

**Serveur Académique Lausannois SERVAL [serval.unil.ch](http://serval.unil.ch)**

## **Author Manuscript**

**Faculty of Biology and Medicine Publication**

**This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.**

Published in final edited form as:

**Title:** Lausannevirus, a giant amoebal virus encoding histone doublets.

**Authors:** Thomas V, Bertelli C, Collyn F, Casson N, Telenti A, Goesmann A, Croxatto A, Greub G

**Journal:** Environmental microbiology

**Year:** 2011 Jun

**Volume:** 13

**Issue:** 6

**Pages:** 1454-66

**DOI:** 10.1111/j.1462-2920.2011.02446.x

In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.

1 **Title**

2 **Lausannevirus, a giant amoebal virus encoding histone doublets**

3

4 **Authors**

5 Vincent Thomas\*<sup>1†</sup>, Claire Bertelli\*<sup>1</sup>, François Collyn<sup>1</sup>, Nicola Casson<sup>1</sup>, Amalio Telenti<sup>1</sup>, Alexander  
6 Goesmann<sup>2</sup>, Antony Croxatto<sup>1</sup>, Gilbert Greub<sup>°1</sup>

7 <sup>†</sup>Equal contribution

8

9 **Affiliations**

10 <sup>1</sup> Institute of Microbiology, University Hospital Center and University of Lausanne, 1011 Lausanne,  
11 Switzerland, <sup>2</sup> Center for Biotechnology (CeBiTec), Bielefeld University, 33501 Bielefeld, Germany

12 <sup>†</sup> Current affiliation: STERIS SA R&D - 18, Route du Panorama, 92260 Fontenay-aux-Roses, France

13

14 **°Corresponding author**

15 Gilbert Greub, MD PhD

16 Institute of Microbiology

17 University of Lausanne

18 1011 Lausanne

19 SWITZERLAND

20 Phone : +41-21-314 49 79

21 Fax : +41-21-341 40 60

22 e-mail : gilbert.greub@chuv.ch

23

24 **Running title**

25 A giant virus encoding histone doublets

**26 Summary**

27 Large viruses infecting algae or amoebae belong to the Nucleocytoplasmic Large DNA Viruses  
28 (NCLDV) and present genotypic and phenotypic characteristics that have raised major interest among  
29 microbiologists. Here, we describe a new large virus discovered in *Acanthamoeba castellanii* co-  
30 culture of an environmental sample. The virus, referred to as Lausannevirus, has a very limited host  
31 range, infecting *Acanthamoeba* spp. but being unable to infect other amoebae and mammalian cell  
32 lines tested. Within *A. castellanii*, this icosahedral virus of about 200 nm exhibits a development cycle  
33 similar to Mimivirus, with an eclipse phase two hours post-infection and a logarithmic growth leading to  
34 amoebal lysis in less than 24 hours. The 346 kbp Lausannevirus genome presents similarities with the  
35 recently described Marseillevirus, sharing 89% of genes, and thus belongs to the same family as  
36 confirmed by core gene phylogeny. Interestingly, Lausannevirus and Marseillevirus genomes both  
37 encode three proteins with predicted histone folds, including two histone doublets, that present  
38 similarities to eukaryotic and archaeal histones. The discovery of Lausannevirus and the analysis of its  
39 genome provide some insight in the evolution of these large amoebae-infecting viruses.

## 40 Introduction

41

42 The presence of virus-like particles (VLPs) in protozoa, algae and fungi has been reported for several  
43 decades (for a review see (Wang and Wang, 1991)). Among these VLPs, large particles up to 200 nm  
44 long and 100 nm wide were observed in *Acanthamoeba* sp. cytoplasm (Vickerman, 1962), in various  
45 *Giardia* species (Sogayar and Gregorio, 1986; Sogayar and Gregório, 1998) and in *Blastocystis*  
46 species (Stenzel and Boreham, 1997). Since these descriptions were based on electron microscopy,  
47 the exact nature of these particles remained unknown.

48

49 The interest in viruses infecting amoebae has been stimulated by the description of new giant viruses  
50 that infect several amoebal species of the genus *Acanthamoeba* (La Scola *et al.*, 2003; Raoult *et al.*,  
51 2004; Boyer *et al.*, 2009). They belong to the NucleoCytoplasmic Large DNA Viruses (NCLDV) that  
52 are divided into two principal lineages: the poxvirus-asfarvirus group, and the iridovirus-phycodnavirus  
53 group (Iyer *et al.*, 2006). Species from these groups share unique features in virology, such as  
54 aminoacyl-tRNA synthases, DNA site-specific endonucleases, glycosylating enzymes, different types  
55 of introns, and numerous RNA polymerase subunits (Van Etten, 2003; Raoult *et al.*, 2004; Allen *et al.*,  
56 2006). The presence of many genes related to those of eukaryotes in the genome of Mimivirus  
57 generated debate on the pre- or post-eukaryotic emergence of these viruses and their potential role in  
58 the origin of eukaryotes (Raoult *et al.*, 2004; Moreira and Brochier-Armanet, 2008). Amoebae are also  
59 infected by various intracellular bacteria (Greub and Raoult, 2004; Lamoth and Greub, 2010) and it  
60 has been reported that these bacteria have exchanged genes with their hosts and with other co-  
61 infecting microorganisms (*i.e.* other intra-amoebal bacteria and viruses), suggesting a role for  
62 amoebae as a “melting-pot” or a “shuttle” facilitating gene exchange through the different domains of  
63 life (Moliner *et al.*, 2010; Thomas and Greub, 2010; Thomas *et al.*, 2010). Due to the extreme  
64 versatility of gene exchanges occurring between amoebae and their intracellular guests, new genome  
65 sequences of viruses infecting amoebae are needed to better understand mechanisms at play.

66 In this paper, we describe a new large amoebae-resisting virus that we discovered in Seine river water  
67 using amoebal co-culture. This virus exhibits significant similarities with the recently described NCLDV  
68 Marseillevirus, allowing detailed comparative genomic analysis and shedding some insight into the  
69 evolution and the biodiversity of giant viruses.

## 70 **Results and discussion**

### 71 ***Discovery of the virus and description of its life cycle***

72 An amoebal co-culture sample collected in 2005 from the Seine river (France) during previous  
73 investigations (Thomas *et al.*, 2008) showed (i) Gimenez-positive rods identified as a *Legionella*-Like  
74 Amoebal Pathogen (LLAP-2) and (ii) very small intracellular Gimenez-positive cocci (Fig. 1A). When  
75 re-infecting fresh amoebae with 0.45- $\mu\text{m}$  filtered supernatant, amoebal cells detached from the bottom  
76 of the flask within 1-2 hours and lysed within 12-16 hours. Eight hours post-infection, amoebae were  
77 filled with typical icosahedral viral particles with an average diameter of 190 to 220 nm (Fig. 1B-C)  
78 mainly distributed within large cytoplasmic vacuoles. No fibrils or tails were observed around the viral  
79 capsid, even when negatively stained (Fig. 1D). In a few amoebae, a large granular area containing  
80 maturing virus particles was observed (Fig. S1), corresponding to the viral factory described for other  
81 NCLDV viruses (Novoa *et al.*, 2005; Boyer *et al.*, 2009). In the periphery of this granular area, most  
82 viral particles were mature (electron-dense).

83 The developmental cycle in amoebae was studied by confocal microscopy with specific polyclonal  
84 antibodies (Fig. 1E). Thirty minutes after infection, a few viral particles were observed in the cytoplasm  
85 of the amoebae ( $\text{MOI} \leq 1$ ). Viral particles were no longer observed within amoebae two hours post-  
86 infection, a time falling in the interval of the eclipse phase. After 4 hours, large vesicles filled with  
87 viruses, as well as isolated viral particles, were present within amoebal cells. After 8 hours, all  
88 amoebae were filled with viral particles. Sixteen hours post-infection, the amoebae were lysed and  
89 clusters of released viruses could be detected.

90

### 91 ***Host range***

92 Cytopathic effects, viral multiplication and rapid cell lysis were observed only when various  
93 *Acanthamoeba* spp. that belong to the genotype T4 were infected with the virus. Interestingly, a  
94 genotype T5 field isolate, identified as *Acanthamoeba lenticulata* due to its typical intron in the 18S  
95 rRNA gene (Coulon *et al.*, 2010), was found to be resistant to infection, suggesting a very narrow host  
96 range. Similarly, no cytopathic effects were observed following infection of various other amoebal  
97 species, including *Hartmannella vermiformis*, *Dictyostelium discoideum* and different *Naegleria*  
98 species. Moreover, none of the various human and animal cell lines tested (see Experimental  
99 Procedures) displayed cytopathic effects after inoculation of the virus. The absence of replication in

100 human cells tested is however not sufficient to conclude that Lausannevirus lacks potential  
101 pathogenicity towards mammals. Indeed, Mimivirus has been demonstrated to penetrate within human  
102 and mice myeloid cells (Ghigo *et al.*, 2008) and to induce pneumonia in human (Raoult *et al.*, 2006)  
103 and mice (Khan *et al.*, 2007) despite having a very limited *in vitro* host cell range (Suzan-Monti *et al.*,  
104 2006).

105

#### 106 ***Lausannevirus, a close neighbour of Marseillevirus***

107 Lausannevirus exhibits a genome of 346'754 bp with a G+C content of 42.9% (Table S1) with two  
108 possible conformations: a linear molecule with terminal repeats or a circularized molecule (published  
109 in GenBank, minimal repeat). A total of 450 genes with an average length of 716 bp and covering  
110 92.6% of the genome are predicted to encode proteins ranging from 44 to 1'526 amino acids. Coding  
111 sequences (CDS) are almost equally distributed on both strands (47.1% and 52.9% on the positive  
112 and negative strand respectively). Gene orientation skew and cumulative GC skew indicate a possible  
113 origin of replication around position 195'000 bp (Fig. S2).

114 Out of the 450 CDS, 332 (73.8%) showed significant sequence similarities to proteins of the non  
115 redundant database. Among these proteins, 320 (71.1%) have their closest homolog in the  
116 Marseillevirus genome, highlighting the close relationship between these two viruses. Among the  
117 remaining 12 CDS with a best hit against other organisms, three were probably missed during gene  
118 prediction in Marseillevirus (Table 1). Seven other proteins present homologs in Marseillevirus and  
119 encode some poorly characterized proteins, a nearly complete dUTPase most similar to the intra-  
120 amoebal symbiont *Candidatus Amoebohilus asiaticus* (see below) and an ubiquitin most similar to  
121 *Acanthamoeba castellanii*. Finally, two genes absent in Marseillevirus encode an ankyrin repeat  
122 protein with some similarity to *Hydra magnipapillata* and a hypothetical protein found in a bacterial  
123 phage. Conversely, 280 proteins of Marseillevirus exhibit best hits in Lausannevirus, 56 best hits with  
124 other Marseillevirus proteins (protein families) and 80 are ORFans without significant hit. The  
125 remaining 12 CDS present a best hit with other organisms (Table 1). Marseillevirus genes not present  
126 in Lausannevirus encode a Dam-like adenine-specific DNA methylase, a HNH-family endonuclease, a  
127 translation initiation factor SUI1, a P-loop ATPase/GTPase, a zinc finger protein and a hypothetical  
128 protein (Table 1).

129 Several families of proteins were identified in Lausannevirus (Table S2), the largest families encoding  
130 (i) MORN-repeats containing proteins, (ii) different endonucleases and (iii) serine/threonine protein  
131 kinases. Twelve additional small families are almost exclusively constituted of proteins of unknown  
132 function, with the exception of ubiquitin. The major families of proteins identified are commonly found  
133 in both viruses, including F-box proteins (data not shown).

134 By TBLASTN, 401 proteins of Lausannevirus present a hit in Marseillevirus genome sequence, thus  
135 rising to 89% the percentage of homologous ORFs. This value is in the range found between closely  
136 related NCLDV viruses: 79% to 98% for *Chlorella* viruses (Fitzgerald *et al.*, 2007b; Fitzgerald *et al.*,  
137 2007a), 87.5% for *Ostreococcus* viruses (Derelle *et al.*, 2008; Weynberg *et al.*, 2009) and 99% for  
138 *Mimiviridae* (La Scola *et al.*, 2008).

139 NCLDV possess a set of core genes inherited from the last common ancestor of extant NCLDV (Iyer  
140 *et al.*, 2001). These genes are divided into four groups from the most to the least evolutionarily  
141 conserved, with group 1 core genes being the most conserved. All NCLDV core genes retrieved in  
142 Marseillevirus (Boyer *et al.*, 2009) present highly similar orthologs in Lausannevirus, including a  
143 thymidine kinase present in both viruses. Phylogenetic analyses using five concatenated group I core  
144 protein sequences clearly affiliated Lausannevirus as a relative of Marseillevirus (Fig. 2), suggesting  
145 that Lausannevirus and Marseillevirus belong to the same viral family that we propose to name  
146 *Marseilleviridae*.

