are similar to the recently described Cedratvirus (Andreani *et al.*, 2016). The almost complete absence of palindromic repeats within the genome of *C. lausannensis* and Cedratvirus A11 together

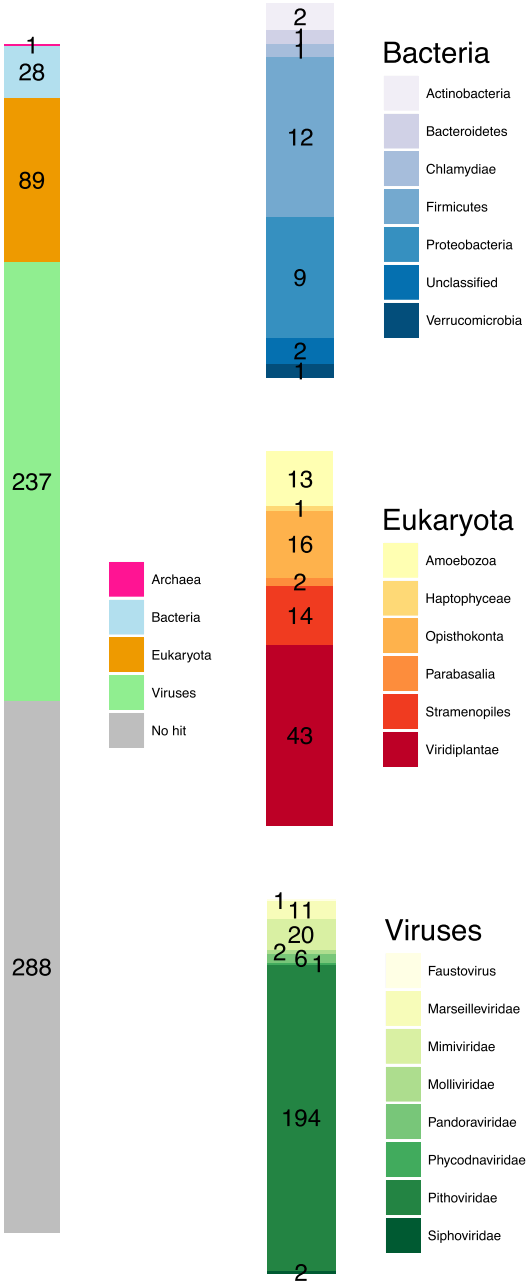

Fig. 5. Taxonomic classification of *C. lausannensis* protein homologs. 45% of *C. lausannensis* proteins have no hit in the non-redundant database as of 15 November 2016 by BLASTP with an e value cutoff of 10^{-5}. The taxonomic affiliation of each hit shows the proximity of *C. lausannensis* to the other Pithoviruses as well as other giant viruses (37%), but also highlights the presence of 89 genes (14%) similar to Eukaryota.

with the similarities in morphology and life cycle to *P. sibericum*, enable to refine the hypotheses made by Legendre *et al.* (Legendre *et al.*, 2014) and we speculate that the unique structure of *P. sibericum* genome could be the result from the multiplication of an invasive selfish DNA sequence rather than due to physical constraints. However, we cannot exclude that these repetitions reflect a specific evolution of the branch leading to *P. sibericum* and *P. massiliensis*, for another still unknown function.

Several studies have previously reported the possibility that amoebae could serve as a niche that would favour horizontal gene transfer (HGT) between its microorganisms (Moliner *et al.*, 2010; Bertelli and Greub, 2012). In addition, the genome of the amoeba *A. castellanii* showed extensive traces of HGTs (Clarke *et al.*, 2013). In this context, it is interesting to note that 89 proteins (14%) of *C. lausannensis* have their closest homologs in Eukaryota – a proportion comparable to the results of Legendre *et al.* (Legendre *et al.*, 2014) for *P. sibericum* – and among those, 13 in *A. castellanii* itself. However, a simple homology investigation does not demonstrate that a HGT has occurred and might also reflect the convergent evolution of viral proteins with their host to better divert its cellular functions.