147 Strikingly, the first 150 kbp of Lausannevirus genome present only poor colinearity, i.e. conservation of  
148 gene order and orientation, with Marseillevirus as shown in the green circle on Fig. 3A by a dot-plot  
149 representation of the genomic position of orthologous proteins. On the contrary, the following 200 kbp  
150 exhibit higher colinearity, with only five main inverted regions detected. The first inverted region  
151 appears to be formed by a simple segmental inversion reversing gene order and orientation, whereas  
152 others show more complex patterns. NCLDV core genes from groups 1, 2 and 3 are found in both  
153 parts of the genome although the colinear area is slightly enriched in such genes (Fig. 3B). The  
154 distribution of annotated genes is not random, the non-colinear fraction of the genome being enriched  
155 in hypothetical proteins (Fig. 3C). When excluding Marseillevirus from BLAST analyses, the taxonomic  
156 affiliation of best hits do not show a particular enrichment in bacterial, eukaryotic or phagic genes in  
157 the non-colinear part (Fig 3D). Interestingly, the percentage identity to the best BLAST hit presents  
158 only few outliers that are listed here in decreasing order: ubiquitin (97%, Eukaryote), deoxyuridine

159 triphosphate nucleotidohydrolase (67%, Bacteria), a hypothetical protein (58%, Phage), the eukaryotic  
160 peptide chain release factor 1 (57%, Eukaryote) and the ribonuclease H (53%, Bacteria).

161

### 162 ***Viral histones***

163 Histones presumably acquired from hosts are reported in several viruses (Fig. 4): H3-H4 protein in  
164 *Heliothis zea* virus 1, H4 protein in bracoviruses and H2B protein of the ostreid herpesvirus integrated  
165 into the amphioxus genome (Cheng *et al.*, 2002; Gad and Kim, 2008; de Souza *et al.*, 2010). In  
166 bracoviruses the H4 protein has been demonstrated to play a critical role in suppressing host immune  
167 responses during parasitism (Gad and Kim, 2008).

168 Lausannevirus and Marseillevirus both encode three histone-like proteins. LAU\_0051 and  
169 MAR\_ORF166 contain a C-terminal H2A-like histone fold and an unknown N-terminal domain (Fig. 4 A  
170 and B). Surprisingly, the two other proteins form histone doublets, i.e. pairs of fused histones, with C-  
171 terminal and N-terminal extremities that are related to different histones: LAU\_0386 and  
172 MAR\_ORF414 contain a N-terminal H2B-like and a C-terminal H2A-like histone, whereas LAU\_0387  
173 and MAR\_ORF413 contain a C-terminal H3-like histone and a deeply-branching N-terminal part  
174 related to histone H4 or archaeal histones (Fig. 4, Fig. S3 and S4). Viral histones may be functional  
175 and able to bind DNA; they may serve to interact with the host cell DNA or regulate the viral DNA  
176 itself.

177 In addition to Marseilleviridae, two superkingdoms harbor histone proteins with different organization  
178 (Fig. 4C). Eukaryotic organisms harbour one to several copies of the four histones H2A, H2B, H3 and  
179 H4 that associate to form the nucleosome and wrap the DNA. If the canonical histones are often found  
180 in a repeated cluster, other variants are often found isolated in the genome (Talbert and Henikoff,  
181 2010). Moreover, histones were identified in all archaeal phyla including the deepest branching  
182 phylum *Thaumarchaeota* (Cubonova *et al.*, 2005; Sandman and Reeve, 2006; Spang *et al.*, 2010).  
183 The various arrangements and similarity patterns between these histone-encoding genes render  
184 difficult to build a parsimonious model for the presence of these histone doublets in Marseilleviridae  
185 through gene exchange, fusion or splitting. A potential gene acquisition by an ancestral  
186 Marseilleviridae from an unknown eukaryote harbouring fused histone doublets cannot be ruled out,  
187 although such an organism has not been discovered. Another possibility would imply a forced  
188 convergence to adapt and manipulate the host cell. These unusual histone doublets challenge our



189 view on the existing diversity of histone encoding genes and detailed identification of structural  
190 variation might give some additional clues on their evolution.

191

### 192 ***A degenerating core protein acquired by horizontal gene transfer***

193 Blast results for the deoxyuridine triphosphate nucleotidohydrolase (dUTPase), a group 2 core protein  
194 found in most NCLDV (Iyer *et al.*, 2006), raised our interest for this gene given its significant role in  
195 hydrolyzing dUTP, generating dUMP for biosynthesis of thymidine nucleotides while decreasing the  
196 availability of dUTP for misincorporation and its association with the ability of viruses to replicate in  
197 non-dividing cells (Chen *et al.*, 2002). Lausannevirus encodes a nearly complete dUTPase containing  
198 all residues from major active sites previously described except Motif1 that is truncated but still retains  
199 the very conserved aspartic acid residue (Fig. 5). Conversely, Marseillevirus encodes a truncated  
200 dUTPase (Boyer *et al.*, 2009) in which four of the five conserved motifs are deleted (Fig. 5),  
201 suggesting that the latter protein is inactive and evolves toward degradation. Both viruses encode for  
202 an uracil DNA glycosylase, the other enzyme of critical importance to control the amount of uracil in  
203 DNA (De Silva and Moss, 2008). Whether Lausannevirus dUTPase is really active and results in  
204 metabolic and/or virulence differences with Marseillevirus such as in vaccinia virus (De Silva and  
205 Moss, 2008) remains to be determined. Although the short length of dUTPase does not provide  
206 sufficient signal to infer a robust phylogeny at species-level, phylogenetic analyses support the notion  
207 that Lausannevirus dUTPase is clustered with bacterial orthologs and are distantly related to NCLDV  
208 and eukaryotic dUTPases (Fig. 5). dUTPases have been reported in various viral species and *dut*  
209 gene transfers among bacteria as well as between bacteria and viruses have already been  
210 documented (Baldo and McClure, 1999).

211

### 212 ***Entire and degenerated inteins***

213 Inteins are mobile genetic elements co-translated with their host protein that can excise themselves by  
214 protein splicing (Gogarten *et al.*, 2002) and represent interesting viral features since viruses were  
215 suggested to be implicated in transmitting inteins across species (Petrokovski, 1998). In  
216 Lausannevirus and Marseillevirus, two homologous inteins found in the D6/D11-like helicases present  
217 highest similarity with an intein found in the SNF2/Rad54 helicase of the cyanobacteria  
218 *Trichodesmium erythraeum*. Similarly, inteins in the ribonucleotide reductase large subunits

219 MAR\_ORF211 and LAU\_0211 present highest similarities with RIR1 inteins of Iridoviruses and  
220 *Chlorella* virus NY2A (Phycodnavirus), and with RIR1 and RIR3 inteins of the cyanobacterial species  
221 *Crocospaera watsonii* and *Trichodesmium erythraeum*. Important differences were found between  
222 ribonucleotide reductase inteins of both amoebal viruses. Similarly to inteins from *Wiseana* and  
223 *Costelytra zealandica* iridescent viruses, Marseillevirus intein lacks blocks C, E and H, whereas  
224 Lausannevirus intein only lacks block C (Fig. S5). The homing cycle of inteins predicts that the  
225 endonuclease domain (blocks C, D, E and H) would be lost before the splicing domain (blocks A, B, F,  
226 G) (Gogarten *et al.*, 2002), which is the case for both amoebal viruses. Loss of intein has been  
227 reported between closely related viruses such as inteins present in the ribonucleotide reductase and  
228 superfamily II helicase of *Paramecium bursaria* *Chlorella* virus NY-2A but absent from PBCV-1 strain.  
229 Conversely, only minor variations were reported between inteins encoded in the DNA polymerases of  
230 algal viruses with distinct intraspecies host specificities (Nagasaki *et al.*, 2005). In Marseilleviridae, the  
231 two inteins evolved differently toward degradation of the endonuclease domain: Marseillevirus has a  
232 minimal protein splicing element (Telenti *et al.*, 1997) and Lausannevirus appears to have initiated a  
233 progressive loss of the core endonuclease domains or blocks.

234

### 235 **Concluding remarks**

236 The identification of Lausannevirus and the sequencing of its genome revealed new features on the  
237 evolution of these unusual amoebae-infecting viruses of the Marseilleviridae family. There are only  
238 four fully sequenced genomes of viruses infecting amoebae reported to date. These four genomes  
239 group in two separate sets of closely related species, with Mimivirus-Mamavirus on one side and  
240 Marseillevirus-Lausannevirus on the other side. This apparent lack of diversity might be due to the  
241 limited number of amoebal species used to date to isolate these viruses from the environment. A  
242 recent study reports the retrieval of 19 new viruses isolated from the environment using *A. polyphaga*  
243 Linc-AP1 (genotype T4) as a cell background (La Scola *et al.*, 2010). Preliminary analyses based on  
244 partial DNA-polymerase B sequences of these 19 viruses suggest that most of them cluster in the  
245 same group as Mimivirus, except one potentially representing a new viral family (La Scola *et al.*,  
246 2010). Full genome sequence analysis of these new viruses will bring invaluable information on the  
247 biology of these amoebae-resisting viruses. The possible recovery of additional large DNA viruses  
248 from other amoebal species such as *Hartmannella* and *Naegleria* spp. will likely provide a greater

249 insight into virus evolution. New giant viruses will also likely be isolated from other ecological niches,  
250 since virus-like particles have been reported in a variety of hosts including dinoflagellates (Tarutani *et*  
251 *al.*, 2001), *Giardia muris* from hamsters, *Giardia duodenalis* from domestic rats and *Blastocystis* sp.  
252 from simian faecal material (Sogayar and Gregorio, 1986; Stenzel and Boreham, 1997; Sogayar and  
253 Gregório, 1998). The description of large viruses in eukaryotic species infecting mammals could  
254 bring additional perspectives in terms of host range and, more importantly, evolution and gene  
255 exchange between viruses and perhaps with their eukaryotic hosts. The role of viruses, giant or not, in  
256 horizontal gene transfer should not be underestimated (Thomas and Greub, 2010) and the analysis of  
257 new genomes of amoebae-infecting viruses will provide insight in the evolutionary history of viruses  
258 and their interactions with eukaryotes.

## 259 **Experimental Procedures**

### 260 **Amoebal co-culture**

261 To recover amoebae-resisting bacteria potentially present in Seine River, 1 L water samples were  
262 collected at the entry of the Morsang-sur-Seine (France) drinking water plant. Samples were obtained  
263 in January, May, July and October 2005, and investigated for the presence of amoebae-resisting  
264 micro-organisms using amoebal co-culture onto axenic *Acanthamoeba castellanii* amoeba strain  
265 ATCC 30010 as described in Thomas *et al.* (Thomas *et al.*, 2008).

266

### 267 **Electron microscopy**

268 Sixteen hours post-infection, co-cultures of Gimenez-positive cocci with *A. castellanii* ATCC 30010  
269 were harvested, washed in PAS buffer, and fixed in 4% glutaraldehyde (Fluka Biochemika, Buchs,  
270 Switzerland) for 3h at 4°C. Fixed samples were then washed in PBS and fixed for 1h at room  
271 temperature with 1% osmium tetroxide in PBS. Dehydration was performed by successive washes in  
272 increasing acetone concentrations (50 to 100%). Samples were incubated for 1h in a vol/vol  
273 suspension of acetone-epon and overnight in epon. They were then embedded in an epoxy resin  
274 (Fluka). Thin sections were cut from embedded blocks by a LKB 2088 ultratome, deposited on copper  
275 grids coated with formvar (Sigma-Aldrich, Taufkirchen, Germany) and stained for 10 min with a  
276 solution of methanol-uranyl acetate and lead nitrate with sodium citrate in water. Grids were examined  
277 with a transmission electron microscope (PHILIPS EM 201 C, Philips, Eindhoven, Netherlands).

278

### 279 **Virus purification and DNA extraction**

280 A one-week old *A. castellanii* infected flask was 5 µm filtered and 1 ml aliquots of the filtrate from this  
281 suspension were used to re-infect new flasks. After 5 days of incubation at 32°C, the infected flask  
282 supernatants were harvested and centrifuged at 5,000 X g for 15 min to pellet amoebal cells. The  
283 supernatant was 5 µm-filtered to remove residual cells. The suspension was then centrifuged at  
284 35,000 X g for 1h and the pelleted virus resuspended in 1 ml PBS (for antibody production) or DNA-  
285 free water (for DNA sequencing, see below). Genomic DNA was isolated from purified virus particles  
286 with the Wizard Genomic DNA purification kit (Promega Corporation, Madison, USA).

287

288 **Antibody production, immunofluorescence and study of the virus life cycle.**

289 The virus was inactivated at 70 °C for 1h and used for vaccination of two 6-week old Balb-C mice. Two  
 290 weeks after a first intraperitoneal injection of a mixture containing 375 µl of purified, inactivated virus  
 291 and 375 µl of complete Freund's adjuvant (Sigma, Steinheim, Germany), 750 µl purified virus without  
 292 Freund's adjuvant was administered intraperitoneally. One month after the first injection, blood was  
 293 collected, centrifuged at 2,754 X g for 10 min, and the serum was stored at -20°C. The specific  
 294 polyclonal antibodies were then used to detect the virus within amoebal cells. Fifty µl supernatant of a  
 295 one-week old *A. castellanii* infected flask was inoculated in wells of plates seeded with 10<sup>5</sup> *A.*  
 296 *castellanii* cells in 1 ml of PAS. Thirty minutes, 4, 8, 16 and 24h post-infection, cells were fixed for 10  
 297 min with 3% paraformaldehyde. After 2 washes with PBS-0.1% saponin, cover slips were incubated  
 298 for 1h at room temperature with immunoglobulins diluted 1/800 in PBS-0.1% saponin and 10% FCS.  
 299 They were then washed in PBS-0.1% saponin, and incubated with FITC-coupled anti-mouse  
 300 immunoglobulins antibodies diluted 1/200 (BioRad, Reinach, Switzerland) and 1% Evans blue  
 301 (BioMerieux, Marcy-l'Etoile, France) for 1 h at room temperature. Finally, cover slips were washed  
 302 consecutively in PBS-0.1% saponin, PBS and ultrapure water, mounted using Mowiol, and examined  
 303 with a Zeiss LSM 510 confocal microscope.

304

305 **Host range**

306 A one-week old infected flask of *A. castellanii* was scraped; the supernatant was 5µm-filtered and 50µl  
 307 were used to infect *Hartmannella vermiformis* ATCC 50237, *Acanthamoeba polyphaga* Linc-AP1,  
 308 *Dictyostelium discoideum* DH1-10, various *Acanthamoeba* isolates that belong to the T4 and T5  
 309 groups (Thomas *et al.*, 2008; Coulon *et al.*, 2010) and various *Naegleria* isolates recovered from the  
 310 same environmental samples than the virus (Thomas *et al.*, 2008). In addition to human PBMC-  
 311 derived macrophages prepared as in Croxatto *et al.* (Croxatto and Greub, 2010), human (Vero ATCC  
 312 CCL-81, HEL ATCC CCL-137, HEp-2 ATCC CCL-23 and A549 ATCC CCL-185) and animal (BGM  
 313 ECACC 90092601, BT ATCC CRL-1390, MA-104 ATCC CRL-2378, MDCK ATCC CCL-34, MDBK  
 314 ATCC CCL-22, A-72 ATCC CRL-1542, EBL DSMZ ACC192, Bu ECACC 89051704, and Mv1-Lu  
 315 ATCC CCL-64) cell lines were also tested, as well as the insect cell line ATCC CRL-1660 from *Aedes*  
 316 *albopictus*.

317

### 318 **Genome sequencing, assembly and gap closure**

319 A 1.5kb and 5kb library was constructed and sequenced by Sanger method, yielding 6,495 paired  
320 reads for an average 15X coverage. Sanger sequences were assembled in 29 large contigs using  
321 Phrap. Primers were designed with Consed (Gordon *et al.*, 1998) and PCR-based techniques were  
322 used to scaffold the contigs followed by Sanger sequencing of the products to fill the gaps. DNA was  
323 also sequenced using Genome Sequencer FLX<sup>TM</sup> (Droege and Hill, 2008). GS FLX reads were  
324 mapped to the final assembly with Phrap.

### 325 **Genome annotation**

326 Coding sequence (CDS) prediction and annotation was performed using the genome annotation  
327 system GenDB 2.4 (Meyer *et al.*, 2003). CDS were predicted with the softwares Critica (Badger and  
328 Olsen, 1999), Glimmer (Delcher *et al.*, 2007) and Reganor (Linke *et al.*, 2006) and automatically  
329 submitted to similarity searches against nr (non-redundant database), Swissprot, TIGRfam, Pfam,  
330 InterPro and KEGG databases. BLAST searches of the CDS were also performed *versus*  
331 representative genomes of several NCLDV families (Marseillevirus, *A. polyphaga* Mimivirus, *E. huxleyi*  
332 virus 86, *P. bursaria* Chlorella virus NY2A). Putative signal peptides, transmembrane helices and  
333 nucleic acid binding domains were respectively predicted using SignalP (Bendtsen *et al.*, 2004),  
334 TMHMM (Krogh *et al.*, 2001) and Helix-Turn-Helix (Dodd and Egan, 1987). In case of completely or  
335 largely overlapping genes, preference was given to the genes showing a homolog in Marseillevirus  
336 genome. For each CDS, the automatic annotation was manually checked and corrected according to  
337 the most congruent tool results.

### 338 **Genome sequence availability**

339 The complete annotated genome sequence is available at NCBI under the accession number  
340 HQ113105.

### 341 **Genome analysis**

342 Lausannevirus and Marseillevirus translated protein sequences were compared against the NCBI nr  
343 database (July 2010) using BLAST searches (blastp, tblastn) with an e-value cutoff of 0.01. Orthologs  
344 between the two genomes were identified by reciprocal best blast hit and their respective positions on  
345 the genomes were used to draw the dot-plot in Figure 3. Best hits against other organisms were

346 retrieved by excluding Marseillevirus hits and taxonomic position was retrieved according to NCBI  
347 taxonomic tables. Protein families of Lausannevirus were identified by similarity-based clustering with  
348 BLASTCLUST (Altschul *et al.*, 1990) using parameters of coverage (L=40), sequence conservation  
349 (s=30) and e-value (0.01). Families presenting a similar annotation were not merged as this  
350 sometimes reflected the presence of truncated proteins of a larger family (e.g. clusters 3 and 24).  
351 Inteins were detected and analyzed by BLAST against InBase (Perler, 2002)  
352 (<http://www.neb.com/neb/inteins.html>). Intein sequence motifs were identified and annotated according  
353 to Perler *et al.* (Perler *et al.*, 1997)

354

### 355 **Phylogenetic analyses**

356 All multiple sequence alignments were performed with Muscle V3.8.31 (Edgar, 2004) and phylogenetic  
357 trees were built with PhyML (Guindon and Gascuel, 2003) (parameters: LG model, estimated gamma  
358 parameter, best of SPR and NNI, 100 bootstraps) on the following datasets: Five core I proteins of  
359 selected viruses representative of NCLDV families were aligned and then concatenated to construct a  
360 tree. Similarly, proteins exhibiting best hits with Lausannevirus dUTPase by BLASTP vs nr were  
361 retrieved, together with dUTPase from various representative NCLDV viruses, phages, eukaryotes and  
362 other prokaryotes.

363 Histone-like proteins of *Archaea* and all histones available from a large panel of species covering the  
364 different eukaryotic phyla (Hampel *et al.*, 2009) (*Homo sapiens*, *Drosophila melanogaster*, *Danio rerio*,  
365 *Schizosaccharomyces pombe*, *Aspergillus niger*, *Candida albicans* [Opisthokonta], *Mytilus edulis*,  
366 *Encephalitozoon cuniculi*, *Dictyostelium discoideum*, *Entamoeba histolytica* [Amoebozoa],  
367 *Trypanosoma cruzi*, *Leishmania major* [Excavata], *Arabidopsis thaliana*, *Oryza sativa* [Archaeplastida],  
368 *Toxoplasma gondii*, *Cryptosporidium parvum*, *Plasmodium falciparum* [Chromalveolates]) were  
369 retrieved from NCBI protein database or from the genome project website for *Acanthamoeba*  
370 *castellanii* Neff (<ftp://ftp.hgsc.bcm.tmc.edu/pub/data/AcastellaniiNeff/>) and the EST database for  
371 *Nuclearia simplex* (<http://amoebidia.bcm.umontreal.ca/pepdb/searches/welcome.php>). These  
372 sequences were pre-aligned by histone type before being combined using Muscle –profile option. Viral  
373 genes exhibiting histone folds, as recognized by Pfam, were splitted to keep each predicted histone  
374 fold in one unique protein sequence. These amino acid sequences were then singularly or iteratively  
375 aligned to the pre-computed alignment of histones. To reduce the complexity of the phylogenetic tree

376 presented, one representative sequence was chosen for closely-related branches that were appearing  
377 far from the viral histones in an iterative process. Phylogenetic reconstructions were performed with  
378 the maximum likelihood, Bayesian and neighbor-joining algorithms using respectively PhyML  
379 (parameters: LG model, estimated gamma parameter, best of SPR and NNI, 100 bootstraps),  
380 MrBayes (Ronquist and Huelsenbeck, 2003) (1'500'000 generations) and Mega V4 (Tamura *et al.*,  
381 2007) (Neighbor Joining, pairwise deletion, poisson distribution, gamma parameter=1).  
382



383 **Acknowledgments**

384 We warmly thank Dr. Roland Sahli for helpful discussions. This work was partially supported by the  
385 Infectigen foundation (project In010), by the Institute of Microbiology of the University of Lausanne and  
386 by the ESCMID Young Investigator Award (G. Greub, 2006). We thank J.L. Barblan (Department of  
387 Fundamental Microbiology, Lausanne, Switzerland) and the PFMU at the Medical Faculty of Geneva  
388 for assisting with electron microscopy analyses. We also thank S. Aeby and B. Linke for their excellent  
389 technical assistance. G. Greub is supported by the Leenards Foundation through a career award  
390 entitled 'Bourse Leenards pour la relève académique en médecine clinique à Lausanne'.

391 **References**

392

393 Allen, M.J., Schroeder, D.C., Holden, M.T., and Wilson, W.H. (2006) Evolutionary history of the  
394 *Coccolithoviridae*. *Mol Biol Evol* **23**: 86-92.

395  
396 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search  
397 tool. *J Mol Biol* **215**: 403-410.

398  
399 Badger, J.H., and Olsen, G.J. (1999) CRITICA: coding region identification tool invoking comparative  
400 analysis. *Mol Biol Evol* **16**: 512-524.

401  
402 Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal  
403 peptides: SignalP 3.0. *J Mol Biol* **340**: 783-795.

404  
405 Baldo, A.M., and McClure, M.A. (1999) Evolution and horizontal transfer of dUTPase-encoding genes  
406 in viruses and their hosts. *J Virol* **73**: 7710-7721.

407  
408 Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L. et al. (2009) Giant  
409 Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric  
410 microorganisms. *Proc Natl Acad Sci U S A* **106**: 21848-21853.

411  
412 Chen, R., Wang, H., and Mansky, L.M. (2002) Roles of uracil-DNA glycosylase and dUTPase in virus  
413 replication. *J Gen Virol* **83**: 2339-2345.

414  
415 Cheng, C.H., Liu, S.M., Chow, T.Y., Hsiao, Y.Y., Wang, D.P., Huang, J.J., and Chen, H.H. (2002)  
416 Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of  
417 the Baculoviridae. *J Virol* **76**: 9024-9034.

418  
419 Coulon, C., Collignon, A., McDonnell, G., and Thomas, V. (2010) Resistance of *Acanthamoeba* cysts  
420 to disinfection treatments used in health care settings. *J Clin Microbiol* **48**: 2689-2697.

421  
422 Croxatto, A., and Greub, G. (2010) Early intracellular trafficking of *Waddlia chondrophila* in human  
423 macrophages. *Microbiology* **156**: 340-355.

424  
425 Cubonova, L., Sandman, K., Hallam, S.J., DeLong, E.F., and Reeve, J.N. (2005) Histones in  
426 crenarchaea. *J Bacteriol* **187**: 5482-5485.

427  
428 De Silva, F.S., and Moss, B. (2008) Effects of vaccinia virus uracil DNA glycosylase catalytic site and  
429 deoxyuridine triphosphatase deletion mutations individually and together on replication in active and  
430 quiescent cells and pathogenesis in mice. *Virology* **5**: 145.

431  
432 de Souza, R.F., Iyer, L.M., and Aravind, L. (2010) Diversity and evolution of chromatin proteins  
433 encoded by DNA viruses. *Biochim Biophys Acta* **1799**: 302-318.

434  
435 Delcher, A.L., Bratke, K.A., Powers, E.C., and Salzberg, S.L. (2007) Identifying bacterial genes and  
436 endosymbiont DNA with Glimmer. *Bioinformatics* **23**: 673-679.

437  
438 Derelle, E., Ferraz, C., Escande, M.L., Eychenie, S., Cooke, R., Piganeau, G. et al. (2008) Life-cycle  
439 and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus*  
440 *tauri*. *PLoS One* **3**: e2250.