Overall, *C. lausannensis* and Cedratvirus A11 share a significant proportion of genes with the two Pithoviruses compared to other families of the *Megavirales*. Furthermore, Cedratviruses cluster with Pithoviruses in various phylogenetic reconstructions and therefore form a novel distinct branch in the family *Pithoviridae*. The discovery and the sequencing of other isolates, such as the previously described KC5/2 (Michel *et al.*, 2003), will be of major importance to better define the intrinsic genomic characteristics of this family. Finally, the definition of sets of conserved proteins unique to the *Pithoviridae*, albeit mostly of unknown function, might help mining environmental data and designing probes to identify the presence of *Pithoviridae*.

Interestingly, *C. lausannensis* presents a high resistance to selected inactivation treatments, including ethanol and hydrogen peroxide. It has been reported that other amoeba-infecting viruses such as *Acanthamoeba polyphaga* mimivirus (APMV) are able to survive in various environmental and clinical substrates (Dornas *et al.*, 2014). This could become a concern since it has also been suggested that APMV as well as other amoebae-infecting viruses that belong to *Marseilleviridae* could be associated with human infections (La Scola, 2014; Aherfi *et al.*, 2016a,b). From another perspective, the capacity of amoeba-infecting viruses to survive in freshwater and drinking water could have a strong impact on the colonization of these ecosystems by specific species of amoebae. Since amoebae are well-known hosts of pathogenic bacterial species such as *Legionella pneumophila*, amoebae-infecting viruses could potentially have a negative impact on proliferation of these pathogenic bacterial species in hydric environments. For these reasons, we believe that a systematic investigation of amoebae-infecting viruses susceptibility to a variety of biocidal solutions, as well as an evaluation of their survival capacity in various conditions are warranted.