441  
442 Dodd, I.B., and Egan, J.B. (1987) Systematic method for the detection of potential lambda Cro-like  
443 DNA-binding regions in proteins. *J Mol Biol* **194**: 557-564.

444  
445 Droege, M., and Hill, B. (2008) The Genome Sequencer FLX System--longer reads, more applications,  
446 straight forward bioinformatics and more complete data sets. *J Biotechnol* **136**: 3-10.

447

- 448 Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput.  
449 *Nucleic Acids Res* **32**: 1792-1797.  
450
- 451 Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Nierman, W.C., and Van Etten, J.L. (2007a)  
452 Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella*  
453 NC64A. *Virology* **358**: 472-484.  
454
- 455 Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Hartigan, J., and Van Etten, J.L. (2007b)  
456 Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella*  
457 Pbi. *Virology* **358**: 459-471.  
458
- 459 Gad, W., and Kim, Y. (2008) A viral histone H4 encoded by *Cotesia plutellae* bracovirus inhibits  
460 haemocyte-spreading behaviour of the diamondback moth, *Plutella xylostella*. *J Gen Virol* **89**: 931-  
461 938.  
462
- 463 Ghigo, E., Kartenbeck, J., Lien, P., Pelkmans, L., Capo, C., Mege, J.L., and Raoult, D. (2008)  
464 Ameobal pathogen mimivirus infects macrophages through phagocytosis. *PLoS Pathog* **4**: e1000087.  
465
- 466 Gogarten, J.P., Senejani, A.G., Zhaxybayeva, O., Olendzenski, L., and Hilario, E. (2002) Inteins:  
467 structure, function, and evolution. *Annu Rev Microbiol* **56**: 263-287.  
468
- 469 Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing.  
470 *Genome Res* **8**: 195-202.  
471
- 472 Greub, G., and Raoult, D. (2004) Microorganisms resistant to free-living amoebae. *Clin Microbiol Rev*  
473 **17**: 413-433.  
474
- 475 Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large  
476 phylogenies by maximum likelihood. *Syst Biol* **52**: 696-704.  
477
- 478 Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G., and Roger, A.J. (2009)  
479 Phylogenomic analyses support the monophyly of Excavata and resolve relationships among  
480 eukaryotic "supergroups". *Proc Natl Acad Sci U S A* **106**: 3859-3864.  
481
- 482 Iyer, L.M., Aravind, L., and Koonin, E.V. (2001) Common origin of four diverse families of large  
483 eukaryotic DNA viruses. *J Virol* **75**: 11720-11734.  
484
- 485 Iyer, L.M., Balaji, S., Koonin, E.V., and Aravind, L. (2006) Evolutionary genomics of nucleo-  
486 cytoplasmic large DNA viruses. *Virus Res* **117**: 156-184.  
487
- 488 Khan, M., La Scola, B., Lepidi, H., and Raoult, D. (2007) Pneumonia in mice inoculated experimentally  
489 with *Acanthamoeba polyphaga* mimivirus. *Microb Pathog* **42**: 56-61.  
490
- 491 Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane  
492 protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-  
493 580.  
494
- 495 La Scola, B., Campocasso, A., N'Dong, R., Fournous, G., Barrassi, L., Flaudrops, C., and Raoult, D.  
496 (2010) Tentative characterization of new environmental giant viruses by MALDI-TOF mass  
497 spectrometry. *Intervirology* **53**: 344-353.  
498
- 499 La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M. et al. (2003) A giant  
500 virus in amoebae. *Science* **299**: 2033.  
501
- 502 La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G. et al. (2008) The  
503 virophage as a unique parasite of the giant mimivirus. *Nature* **455**: 100-104.  
504
- 505 Lamoth, F., and Greub, G. (2010) Amoebal pathogens as emerging causal agents of pneumonia.  
506 *FEMS Microbiol Rev* **34**: 260-280.  
507

- 508 Linke, B., McHardy, A.C., Neuweger, H., Krause, L., and Meyer, F. (2006) REGANOR: a gene  
 509 prediction server for prokaryotic genomes and a database of high quality gene predictions for  
 510 prokaryotes. *Appl Bioinformatics* **5**: 193-198.  
 511
- 512 Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J. et al. (2003) GenDB--an  
 513 open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31**: 2187-2195.  
 514
- 515 Moliner, C., Fournier, P.E., and Raoult, D. (2010) Genome analysis of microorganisms living in  
 516 amoebae reveals a melting pot of evolution. *FEMS Microbiol Rev* **34**: 281-294.  
 517
- 518 Moreira, D., and Brochier-Armanet, C. (2008) Giant viruses, giant chimeras: the multiple evolutionary  
 519 histories of Mimivirus genes. *BMC Evol Biol* **8**: 12.  
 520
- 521 Nagasaki, K., Shirai, Y., Tomaru, Y., Nishida, K., and Pietrokovski, S. (2005) Algal viruses with distinct  
 522 intraspecies host specificities include identical intein elements. *Appl Environ Microbiol* **71**: 3599-3607.  
 523
- 524 Novoa, R.R., Calderita, G., Arranz, R., Fontana, J., Granzow, H., and Risco, C. (2005) Virus factories:  
 525 associations of cell organelles for viral replication and morphogenesis. *Biol Cell* **97**: 147-172.  
 526
- 527 Perler, F.B. (2002) InBase: the Intein Database. *Nucleic Acids Res* **30**: 383-384.  
 528
- 529 Perler, F.B., Olsen, G.J., and Adam, E. (1997) Compilation and analysis of intein sequences. *Nucleic  
 530 Acids Res* **25**: 1087-1093.  
 531
- 532 Pietrokovski, S. (1998) Modular organization of inteins and C-terminal autocatalytic domains. *Protein  
 533 Sci* **7**: 64-71.  
 534
- 535 Raoult, D., Renesto, P., and Brouqui, P. (2006) Laboratory infection of a technician by mimivirus. *Ann  
 536 Intern Med* **144**: 702-703.  
 537
- 538 Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H. et al. (2004) The 1.2-megabase  
 539 genome sequence of Mimivirus. *Science* **306**: 1344-1350.  
 540
- 541 Ronquist, F., and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed  
 542 models. *Bioinformatics* **19**: 1572-1574.  
 543
- 544 Sandman, K., and Reeve, J.N. (2006) Archaeal histones and the origin of the histone fold. *Curr Opin  
 545 Microbiol* **9**: 520-525.  
 546
- 547 Sogayar, M.I., and Gregorio, E.A. (1986) Cytoplasmic inclusions in *Giardia*: an electron microscopy  
 548 study. *Ann Trop Med Parasitol* **80**: 49-52.  
 549
- 550 Sogayar, M.I., and Gregório, E.A. (1998) *Giardia agilis*: Ultrastructure of the Trophozoites in the Frog  
 551 Intestine. In *Mem Inst Oswaldo Cruz*. Rio de Janeiro, Bresil: Instituto Oswaldo Cruz, pp. 357-361.  
 552
- 553 Spang, A., Hatzepichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E. et al. (2010)  
 554 Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum  
 555 *Thaumarchaeota*. *Trends Microbiol* **18**: 331-340.  
 556
- 557 Stenzel, D.J., and Boreham, P.F. (1997) Virus-like particles in *Blastocystis* sp. from simian faecal  
 558 material. *Int J Parasitol* **27**: 345-348.  
 559
- 560 Suzan-Monti, M., La Scola, B., and Raoult, D. (2006) Genomic and evolutionary aspects of Mimivirus.  
 561 *Virus Res* **117**: 145-155.  
 562
- 563 Talbert, P.B., and Henikoff, S. (2010) Histone variants--ancient wrap artists of the epigenome. *Nat Rev  
 564 Mol Cell Biol* **11**: 264-275.  
 565
- 566 Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics  
 567 Analysis (MEGA) software version 4.0. *Mol Biol Evol*.

- 568  
569 Tarutani, K., Nagasaki, K., Itakura, S., and Yamaguchi, M. (2001) Isolation of a virus infecting the  
570 novel shellfish-killing dinoflagellate *Heterocapsa circularisquama*. *Aquat Microb Ecol* **23**: 103-111.  
571
- 572 Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W.R., Jr., and Perler, F.B. (1997) The  
573 *Mycobacterium xenopi* GyrA protein splicing element: characterization of a minimal intein. *J Bacteriol*  
574 **179**: 6378-6382.  
575
- 576 Thomas, V., and Greub, G. (2010) Amoebae/amoebal symbionts genetic transfers: lessons from giant  
577 viruses neighbours. *Intervirology* **53**: 254-267.  
578
- 579 Thomas, V., Loret, J.F., Jousset, M., and Greub, G. (2008) Biodiversity of amoebae and amoebae-  
580 resisting bacteria in a drinking water treatment plant. *Environ Microbiol* **10**: 2728-2745.  
581
- 582 Thomas, V., McDonnell, G., Denyer, S.P., and Maillard, J.-Y. (2010) Free-living amoebae and their  
583 intracellular pathogenic microorganisms: risks for water quality. *FEMS Microbiol Rev* **34**: 231-259.  
584
- 585 Van Etten, J.L. (2003) Unusual life style of giant chlorella viruses. *Annu Rev Genet* **37**: 153-195.  
586
- 587 Vickerman, K. (1962) Patterns of cellular organisation in *Limax* amoebae. An electron microscope  
588 study. *Exp Cell Res* **26**: 497-519.  
589
- 590 Wang, A.L., and Wang, C.C. (1991) Viruses of the protozoa. *Annu Rev Microbiol* **45**: 251-263.  
591
- 592 Weynberg, K.D., Allen, M.J., Ashelford, K., Scanlan, D.J., and Wilson, W.H. (2009) From small hosts  
593 come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environ*  
594 *Microbiol* **11**: 2821-2839.  
595  
596  
597

598  
599

## Table and Figure Legends

Query protein		Best hit protein			BLAST features		Presence of a hit in	
Protein ID	Product annotation	Organism	Taxonomic classification	Protein product	% identity	% coverage	BLASTP	TBLASTN
LAU_0069	Hypothetical protein	<i>Sebaldeella termitidis</i> ATCC 33386	Fusobacteria	MORN variant repeat protein	32.8	73.9	yes	yes
LAU_0073	Hypothetical protein	<i>Burkholderia phymatum</i> STM815	Betaproteobacteria	Hypothetical protein Bphy 5986	32.4	69.1	yes	yes
LAU_0075	Dual specificity protein phosphatase	<i>Aedes aegypti</i>	Insecta	Puckered protein, putative	40.3	82.7	no	yes
LAU_0098	Deoxyuridine 5'-triphosphate nucleotidohydrolase	<i>Candidatus Amoebophilus asiaticus</i> 5a2	Bacteroidetes	Deoxyuridine 5'-triphosphate nucleotidohydrolase Dut	66.9	100	yes	yes
LAU_0198	Hypothetical protein	<i>Helicobacter bilis</i> ATCC 43879	Epsilonproteobacteria	Conserved hypothetical protein	33.9	76.9	yes	yes
LAU_0200	putative ankyrin repeat protein	<i>Hydra magnipapillata</i>	Cnidaria	Similar to predicted protein	25.9	63.3	no	no
LAU_0245	Hypothetical protein	gamma proteobacterium NOR51-B	Gammaproteobacteria	MORN variant repeat protein	32.6	57.9	yes	yes
LAU_0250	Ubiquitin	<i>Acanthamoeba castellanii</i>	Amoebozoa	Polyubiquitin	97.3	100	yes	yes
LAU_0296	Conserved hypothetical protein	uncultured bacterium pFosPlaG	Fusobacteria	Hypothetical exported 24-amino acid repeat protein	33.3	62.3	yes	yes
LAU_0319	Hypothetical protein	<i>Escherichia</i> phage rv5	Bacteriophage	Hypothetical protein rv5 gp094	57.9	61.3	no	no
LAU_0331	Hypothetical protein	<i>Ostreococcus tauri</i> virus 1	Virus	Hypothetical protein OTV1 149	40.3	79.5	no	yes
LAU_0345	DNA polymerase delta catalytic subunit	<i>Aedes taeniorhynchus</i> iridescent virus	Virus	Hypothetical protein MIV120R	33.0	81.7	no	yes
Query protein		Best hit protein			BLAST features		Presence of a hit in	
Protein ID	Product annotation	Organism	Taxonomic classification	Protein product	% identity	% coverage	BLASTP	TBLASTN
MAR ORF015	Truncated dUTPase	<i>Peptoniphilus</i> sp. oral taxon 386 str. F0131	Firmicutes	Deoxyuridine 5'triphosphate nucleotidohydrolase	66.7	80.3	yes	yes
MAR ORF016	Hypothetical protein	<i>Paramecium bursaria Chlorella</i> virus 1	Virus	Hypothetical protein	35.3	71.8	no	no
MAR ORF017	Dam-like adenine-specific DNA methylase	<i>Paramecium bursaria Chlorella</i> virus AR158	Virus	Hypothetical protein AR158 C701R	46.1	86.8	no	no
MAR ORF079	Translation initiation factor SUI1	<i>Medicago truncatula</i>	Viridiplantae	Unknown	40.8	80	no	no
MAR ORF103	Hypothetical protein	<i>Leptotrichia goodfellowii</i> F0264	Fusobacteria	MORN repeat-containing protein	29.8	80.3	yes	yes
MAR ORF159	Hypothetical protein	<i>Helicobacter bilis</i> ATCC 43879	Epsilonproteobacteria	Glucosyltransferase-S	33.6	44.8	yes	yes
MAR ORF181	Putative nuclease	<i>Coprinopsis cinerea</i> okayama7 130	Fungi	Hypothetical protein CC1G 02718	26.8	67.7	yes	yes
MAR ORF247	Ubiquitin	<i>Cercozoa</i> sp. CC-2009e	Rhizaria	Polyubiquitin	89.0	97.3	yes	yes
MAR ORF295	HNH-family endonuclease	<i>Cryptosporidium muris</i> RN66	Alveolata	AP2 domain-containing protein	36.1	15.5	no	no

MAR ORF304	Eukaryotic translation initiation factor 5	<i>Ajellomyces capsulatus</i> H143	Fungi	Eukaryotic translation initiation factor 5	35.2	85.3	yes	yes
MAR ORF326	P-loop ATPase/GTPase	<i>Acanthamoeba polyphaga</i> mimivirus	Virus	Hypothetical protein MIMI L611	37.7	86.7	no	no
MAR ORF391	Multiple zinc ribbon protein	<i>Aedes aegypti</i>	Metazoa	Zinc finger protein	38.5	21.8	no	no

600

601 Table 1. Gene content comparison

602 This table shows Lausannevirus and Marseillevirus proteins exhibiting a best hit *versus* the non-redundant database other than Marseillevirus, respectively  
603 Lausannevirus. The three Lausannevirus proteins showing a hit in Marseillevirus by tblastn only, probably missed by the gene prediction software, are colored  
604 in light grey. Genes unique to one or the other virus, but found in other organisms are highlighted in dark grey whereas those present in both viruses but  
605 exhibiting higher similarity to another organism are shown in white.

606

607 Figure 1. Discovery and characterization of Lausannevirus.

608 A (bar: 10 $\mu$ m): Gimenez staining of the initial co-culture sample demonstrating the presence of  
609 bacterial rods (green arrows, later identified as LLAP-2) and small dots (purple arrows, Lausannevirus)  
610 within the amoebae. B (bar: 2 $\mu$ m): electron microscopy picture of numerous viral particles within the  
611 amoebal cytoplasm. C (bar: 500nm): electron microscopy picture demonstrating the presence of full  
612 (electron-dense, black arrows) and empty (electron-lucent, white arrows) viral particles in the  
613 cytoplasm of infected amoebae. D (bar: 40nm): negative staining electron microscopy picture showing  
614 the presence of several layers delimiting the condensed core. E-I (bar: 10 $\mu$ m): development cycle of  
615 Lausannevirus in *Acanthamoeba castellanii*, assessed using immunofluorescence and confocal  
616 microscopy; viral particles (green) were detected using polyclonal antibodies raised in mice and FITC-  
617 coupled anti-mouse immunoglobulins antibodies, and amoebae were stained with Evans blue (red).  
618 Two hours after infection, during the eclipse phase, Lausannevirus is not detected. Amoebal lysis  
619 occurs 16h post-infection.

620

621 Figure 2. Core gene phylogeny

622 Lausannevirus is closely related to Marseillevirus as shown in this phylogenetic maximum-likelihood  
623 tree of representative NCLDV viruses based on the concatenated alignment of 5 core proteins (A32  
624 ATPase, major capsid protein, D5 helicase, DNA polymerase B and A1L/VLTF2 transcription factor).

625

626 Figure 3. Genome comparison of Lausannevirus and Marseillevirus

627 (A) Dot-plot based on the genomic position of orthologous proteins, as defined by reciprocal best blast  
628 hit between Lausannevirus and Marseillevirus. Genes presenting the same strand orientation are  
629 shown in blue whereas those encoded on opposite strands are shown in red. The first part of the  
630 genome exhibits little colinearity, i.e. same gene order and orientation (green circle), whereas the  
631 second part is largely colinear, showing 5 main inverted segments (brown arrows). (B) Similar  
632 representation highlighting the position of Group I, II and III core genes in pink, blue and green  
633 respectively. (C) Distribution along the genome of hypothetical proteins (grey), conserved hypothetical  
634 proteins (red) and proteins with annotated family or gene function (black). (D) Genome position and  
635 percentage identity of proteins with best blast hit against viruses (cyan "V"), phages (dark green "P"),  
636 bacteria (blue "B") and eukarya (pink "E").



637

638 Figure 4. Description of viral histones.

639 (A) Bayesian tree of eukaryote and viral histones as well as archaeal histone-like proteins.  
640 Lausannevirus and Marseillevirus proteins presenting a predicted histone fold (purple) branch deeply  
641 with the different types of eukaryotic histones. The position of LAU\_0387 N-terminal histone fold is  
642 basal and cannot be confidently inferred, changing according to the model used between Archaea and  
643 H4. On the contrary, histones from other viruses (pink) are found branching very close to classical  
644 eukaryotic histones. (B) Schematic representation of Lausannevirus proteins with their Pfam hits and  
645 the annotation of their best BLAST hit (excluding Marseillevirus) indicated respectively above and  
646 under each domain. (C) Existing arrangement of histone proteins. In Archaea, histones are found as  
647 single gene for example in *Methanosarcina acetivorans*, as fused genes forming doublets e.g. in  
648 *Methanopyrus kandleri* or two different histones e.g. in *Nanoarchaeum equitans*. Marseilleviridae  
649 encode a protein with one H2A-like histone domain and two histone doublets that were colored  
650 according to the annotation of the best blast hit. Eukaryota do have multiple histone encoding genes  
651 for each H2A, H2B, H3 and H4 family that can be found close to each other or in different  
652 chromosomal loci.

653

654 Figure 5. Truncated dUTPase

655 Alignment (A) and maximum-likelihood tree (B) of the dUTPase proteins from selected NCLDV,  
656 phages, representative eukaryotes and bacteria as well as the best nr hits of  
657 Lausannevirus/Marseillevirus dUTPase. Respectively, the N-terminus and C-terminus of  
658 Lausannevirus and Marseillevirus are truncated. Whereas other NCLDV proteins cluster together with  
659 eukaryotic dUTPase, Lausannevirus protein clearly clusters with prokaryotic dUTPase, although its  
660 relative position within bacteria cannot be robustly inferred.

661

662

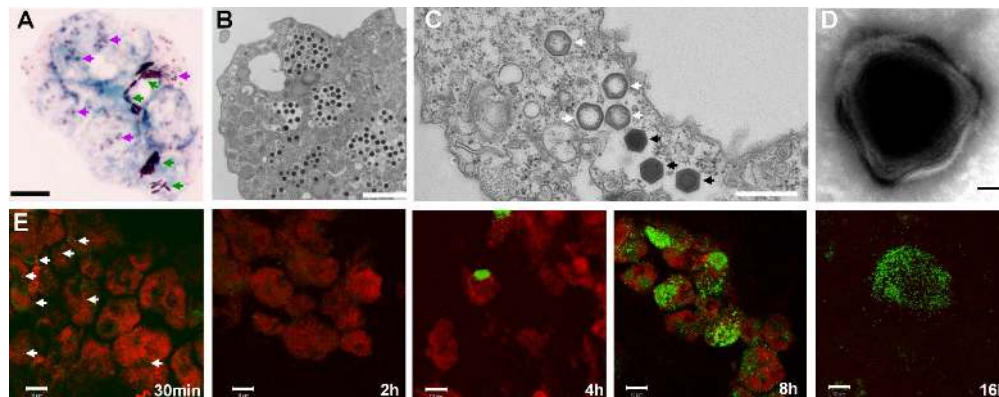


Figure 1. Discovery and characterization of Lausannevirus.

A (bar: 10 $\mu$ m): Gimenez staining of the initial co-culture sample demonstrating the presence of bacterial rods (green arrows, later identified as LLAP-2) and small dots (purple arrows, Lausannevirus) within the amoebae. B (bar: 2 $\mu$ m): electron microscopy picture of numerous viral particles within the amoebal cytoplasm. C (bar: 500nm): electron microscopy picture demonstrating the presence of full (electron-dense, black arrows) and empty (electron-lucent, white arrows) viral particles in the cytoplasm of infected amoebae. D (bar: 40nm): negative staining electron microscopy picture showing the presence of several layers delimiting the condensed core. E-I (bar: 10 $\mu$ m): development cycle of Lausannevirus in *Acanthamoeba castellanii*, assessed using immunofluorescence and confocal microscopy; viral particles (green) were detected using polyclonal antibodies raised in mice and FITC-coupled anti-mouse immunoglobulins antibodies, and amoebae were stained with Evans blue (red). An eclipse phase can be observed 2h post-infection and amoebal lysis occurs 16h post-infection.

168x66mm (300 x 300 DPI)

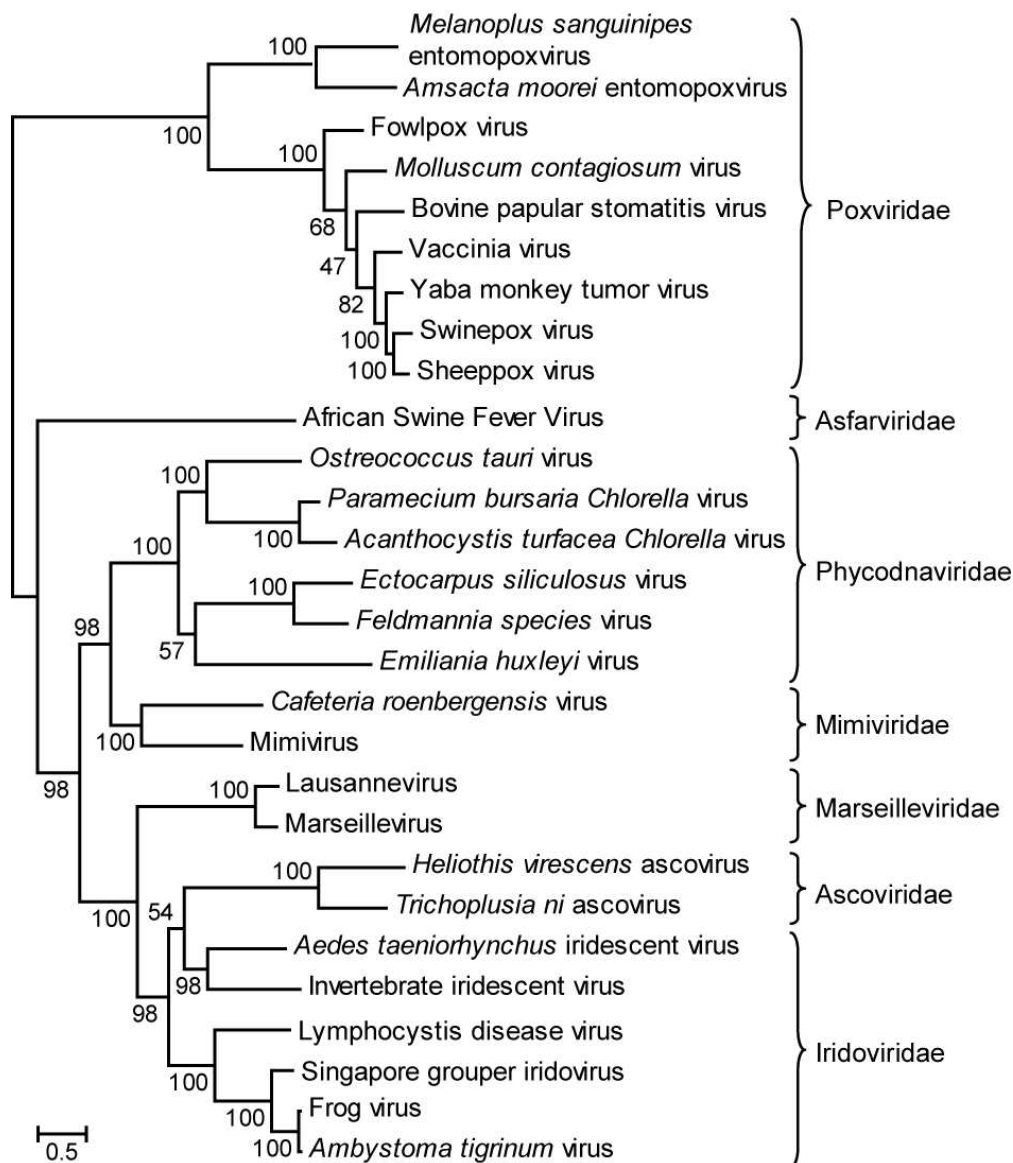


Figure 2. Core gene phylogeny

Lausannevirus is closely related to Marseillevirus as shown in this phylogenetic maximum-likelihood tree of representative NCLDV viruses based on the concatenated alignment of 5 core proteins (A32 ATPase, major capsid protein, D5 helicase, DNA polymerase B and A1L/VLTF2 transcription factor).