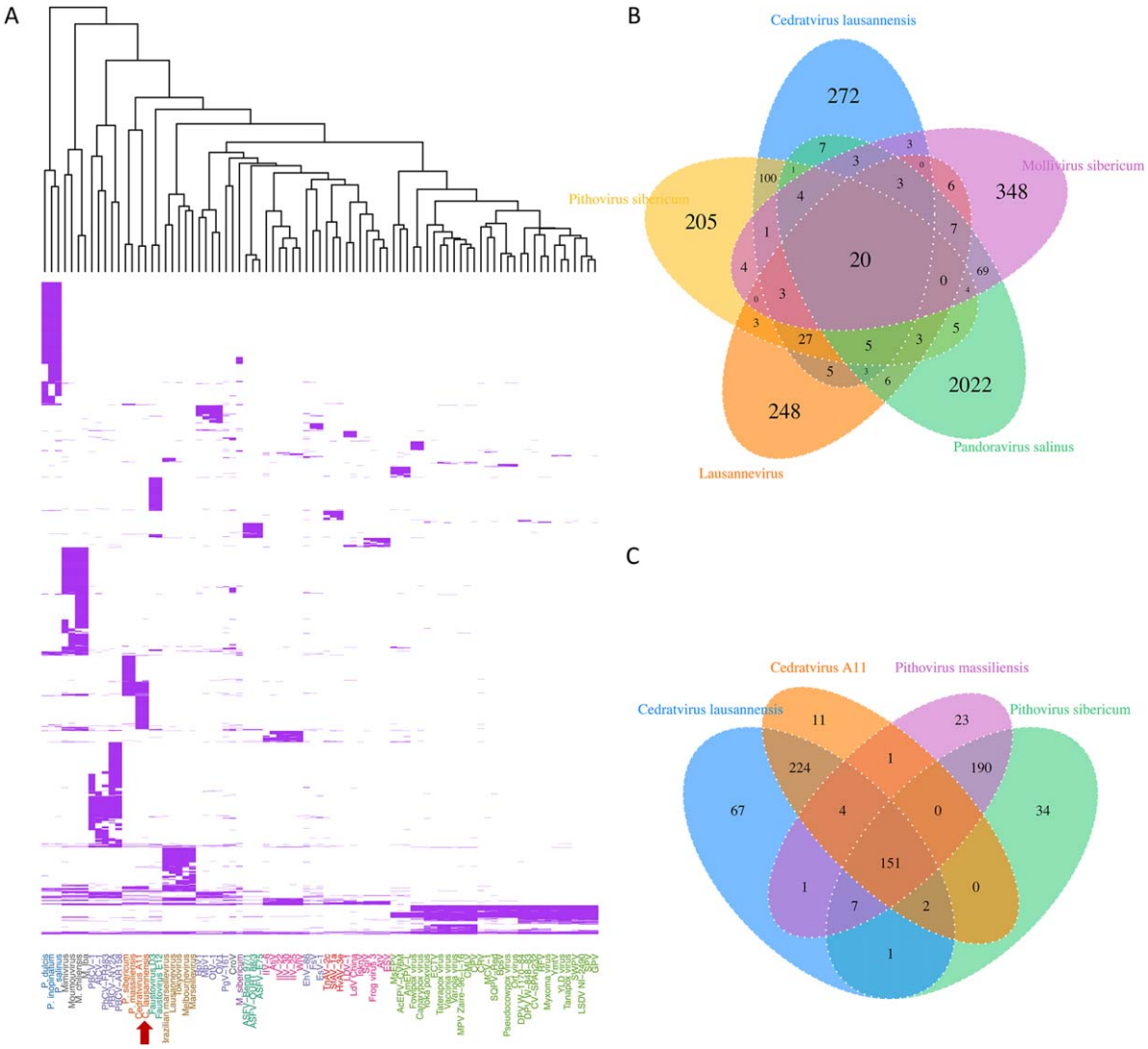

Fig. 6. Groups of orthologous proteins in Megavirales.
A. Hierarchical clustering of viruses based on the presence/absence of groups of orthologous proteins. Venn diagrams with the number of conserved groups of orthologous proteins between recently discovered amoeba-infecting viruses (B) and between Pithovirus isolates, Cedratvirus A11 and C. lausannensis (C).

Experimental procedures

Sample origin

To recover amoebae-resisting microorganisms potentially present in Seine River, 1 L water samples were collected at the entry of the Morsang-sur-Seine (France) drinking water plant. Samples were investigated for the presence of amoebae-resisting microorganisms using amoebal co-culture onto axenic *Acanthamoeba castellanii* amoeba strain ATCC 30010 as described in Thomas *et al.* (10). When

Table 2. Comparison of members of the proposed *Pithoviridae* family.

	Cedratvirus lausannensis CRIB-75	Cedratvirus A11	*Pithovirus sibericum*	*Pithovirus massiliensis*
Particle morphology	Ovoid, amphora-shaped	Ovoid, amphora-shaped	Ovoid, amphora-shaped	Ovoid, amphora-shaped
Cork	Double	Double	Single	Single
Genome size	575,161	589,068	610,033	686,015
GC content	42.8%	42.6%	35.8%	35.4%
Genes	643	574	467	476
Coding density	83%	78.5%	69%	64%

recovered, the viral particles were purified by filtration (5 μm) and ultracentrifugation (35 000 rpm) and frozen at −80°C.

Cell culture

Acanthamoeba castellanii ATCC 30010 and *A. comandoni* strain WBT were cultured in Peptone Yeast-extract Glucose (PYG) medium at 25°C. *Dictyostelium discoideum* DH1–10 was grown in HL5 medium at room temperature. *Aedes albopictus* cells ATCC CRL-1660 were cultured in minimal essential medium (MEM; Gibco Invitrogen, Luzern, Switzerland) supplemented with 10% foetal calf serum (Biochrom, Berlin, Germany) at 28°C + 5% CO_2. Sf9 cells, kindly provided by Prof. F. Martinon (Lausanne, Switzerland), were maintained at 27°C in Grace's Insect Media (Gibco Life Technologies, Luzern, Switzerland) supplemented with 10% foetal calf serum (Biochrom) and diluted at 1:3 once per week. Rainbow trout (Oncorhynchus mykiss) gonad cells (RTG-2), obtained from Prof. H. Segner (Bern, Switzerland) were cultured at 25°C in MEM supplemented with 10% foetal calf serum (Biochrom), 1% nonessential amino acids (Biochrom) and 1% Hepes (BioConcept, Allschwil, Switzerland). Human lung carcinoma cell line A549 ATCC CCL-185 and Vero cells ATCC CCL-81 derived from kidney epithelial cells of the African Green Monkey, were cultured in Dulbecco's modified minimal essential medium (DMEM; Gibco Life Technologies) supplemented with 10% foetal calf serum (Biochrom) at 37°C + 5% CO_2. Human monocytic cell line THP1, kindly provided by Prof. T. Roger (Lausanne, Switzerland), was maintained in RPMI 1640 media (PAA) with L-Glutamine (GE HealthCare, Chicago, USA), supplemented with 10% foetal bovine serum (FBS, Connectorate, Dietikon, Switzerland) at 37°C + 5% CO_2.

Virus culture and purification

Confluent flasks of *A. castellanii* ATCC 30010 grown in Peptone Yeast-extract Glucose (PYG) at 32°C were infected with one frozen aliquot of *C. lausannensis*. One week after infection, flasks were filtered at 5 μm and 1 ml of the filtrate was used to infect fresh amoebal flasks. Five days later, viruses were harvested and flask supernatant was centrifuged at 5000 g for 15 min. The supernatant was then collected and filtered at 5 μm to remove residual amoebal cells. After filtration, the viruses were centrifuged at 35 000 g for 1 h and the pellet was resuspended in 1 ml of DNA-free water. Genomic DNA was isolated from purified virus particles with the Wizard Genomic DNA purification kit (Promega Corporation, Madison, WI, USA).

Infection procedure

Amoebal cells were harvested from Falcon culture flasks by vigorous shaking, washed with PAS medium and seeded in RPMI 1640 (PAA) supplemented with 10% foetal calf serum (Biochrom) (*A. castellanii* and *A. comandoni*) or in HL5 (*D. discoideum*) at 5×10^5 cells per well in 24-wells microplates.

RTG-2, *Ae. albopictus*, Vero and A549 cells were harvested from Corning culture flasks with 0.25% trypsin (Sigma, Buchs, Switzerland), washed with fresh medium and seeded at 5×10^5 cells per well in 24-wells microplates. Sf9 cells were harvested from Corning culture flask by pipet flushing and seeded at 2.5×10^5 cells in 24-wells plate (Corning). THP-1 cells were seeded in 24-wells plate (Corning) at a density of 2.5×10^5 cells/well, and differentiated in macrophages with 50 nM of phorbol 12-myristate 13-acetate (PMA, Sigma-Aldrich, Buchs, Switzerland) during 48 h at 37°C. Differentiated cells were washed three times with RPMI 1640 (PAA) before infection.

Cells were infected with *C. lausannensis* diluted in each appropriate culture media described above.

Plates were then centrifuged at 1790 g for 10 min at room temperature. After 30 min of incubation at the appropriate culture temperature (see cells culture), cells were washed with fresh media to remove non-internalized viruses and time point 0 h was collected. Plates were then incubated at the appropriate culture temperature. Harvest was performed at different times post infection and samples were stored at −20°C until genomic DNA extraction was performed.

Growth kinetics of C. lausannensis *in* A. castellanii *ATCC 30010 strain*

A. castellanii ATCC 30010 were infected with *C. lausannensis* in 24-well microplates as described above. Three different *C. lausannensis* dilution in RPMI 1640 (PAA) were tested: 1/10, 1/100 and 1/1000. Harvest was performed at different times post infection and samples were stored at −20°C. DNA was extracted from 50 μl of the co-culture with the Wizard Genomic DNA purification kit (Promega Corporation) and eluted in 100 μl of elution buffer. Further, *C. lausannensis* DNA was detected by quantitative real-time PCR. Primers and probe for these assays were designed with Geneious software v.6 (Biomatters, Aukland, New Zealand). The amplified DNA fragment is 382 bp long, located in the region of the D5 helicase gene. Primers and probe (Eurogentec, Seraing, Belgium) used were CvF (5′-CGTGGCCATCCTCTTGCTTA-3′), CvR (5′-CTCTCCCTCCAGACGGGTAA-3′) and CvS (5′-FAM-TGCCACTTTCTGTTCTCCCG-BHQ1–3′). Quantitative real-time PCR assays were performed in a total volume of 20 μl, with 10 μl of iTaq supermix (Bio-Rad, Reinach, Switzerland), 100 nM of probe, 200 nM of forward primer, 200 nM of reverse primer, molecular biology grade water (Sigma-Aldrich) and 5 μl of DNA sample. Cycling conditions were enzyme activation at 95°C during 3 min and 45

cycles with 15 s at 95°C and 1 min at 60°C. Quantitative PCR assays were performed on the Step One (Applied Biosystems, Zug, Switzerland) PCR system. Water was used as a negative control. Positive control and quantification were assessed using a plasmid harbouring the D5 helicase gene. The plasmid was constructed with TOPO® TA Cloning® Kit (Invitrogen, Luzern, Switzerland) as provided by the manufacturer. All samples, standard and controls were analysed in duplicate. Growth kinetics were also analysed by electron microscopy, immunofluorescence and confocal microscopy.