79x92mm (300 x 300 DPI)

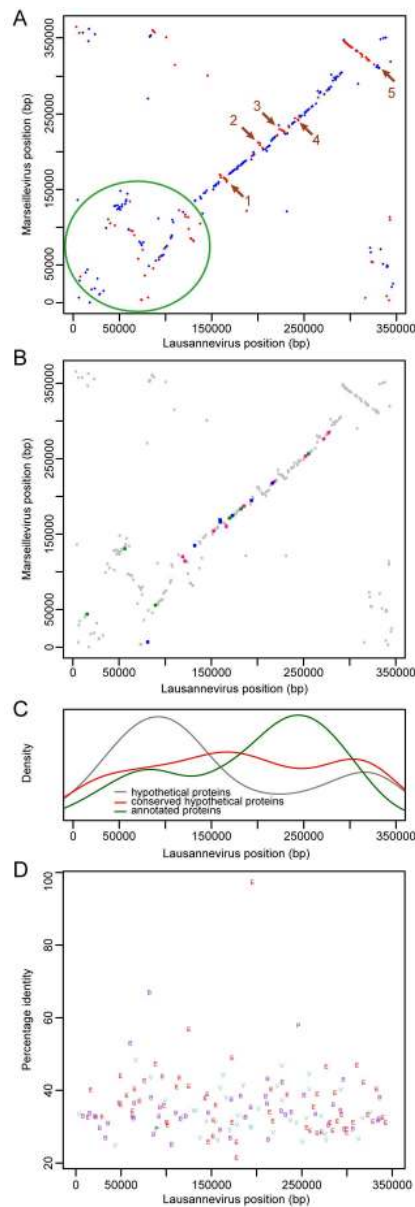


Figure 3. Genome comparison of Lausannevirus and Marseillevirus

(A) Dot-plot based on the genomic position of orthologous proteins, as defined by reciprocal best blast hit between Lausannevirus and Marseillevirus. Genes presenting the same strand orientation are shown in blue whereas those encoded on opposite strands are shown in red. The first part of the genome exhibits little colinearity, i.e. same gene order and orientation (green circle), whereas the second part is largely colinear, showing 5 main inverted segments (brown arrows). (B) Similar representation highlighting the position of Group I, II and III core genes in pink, blue and green respectively. (C) Distribution along the genome of hypothetical proteins (grey), conserved hypothetical proteins (red) and proteins with annotated family or gene function (black). (D) Genome position and percentage identity of proteins with best blast hit against viruses (cyan "V"), phages (dark green "P"), bacteria (blue "B") and eukarya (pink "E").

84x234mm (300 x 300 DPI)

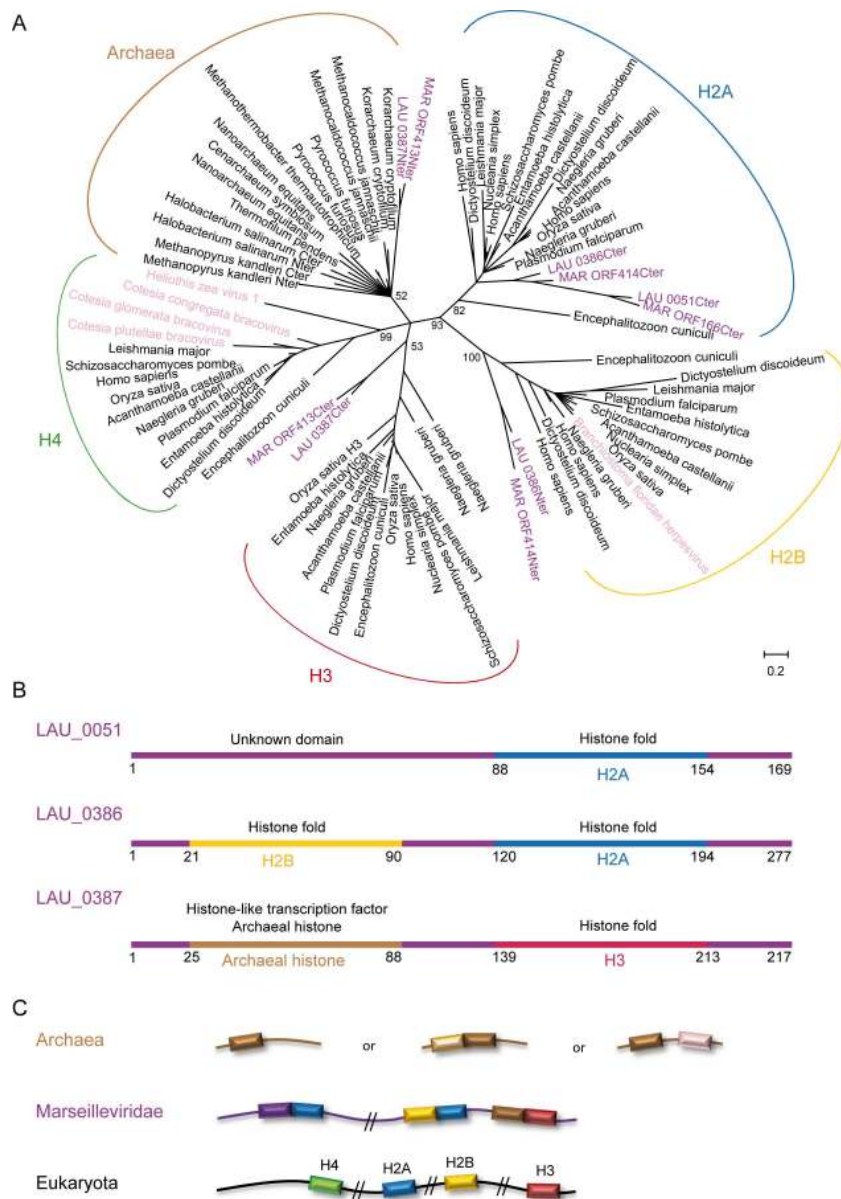


Figure 4. Description of viral histones.

(A) Bayesian tree of eukaryote and viral histones as well as archaeal histone-like proteins. Lausannevirus and Marseillevirus proteins presenting a predicted histone fold (purple) branch deeply with the different types of eukaryotic histones. The position of LAU\_0387 N-terminal histone fold is basal and cannot be confidently inferred, changing according to the model used between Archaea and H4. On the contrary, histones from other viruses (pink) are found branching very close to classical eukaryotic histones. (B) Schematic representation of Lausannevirus proteins with their Pfam hits and the annotation of their best BLAST hit (excluding Marseillevirus) indicated respectively above and under each domain. (C) Existing arrangement of histone proteins. In Archaea, histones are found as single gene for example in *Methanosarcina acetivorans*, as fused genes forming doublets e.g. in *Methanopyrus kandleri* or two different histones e.g. in *Nanoarchaeum equitans*. Marseilleviridae encode a protein with one H2A-like histone domain and two histone doublets that were colored according to the annotation of the best blast hit. Eukaryota do have multiple histone

encoding genes for each H2A, H2B, H3 and H4 family that can be found close to each other or in different chromosomal loci.

161x230mm (300 x 300 DPI)

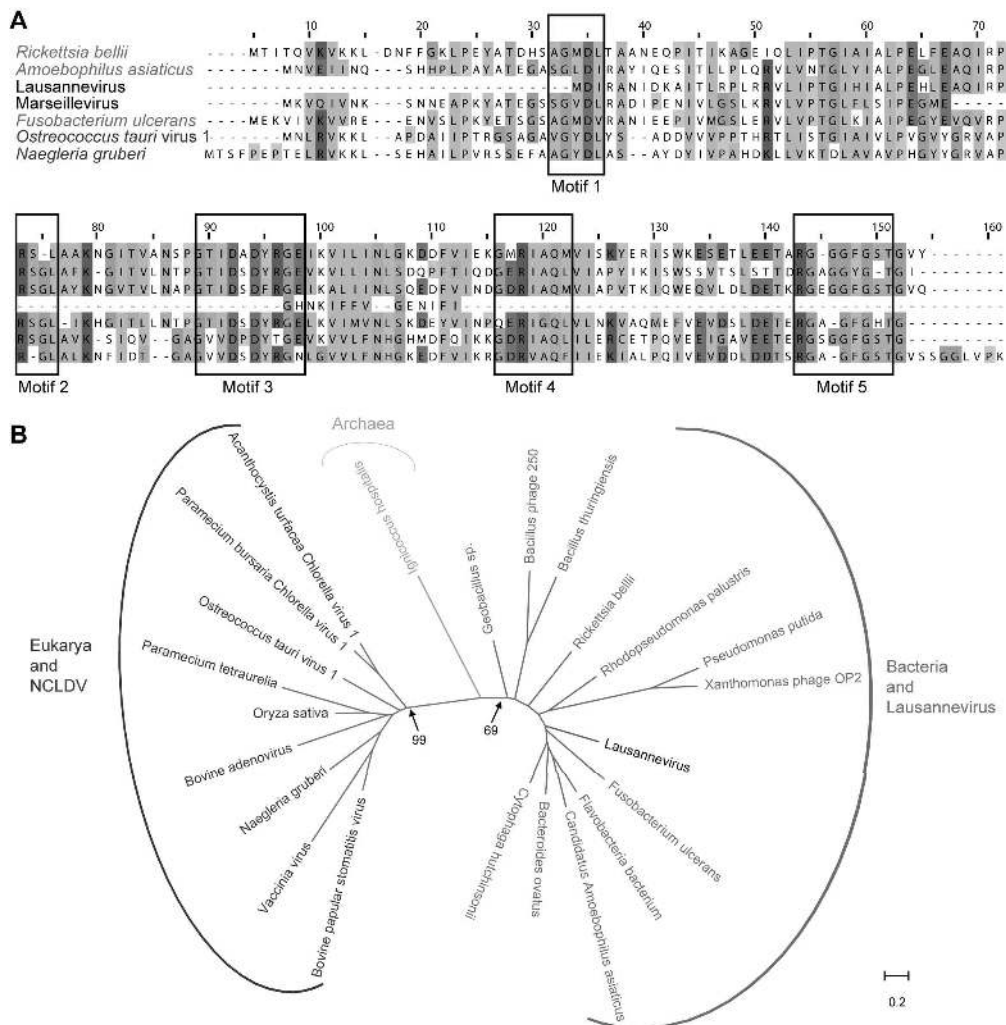


Figure 5. Truncated dUTPase  
 Alignment (A) and maximum-likelihood tree (B) of the dUTPase proteins from selected NCLDV, phages, representative eukaryotes and bacteria as well as the best nr hits of Lausannevirus/Marseillevirus dUTPase. Respectively, the N-terminus and C-terminus of Lausannevirus and Marseillevirus are truncated. Whereas other NCLDV proteins cluster together with eukaryotic dUTPase, Lausannevirus protein clearly clusters with prokaryotic dUTPase, although its relative position within bacteria cannot be robustly inferred.