Antibodies, immunofluorescence and confocal microscopy

Anti-*C. lausannensis* antibodies were purchased from Eurogentec, using the heat-inactivated virus as antigen for the 3-month standard program. Infected *A. castellanii* ATCC 30010 were fixed with MetOH at −20°C for 5 min and washed 3 times with PBS. After 20 min of permeabilization in blocking solution (PBS, 10% FCS, 0.04% NaN_3 and 0.1% saponin), cells were incubated for 1 h in mouse anti-*C. lausannensis* antibodies at a 1/1000 dilution in blocking solution. Then, cells were washed 3 times in PBS and incubated for 1 h in AlexaFluor 488 anti-mouse antibodies diluted 1:1000 (Invitrogen), concanavalin A diluted 1/50 (Invitrogen) and DAPI. Finally, after 3 washes with PBS and 1 with H_2O, we mounted coverslips with Moewiol. Fluorescence was analysed with a confocal fluorescence microscope (Zeiss LSM 510 Meta, Jena, Germany).

Electron microscopy

A. castellanii ATCC 30010 cells were infected with 1/10 *C. lausannensis* dilution as described above. After 0, 2, 4, 8, 12, 16 and 24 h cells were harvested, centrifuged 5 min at 1790 g and resuspended in phosphate buffer 0.1 M, 4% PFA (Electron Microscope Science, Hatfield, England) and 0.2% glutaraldehyde (Fluka Biochemika, Buchs, Switzerland) overnight. The day after, samples were washed and resuspended in phosphate buffer 0.1 M. After dehydration by successive washes in increasing acetone concentrations (50–100%), samples were incubated 1 h in equal volumes of acetone and epon, and were further incubated overnight only in epon. Epoxy resin (Fluka Biochemika) was added to samples to form blocks which were cut in thin sections by LKB 2088 Ultrotome microtome, deposited on copper grids coated with formvar (Sigma-Aldrich) and stained for 10 min with a solution of methanol–uranyl acetate and lead nitrate with sodium citrate in water. Transmission electron microscope Philips EM 201 C (Philips, Eindhoven, Netherlands) was used to analyse grids.

Scanning electron microscopy

For SEM, co-cultures of virus-infected *A. castellanii* ATCC 30010 were harvested two hours post-infection, washed in PAS buffer and immersed in 2.5% (v/v) glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.4, overnight at 4°C. The fixative was removed and samples were rinsed three times for 10 min in the 0.1 M sodium cacodylate solution, pH 7.4. The samples underwent progressive dehydration by soaking in a graded series of ethanol (i.e., 50–100%) before critical-point drying under CO_2. Samples were mounted on aluminium stubs with double-sided adhesive carbon discs (10 mm diameter) and sputter coated with gold-palladium (Polaron SC7640; Elexience, Verrières-le-buisson, France) for 200 s at 10 mA. Samples were visualized at 2 kV by field emission gun scanning electron microscopy as secondary electron images with a Hitachi S4500 instrument (Elexience, Verrières-le-buisson, France).

Host-free virus inactivation tests

For chemical inactivation tests, titrated suspensions of viruses (10^7–10^8 infecting units/ml) in sterile distilled water were diluted 1/10 in disinfectant to be tested and incubated for 10 min at room temperature. Viruses were then resuspended in Dey-Engley neutralizing broth for 5 min (12). For heat inactivation test, viruses in sterile distilled water were diluted 1/10 in pre-heated sterile distilled water and incubated for 10 min at 55°C or 65°C. They were then resuspended in 4°C PBS. In both cases viruses were recovered by centrifugation at 14 000 rpm for 15 min, resuspended in PYG and serial-diluted onto *A. castellanii* ATCC 30010 in 96 wells plates (1 column per dilution) and plates were centrifuged at 1500 rpm for 30 min. After 5 day incubation at 33°C, plates were observed for amoebal lysis and log reductions were calculated using the Spearman Kärber method. Each test was repeated three times.

Genome sequencing and assembly

C. lausannensis dsDNA was extracted from frozen particles using Wizard Genomic DNA purification kit (Promega Corporation) and eluted in 100 μl of elution buffer. Nextera DNA Sample Preparation Kit and Nextera Index Kit (Illumina, San Diego, CA, USA) were used to prepare a library of genomic DNA (gDNA). Paired-end 2 × 150 bp sequencing of gDNA libraries was performed on a MiSeq Desktop Sequencer (Illumina, San Diego). Read quality was checked with fastqc (Andrews, 2010) and reads were trimmed and filtered by fastq-mcf (*code.google.com/p/ea-utils/; parameters –max-ns 1 -l 150 - L150 –mean-qual 36 –q 32*). The resulting reads were assembled by SPAdes (Bankevich *et al.*, 2012) with varying

parameters (kmer: 31–71) yielding a best assembly with one major contig and two smaller contigs. PCR primers designed at contig extremities were used to test contig orientation by PCR and the resulting products were sequenced by Sanger technology on a ABI3130xl instrument (Applied Biosystems, Life technologies). Reads reassembly was performed with Phrap and visualized with Consed (Gordon *et al*., 1998) to allow curation and editing of the final sequence. To estimate the number of terminal repetitions, primers designed within each repeat were used by PCR enabling to count at least 10 repeats on 5′ end and 9 repeats on 3' end. Although the different terminal repetitions do not enable circularization *in silico*, PCRs designed to test the genome circularization were positive. The terminal repetitions might not be completely solved due to the short read length of sequencing and the highly repetitive nature of these regions. Palindromic repeats were searched for using EMBOSS palindrome (Rice *et al*., 2000). The genome was aligned to other Pithovirus using Mauve (Darling *et al*., 2004). The final genome sequence was annotated with Prokka (Seemann, 2014) using a database of 74 available genomes of NCLDVs (Supporting Information Table S1, 1–74).

Genomic and phylogenetic analyses

All protein-coding genes of *C. lausannensis* were submitted to a BLASTP (Altschul *et al*., 1997) against the non-redundant database as of 17 November 2016 with an *e*-value cutoff of 10^{-5}. For each best hit, NCBI taxonomy was automatically retrieved using homemade scripts. Proteins of 83 virus genomes (Supporting Information Table S1) were clustered into 4203 groups of orthologs using OrthoFinder 0.4 (Emms and Kelly, 2015). To confirm the phylogenetic positioning of *C. lausannensis*, the presence/absence of groups of orthologous proteins in all genomes were used for hierarchical clustering and transformed into a distance matrix to infer Neighbor-Joining phylogenetic trees using R (Cran, 2010) with package Phangorn 2.0.4 (Schliep, 2011), ape 3.5 (Popescu *et al*., 2012), gplots 3.0.1 (http://CRAN.R-project.org/package=gplots). The number of groups of orthologs shared divided by the number of proteins encoded by the smallest genome in the pair was taken as a measure of similarity *s*, as suggested by (Snel *et al*., 1999). A Neighbor-joining tree was then inferred based on a distance measure calculated as $1 - s$. Finally, a concatenated alignment of the 8 proteins conserved among the largest number of viruses was used to reconstruct a maximum likelihood phylogeny using PhyML 3.1 (Guindon *et al*., 2010) using the model VT, alpha 1.768, inv 0.008 estimated as the best model using the Bayesian information Criteria in ProtTest 3.4.2 (Darriba *et al*., 2011).

Availability of data and materials

The genome sequence of *C. lausannensis* CRIB-75 has been deposited in the European Nucleotide Archive (ENA) under the project number PRJEB18669.

Acknowledgements

We would like to thank Sébastien Aeby for his technical help and the sequencing of *C. lausannensis* genome. We also acknowledge the Electron Microscopy Facility and the Cellular Imaging Facility platform of the University of Lausanne for assisting with electron and confocal microscopy. We thank Gemma R. Hoad for editing the final version of the manuscript to correct remaining English errors. This work was supported by a grant from the Faculty of Biology and Medicine of the University of Lausanne, as well as by support from Suez-ONDEO (Paris, France). The authors declare that they have no competing interests.