167x170mm (300 x 300 DPI)

## Supplementary Tables

Virus species	Family	Genome size	Virus size (diameter or diameter x length)	GC%	Predicted CDSs
Marseillevirus	Marseilleviridae	368,454	approx 250 nm	45	457
Lausannevirus	Marseilleviridae	346,754	190-220 nm	43	450
Acanthamoeba polyphaga Mimivirus	Mimiviridae	1,181,404	approx. 500 nm *	28	911 (+75)
Cafeteria roenbergensis virus 1	Phycodnaviridae	617,453 **	300 nm	23	544
Paramecium bursaria Chlorella virus 1	Phycodnaviridae	330,743	175-190 nm	40	367
Paramecium bursaria Chlorella virus NY-2A	Phycodnaviridae	368,683	175-190 nm	41	404
Ostreococcus tauri virus 5	Phycodnaviridae	186,234	100-120 nm	45	268
Ostreococcus tauri virus 1	Phycodnaviridae	191,761	100-120 nm	45	232
Emiliana huxleyi virus 86	Phycodnaviridae	407,339	170-190 nm	40	472
Ectocarpus siliculosus virus 1	Phycodnaviridae	335,593	150-190 nm	52	240
Feldmannia sp. virus 158	Phycodnaviridae	154,641	150 nm	53	150
Ambystoma tigrinum virus	Iridoviridae	106,332	160-180 nm	54	96
Tiger Frog virus 3	Iridoviridae	105,903	120-200 nm	55	98
Singapore grouper iridovirus	Iridoviridae	140,131	120-200 nm	49	162
Lymphocystis disease virus 1	Iridoviridae	102,653	120-200 nm	29	195
Invertebrate iridescent virus 6	Iridoviridae	212,482	120-140 nm	29	234
Invertebrate iridescent virus 3	Iridoviridae	190,132	120-200 nm	48	126
Heliothis virescens ascovirus	Ascoviridae	186,262	130x400 nm	46	180
Trichoplusia ni ascovirus	Ascoviridae	174,059	130x400 nm	35	165
Spodoptera frugiperda ascovirus	Ascoviridae	156,922	150x400 nm	49	123
Diadromus pulchellus ascovirus	Ascoviridae	119,343	150x250 nm	50	119
African swine fever virus	Asfaviridae	170,101	175-215 nm	38	151
Canarypox virus	Chordopoxvirinae	359,853	160-190 nm	30	328
Fowlpox virus	Chordopoxvirinae	288,539	approx 200x300 nm	30	261
Lumpy skin disease virus NI-2490	Chordopoxvirinae	150,773	approx 200x300 nm	25	156
Myxoma virus	Chordopoxvirinae	161,773	approx 200x300 nm	43	170
Molluscum contagiosum virus	Chordopoxvirinae	190,289	approx 200x300 nm	63	163
Vaccinia virus	Chordopoxvirinae	194,711	200-400 nm	33	223
Orf virus	Chordopoxvirinae	139,962	approx 200x300 nm	63	130
Swinepox virus	Chordopoxvirinae	146,454	approx 200x300 nm	27	150
Yaba monkey tumor virus	Chordopoxvirinae	134,721	approx 200x300 nm	29	140
Crocodilepox virus	Chordopoxvirinae	190,054	approx 200x300 nm	61	173
Melanoplus sanguinipes entomopoxvirus	Entomopoxvirinae	236,120	200x320 nm	18	267

\* without fibrils

\*\* without large and highly repetitive regions found on both ends of the viral chromosome and increasing the genome size to 730kb

Table S1. Main features comparison of representative NCLDV viruses



Cluster number	Number of proteins in cluster	Cluster annotation	ORFs included	Related Marseillevirus cluster
1	21	MORN repeat-containing protein	LAU_0256, LAU_0261, LAU_0147, LAU_0255, LAU_0187, LAU_0253, LAU_0254, LAU_0148, LAU_0193, LAU_0296, LAU_0044, LAU_0103, LAU_0152, LAU_0063, LAU_0146, LAU_0155, LAU_0074, LAU_0198, LAU_0356, LAU_0185, LAU_0150	cluster1
2	9	Conserved hypothetical protein	LAU_0022, LAU_0429, LAU_0012, LAU_0020, LAU_0011, LAU_0021, LAU_0023, LAU_0430, LAU_0026	cluster5
3	5	Putative restriction endonuclease	LAU_0252, LAU_0097, LAU_0160, LAU_0391, LAU_0335	cluster4
4	5	Conserved hypothetical protein	LAU_0449, LAU_0184, LAU_0039, LAU_0002, LAU_0180	cluster8
5	4	Hypothetical protein	LAU_0156, LAU_0157, LAU_0049, LAU_0288	cluster13
6	3	Putative helicase	LAU_0183, LAU_0001, LAU_0038	cluster7
7	3	Conserved hypothetical protein	LAU_0432, LAU_0117, LAU_0118	cluster16
8	3	Putative Vsr/MutH/archaeal HJR family endonuclease	LAU_0307, LAU_0265, LAU_0374	cluster10
9	3	MORN repeat-containing protein	LAU_0069, LAU_0073, LAU_0365	cluster1
10	3	Hypothetical protein	LAU_0144, LAU_0082, LAU_0090	
11	3	Hypothetical protein	LAU_0446, LAU_0092, LAU_0191	
12	2	Conserved putative membrane protein	LAU_0293, LAU_0282	cluster17
13	2	Hypothetical protein	LAU_0390, LAU_0392	
14	2	Serine/threonine-protein kinase	LAU_0205, LAU_0154	cluster2
15	2	Serine/threonine-protein kinase	LAU_0202, LAU_0376	cluster2
16	2	Hypothetical protein	LAU_0360, LAU_0358	
17	2	MORN repeat-containing protein	LAU_0289, LAU_0245	cluster1
18	2	Hypothetical protein	LAU_0444, LAU_0094	
19	2	Putative serine/threonine protein kinase	LAU_0361, LAU_0359	cluster11
20	2	Hypothetical protein	LAU_0176, LAU_0175	
21	2	Hypothetical protein	LAU_0192, LAU_0179	
22	2	Hypothetical protein	LAU_0085, LAU_0083	
23	2	Hypothetical protein	LAU_0045, LAU_0042	cluster6
24	2	Putative restriction endonuclease	LAU_0336, LAU_0251	cluster4
25	2	Hypothetical protein	LAU_0086, LAU_0096	
26	2	Hypothetical protein	LAU_0239, LAU_0334	
27	2	Ubiquitin	LAU_0263, LAU_0250	
28	2	Hypothetical protein	LAU_0101, LAU_0041	

Table S2: Protein families.

Main families of proteins present in the genome of Lausannevirus as identified by similarity-based clustering. Families with similar annotations have not been merged as this sometimes reflected the presence of truncated proteins of a larger family. Clusters with annotated functions are found in both viruses in addition to a family of conserved hypothetical protein. Families specific to Lausannevirus are composed of proteins of unknown function, with the exception of the ubiquitin that is present in two copies in Lausannevirus and in one copy only in Marseillevirus

### Supplementary Figure legends

Figure S1. The viral factory

Electron microscopy showing the viral factory comprised of a large granular area in the cytoplasm of the amoeba containing numerous empty viral particles (white arrows). Several mitochondria can be observed near the viral factory. Bar: 2 $\mu$ m

Figure S2. Putative origin of replication

Cumulative GC skew and gene orientation skews were performed using a sliding window of 1000 bp with 100 bp overlap. The minimum of the curve and the reversion of the slope occurs around position 195'000 bp for the gene orientation skew. On the contrary, the minimum of the GC skew is located at about 156'000 bp. However, a secondary minimum is present around 194'500 bp in agreement with the gene orientation skew, suggesting that this area corresponds to the origin of replication (grey vertical bar).

Figure S3: Phylogeny of eukaryotic, archaeal and viral histone-like proteins.

(A) Maximum likelihood (100 bootstraps) and (B) neighbor-joining (1000 bootstraps) phylogeny of eukaryotic, archaeal and viral histone-like proteins.

Lausannevirus and Marseillevirus histone-like proteins are highlighted in purple and other viral proteins are shown in pink. Bootstrap values are indicated only for major branches.

Figure S4: Neighbor-joining phylogeny of each viral histone-like protein.

The 10 best BLAST hits of each viral histone-like protein were retrieved to build a neighbor-joining tree (1000 bootstraps). (A) LAU\_0051Cter & MAR\_ORF166Cter, (B) LAU\_0386Nter & MAR\_ORF414Nter, (C) LAU\_0386Cter & MAR\_ORF414Cter, (D) LAU\_0387Nter & MAR\_ORF413Nter, (E) LAU\_0387Cter & MAR\_ORF413Cter. Sequences obtained by BLAST using as query histone-like proteins of Lausannevirus have the prefix LAU whereas those identified by BLAST using histone-like proteins of Marseillevirus have the prefix MAR. Both prefixes are indicated if the sequence belongs to the 10 best hits for both viruses.

Figure S5. Inteins of Lausannevirus and Marseillevirus ribonucleotide reductase.

Alignment of Lausannevirus and Marseillevirus inteins together with other viral and eukaryotic inteins contained in the ribonucleotide reductase large subunit. Among the conserved splicing (A, B F, G) and endonuclease (C, E, H) motifs identified and annotated according to Perler *et al.* (Perler *et al.*, 1997), Lausannevirus only lacks block C whereas Marseillevirus lacks all endonuclease blocks.

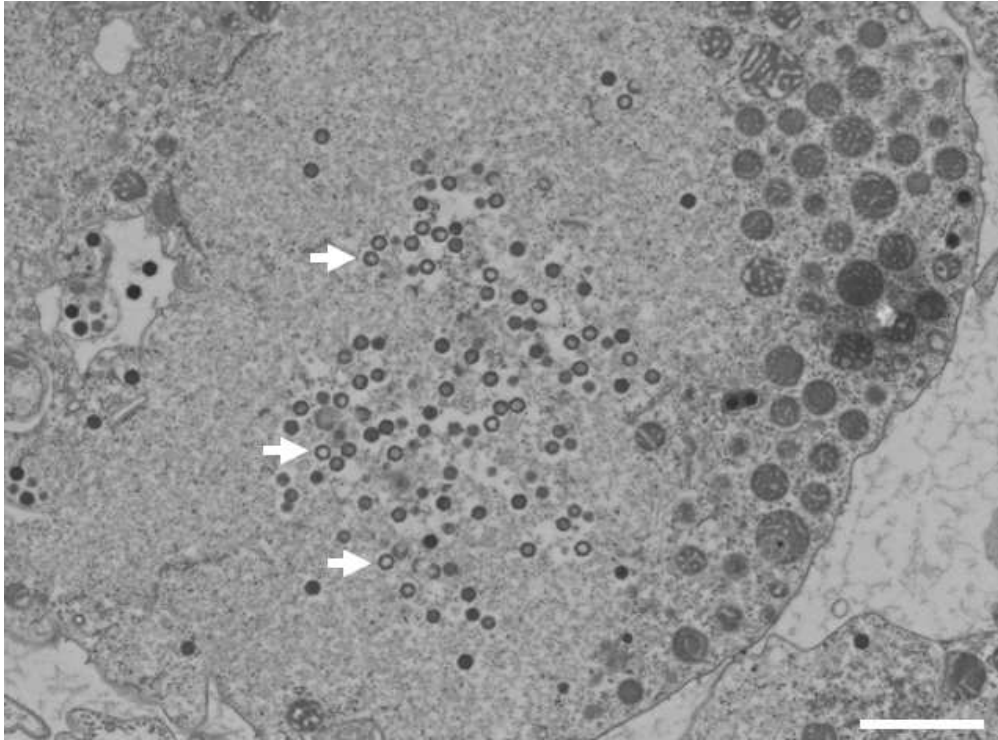


Figure S1. The viral factory  
Electron microscopy showing the viral factory comprised of a large granular area in the cytoplasm of the amoeba containing numerous empty viral particles (white arrows). Several mitochondria can be observed near the viral factory. Bar: 2 $\mu$ m

168x124mm (96 x 96 DPI)

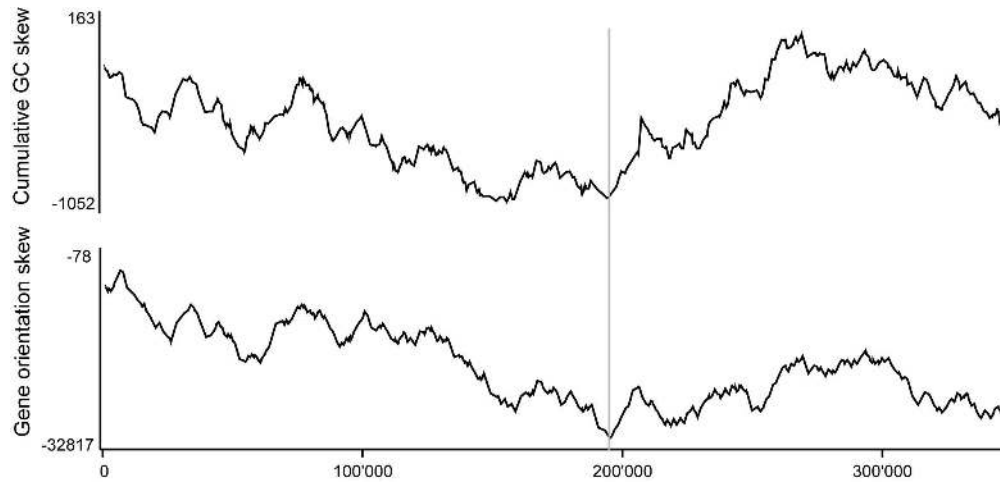


Figure S2. Putative origin of replication

Cumulative GC skew and gene orientation skews were performed using a sliding window of 1000 bp with 100 bp overlap. The minimum of the curve and the reversion of the slope occurs around position 195'000 bp for the gene orientation skew. On the contrary, the minimum of the GC skew is located at about 156'000 bp. However, a secondary minimum is present around 194'500 bp in agreement with the gene orientation skew, suggesting that this area corresponds to the origin of replication (grey vertical bar).