References

Abergel, C., Legendre, M., and Claverie, J.-M. (2015) The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol Rev* **39:** 779–796.

Aherfi, S., Colson, P., Audoly, G., Nappez, C., Xerri, L., Valensi, A., *et al*. (2016a) Marseillevirus in lymphoma: a giant in the lymph node. *Lancet Infect Dis* **16:** e225–e234.

Aherfi, S., Colson, P., and Raoult, D. (2016b) Marseillevirus in the pharynx of a patient with neurologic disorders. *Emerg Infect Dis* **22:** 2008–2010.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Andreani, J., Aherfi, S., Bou Khalil, J., Di Pinto, F., Bitam, I., Raoult, D., *et al*. (2016) Cedratvirus, a double-cork structured giant virus, is a distant relative of Pithoviruses. *Viruses* **8:** 300.

Andrews, S. (2010) FastQC: A quality control tool for high throughput sequence data. *Babraham Bioinf* 1. https://www.bioinformatics.babraham.ac.uk/projects/fastqc

Antwerpen, M.H., Georgi, E., Zoeller, L., Woelfel, R., Stoecker, K., and Scheid, P. (2015) Whole-genome sequencing of a pandoravirus isolated from keratitis-inducing acanthamoeba. *Genome Announc* **3:** e00136–15.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al*. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19:** 455–477.

Bertelli, C., and Greub, G. (2012) Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. *Front Cell Infect Microbiol* **2:** 110.

Boyer, M., Madoui, M.A., Gimenez, G., La Scola, B., and Raoult, D. (2010) Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4th domain of life including giant viruses. *PLoS One* **5:** e15530.

Clarke, M., Lohan, A.J., Liu, B., Lagkouvardos, I., Roy, S., Zafar, N., *et al*. (2013) Genome of Acanthamoeba

castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol* **14:** R11.

Claverie, J.-M., and Abergel, C. (2016) Giant viruses: The difficult breaking of multiple epistemological barriers. *Stud. Hist. Philos. Biol. Biomed. Sci.* **59:** 89–99.

Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D.K., *et al.* (2013) "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch Virol* **158:** 2517–2521.

Cran. (2010) The Comprehensive R Archive Network. Wiley Interdiscip Rev Comput Stat. http://cran.r-project.org/

Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14:** 1394–1403.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27:** 1164–1165.

Dornas, F.P., Silva, L.C.F., de Almeida, G.M., Campos, R.K., Boratto, P.V.M., Franco-Luiz, A.P.M., *et al.* (2014) Acanthamoeba polyphaga mimivirus stability in environmental and clinical substrates: Implications for virus detection and isolation. *PLoS One* **9:** e87811.

Emms, D.M., and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16:** 157.

Erives, A.J. (2015) Eukaryotic core histone diversification in light of the histone doublet and DNA topo II genes of Marseilleviridae. *bioRxiv.* https://doi.org/10.1101/022236

Forterre, P. (2010) Giant viruses: Conflicts in revisiting the virus concept. *Intervirology* **53:** 362–378.

Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res* **8:** 195–202.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate Maximum-Likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* **59:** 307–321.

Jacquier, N., Aeby, S., Lienard, J., and Greub, G. (2013) Discovery of new intracellular pathogens by amoebal coculture and amoebal enrichment approaches. *J Vis Exp* **80:** e51055.

Kebbi-Beghdadi, C., and Greub, G. (2014) Importance of amoebae as a tool to isolate amoeba-resisting microorganisms and for their ecology and evolution: The Chlamydia paradigm. *Environ Microbiol Rep* **6:** 309–324.

La Scola, B. (2014) Looking at protists as a source of pathogenic viruses. *Microb Pathog* **77:** 131–135.

Legendre, M., Arslan, D., Abergel, C., and Claverie, J.-M. (2012) Genomics of Megavirus and the elusive fourth domain of Life. *Commun Integr Biol* **5:** 102–106.

Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., *et al.* (2014) Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci USA* **111:** 4274–4279.

Legendre, M., Lartigue, A., Bertaux, L., Jeudy, S., Bartoli, J., Lescot, M., *et al.* (2015) In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. *Proc Natl Acad Sci USA* **112:** 201510795.

Levasseur, A., Andreani, J., Delerce, J., Bou Khalil, J., Robert, C., La Scola, B., and Raoult, D. (2016) Comparison of a modern and fossil Pithovirus reveals its genetic conservation and evolution. *Genome Biol Evol* **8:** 2333–2339.

López-García, P., and Moreira, D. (2009) Yet viruses cannot be included in the tree of life. *Nat Rev Microbiol* **7:** 615–617.

Michel, R., Müller, K.-D., Schmid, E.N., Zöller, L., and Hoffmann, R. (2003) Endocytobiont KC5/2 induces transformation into sol-like cytoplasm of its host Acanthamoeba sp. as substrate for its own development. *Parasitol Res* **90:** 52–56.

Moliner, C., Fournier, P.-E., and Raoult, D. (2010) Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol Rev* **34:** 281–294.

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., *et al.* (2013) Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* **341:** 281–286.

Popescu, A.-A., Huber, K.T., and Paradis, E. (2012) ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* **28:** 1536–1537.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., *et al.* (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* **306:** 1344–1350.

Reteno, D.G., Benamar, S., Khalil, J.B., Andreani, J., Armstrong, N., Klose, T., *et al.* (2015) Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *J Virol* **89:** 6585–6594.

Rice, P., Longden, I., Bleasby, A., Harper, R., Etzold, T., Argos, P., *et al.* (2000) EMBOSS: The European molecular biology open software suite. *Trends Genet* **16:** 276–277.

Scheid, P. (2015) Viruses in close associations with free-living amoebae. *Parasitol Res* **114:** 3959–3967.

Scheid, P., Zöller, L., Pressmar, S., Richard, G., and Michel, R. (2008) An extraordinary endocytobiont in Acanthamoeba sp. isolated from a patient with keratitis. *Parasitol Res* **102:** 945–950.

Scheid, P., Hauröder, B., and Michel, R. (2010) Investigations of an extraordinary endocytobiont in Acanthamoeba sp.: Development and replication. *Parasitol Res* **106:** 1371–1377.

Schliep, K.P. (2011) phangorn: Phylogenetic analysis in R. *Bioinformatics* **27:** 592–593.

Schuster, F.L. (1969) Intranuclear virus-like bodies in the amoeboflagellate Naegleria gruberi. *J Protozool* **16:** 724–727.

Seemann, T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068–2069.

Snel, B., Bork, P., and Huynen, M.A. (1999) Genome phylogeny based on gene content. *Nat Genet* **21:** 108–110.

Thomas, V., Bertelli, C., Collyn, F., Casson, N., Telenti, A., Goesmann, A., *et al.* (2011) Lausannevirus, a giant amoebal virus encoding histone doublets. *Environ Microbiol* **13:** 1454–1466.

Wang, A.L., and Wang, C.C. (1991) Viruses of the protozoa. *Annu Rev Microbiol* **45:** 251–263.

Williams, T.A., Embley, T.M., and Heinz, E. (2011) Informational gene phylogenies do not support a fourth domain of

life for nucleocytoplasmic large DNA viruses. *PLoS One* **6:** e21080.
Yutin, N., Wolf, Y.I., and Koonin, E.V. (2014) Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* **466–467:** 38–52.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. List of the 83 virus genome sequences included in the present comparative analysis.
Table S2. Groups of orthologous proteins for the 83 viruses.

Fig. S1. *C. lausannensis* host range. *C. lausannensis* permissivity was assessed *in Acanthamoeba castellanii*, *D. discoideum*, *A. comandoni*, *Aedes albopictus*, Sf9, RTG-2, A549, THP1 and Vero cells. Viral growth was measured by quantitative real-time PCR. Infected cells were incubated at different temperatures, according to cell culture conditions. *C. lausannensis* growth was observed only in *A. castellanii* ATCC 30010.
Fig. S2. Pairwise amino acid identity of *Pithoviridae* proteomes. The pairwise amino acid identity was calculated based on multiple sequence of groups of orthologous proteins from the family *Pithoviridae*, excluding those groups that contained several paralogs for any of the four viruses represented here.