166x79mm (300 x 300 DPI)

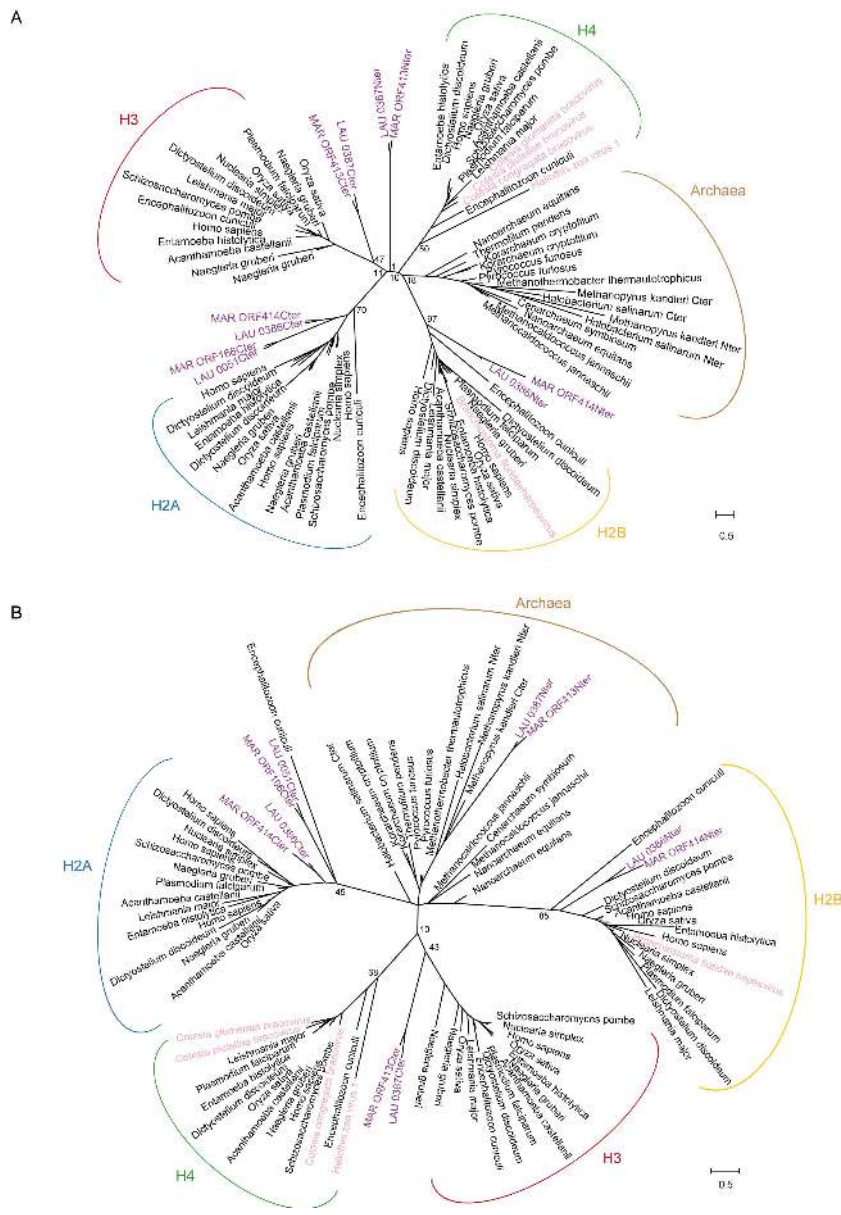


Figure S3: Phylogeny of eukaryotic, archaeal and viral histone-like proteins. (A) Maximum likelihood (100 bootstraps) and (B) neighbor-joining (1000 bootstraps) phylogeny of eukaryotic, archaeal and viral histone-like proteins. Lausannevirus and Marseillevirus histone-like proteins are highlighted in purple and other viral proteins are shown in pink. Bootstrap values are indicated only for major branches.

204x284mm (300 x 300 DPI)

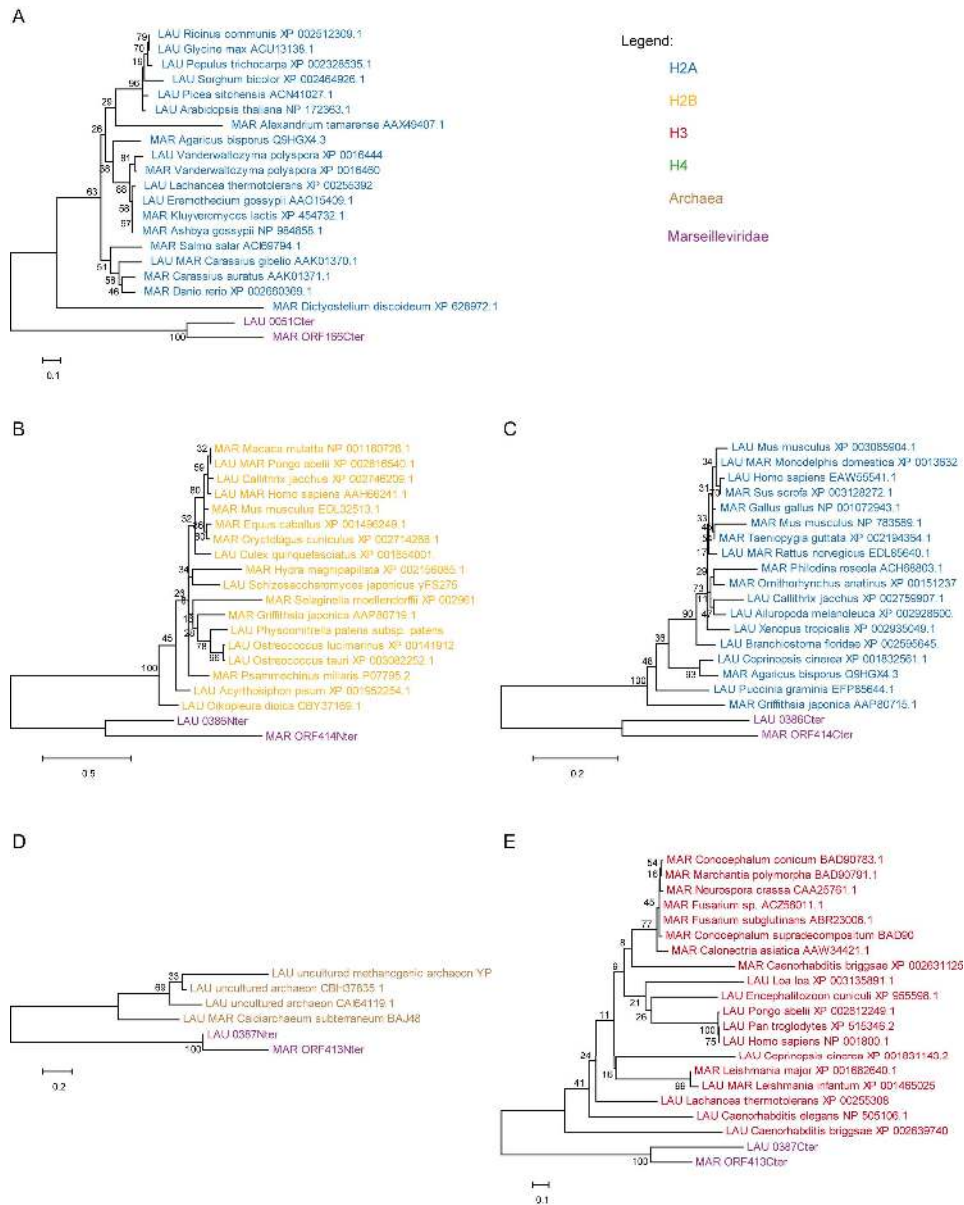


Figure S4: Neighbor-joining phylogeny of each viral histone-like protein. The 10 best BLAST hits of each viral histone-like protein were retrieved to build a neighbor-joining tree (1000 bootstraps). (A) LAU\_0051Cter & MAR\_ORF166Cter, (B) LAU\_0386Nter & MAR\_ORF414Nter, (C) LAU\_0386Cter & MAR\_ORF414Cter, (D) LAU\_0387Nter & MAR\_ORF413Nter, (E) LAU\_0387Cter & MAR\_ORF413Cter. Sequences obtained by BLAST using as query histone-like proteins of Lausannevirus have the prefix LAU whereas those identified by BLAST using histone-like proteins of Marseillevirus have the prefix MAR. Both prefixes are indicated if the sequence belongs to the 10 best hits for both viruses.

208x261mm (300 x 300 DPI)

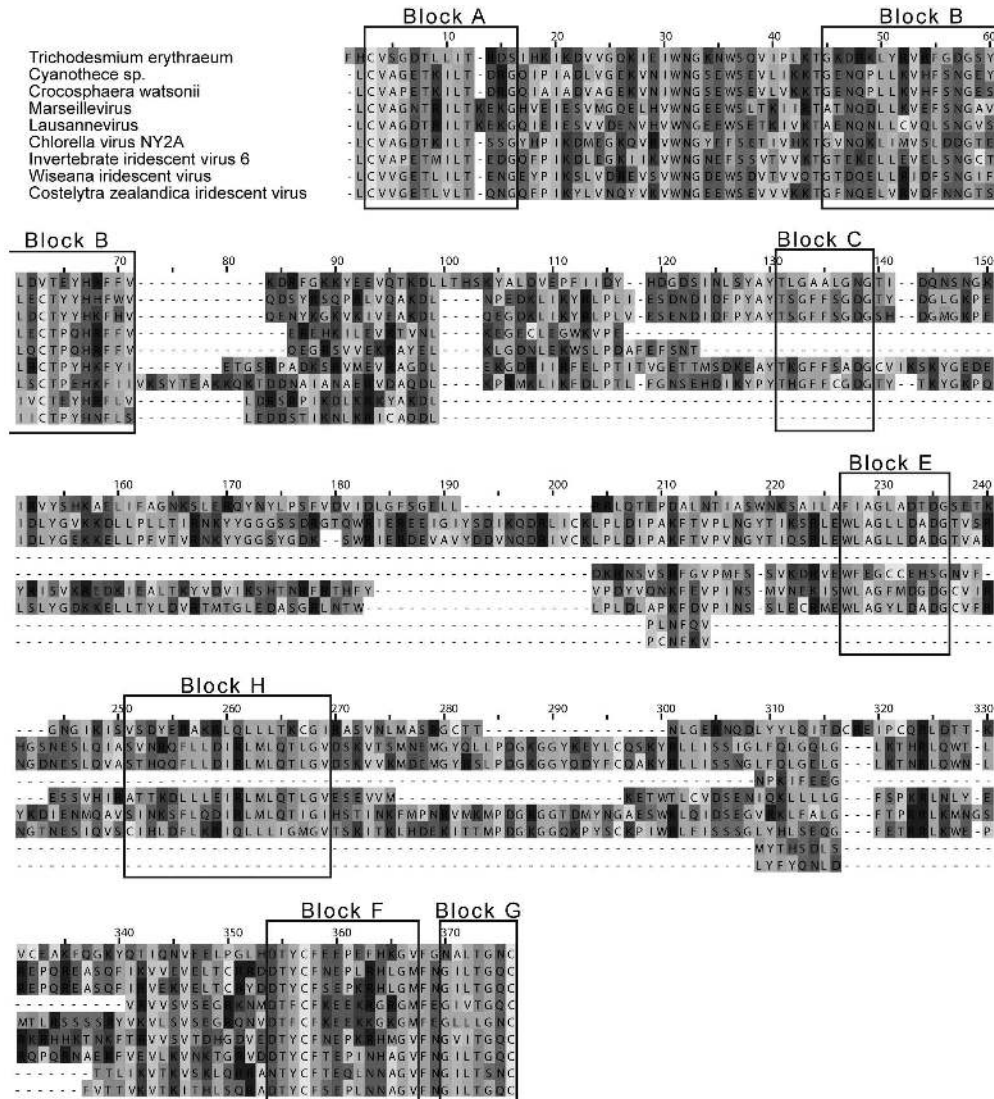


Figure S5. Inteins of Lausannevirus and Marseillevirus ribonucleotide reductase. Alignment of Lausannevirus and Marseillevirus inteins together with other viral and eukaryotic inteins contained in the ribonucleotide reductase large subunit. Among the conserved splicing (A, B, F, G) and endonuclease (C, E, H) motifs identified and annotated according to Perler et al. (Perler et al., 1997), Lausannevirus only lacks block C whereas Marseillevirus lacks all endonuclease blocks.

167x183mm (300 x 300 DPI)