Article

Kaumoebavirus, a New Virus That Clusters with Faustoviruses and *Asfarviridae*

Leena H. Bajrai [1,2], Samia Benamar [1], Esam I. Azhar [3], Catherine Robert [1], Anthony Levasseur [1], Didier Raoult [1] and Bernard La Scola [1,*]

[1] Unité des Rickettsies, URMITE UMR CNRS 7278 IRD 198 INSERM U1095, Facultés de Médecine et de Pharmacie, IHU Méditerranée Infection, Aix-Marseille Université, Marseille 13005, France; leenaasd@gmail.com (L.H.B.); benamarsamia@yahoo.com (S.B.); catherine.robert@univ-amu.fr (C.R.); anthony.levasseur@univ-amu.fr (A.L.); didier.raoult@gmail.com (D.R.)

[2] Department of Biochemistry, Faculty of Science, King Abdulaziz University, 21362 Jeddah, Saudi Arabia

[3] Special Infectious Agents Unit, King Fahd Medical Research Center, and Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, 21362 Jeddah, Saudi Arabia; eazhar@kau.edu.sa

[*] Correspondence: bernard.la-scola@univ-amu.fr; Tel.: +33-4-91-38-55-17

Abstract: In this study, we report the isolation of a new giant virus found in sewage water from the southern area of Jeddah (Saudi Arabia), with morphological and genomic resemblance to Faustoviruses. This new giant virus, named Kaumoebavirus, was obtained from co-culture with *Vermamoeba vermiformis*, an amoeboid protozoa considered to be of special interest to human health and the environment. This new virus has ~250 nm icosahedral capsids and a 350,731 bp DNA genome length. The genome of Kaumoebavirus has a coding density of 86%, corresponding to 465 genes. Most of these genes (59%) are closely related to genes from members of the proposed order *Megavirales*, and the best matches to its proteins with other members of the *Megavirales* are Faustoviruses (43%) and Asfarviruses (23%). Unsurprisingly, phylogenetic reconstruction places Kaumoebavirus as a distant relative of Faustoviruses and Asfarviruses.

1. Introduction

The story of giant viruses began with the discovery of Mimivirus [1,2], the first giant virus to be described and which has now been isolated from many environmental and human samples [3–9]. Subsequently, other viruses such as Marseilleviruses [4,10–12], *Cafeteria roenbergensis* virus (CroV) [5], Pandoraviruses [13], *Pithovirus sibericum* [14] and *Mollivirus sibericum* [15] were discovered using *Acanthamoeba* sp. as a support for co-culture, with the exception of CroV, which was co-isolated alongside its host, the marine flagellate *Cafeteria roenbergensis*. These viruses were recognized as members of the nucleocytoplasmic large DNA viruses (NCLDVs) clade, and it was recently proposed that they should be merged into *Megavirales*, a new viral order [16]. In 2015, Faustovirus, a new Asfarvirus lineage of giant viruses, was discovered [17]. It was the first to be isolated by co-culture with an amoeba other than *Acanthamoeba* sp., *Vermamoeba vermiformis*, and later additional isolates were obtained [18]. Using the same support for co-culture, in this paper, we report the isolation and genome analysis of a new virus we named Kaumoebavirus (meaning King Abdulaziz University amoeba virus), isolated from an environmental sample (sewage water) obtained in the southern area of Jeddah, Saudi Arabia.

2. Materials and Methods

2.1. Isolation and Routine Subculture

A sewage water sample collected from the southern area of Jeddah was inoculated onto a *V. vermiformis* (strain CDC19) monolayer in starvation medium at 30 °C, as previously described in detail [19]. Viral particles observed in the culture supernatant were serially diluted from 10^{-1} to 10^{-11} using the end-point dilution method. Briefly, microplates were inoculated with each dilution in four wells that already contained rinsed *V. vermiformis* at 10^6 per mL. Only one well that simultaneously had a complete lysis at 30 °C in the highest dilution was used for virus production [20]. Viral particles were purified as previously described [19].

2.2. Studying the Viral Cycle Using TEM

A co-culture was prepared by inoculating a 25 cm^2 cell culture flask in *V. vermiformis* starvation medium at 10^6 cells per mL [17] with Kaumoebavirus at 30 °C at a ratio of 1:10. Thirty minutes post-infection (PI), the *V. vermiformis*-infected monolayer was recovered and pelleted for 10 min at $5000 \times g$, which was considered as 0 h. The pellet was suspended in 1 mL of phosphate buffered saline (PBS), fixed in 1 mL of 2% glutaraldehyde 0.1 M cacodylate buffer, and then incubated for at least 1 h at 4 °C, preparing for transmission electron microscopy (TEM). Then, 200 µL infected cells were cytocentrifugated on two slides for Hemacolor and DAPI for nucleic acid labeling (Molecular Probes, Oregon, USA). This step was repeated every 4 h for the next 28 h PI. All samples were prepared for electron microscopy as previously described [17]. Grids were observed under Tecnai G2 transmission electron microscope (FEI, Oregon, USA) at 200 keV through ultrathin 70 nm sections.

2.3. Genome Study

The Kaumoebavirus genome was sequenced as previously described [17] by MiSeq Technology (Illumina, Inc., San Diego, CA, USA) using paired-end and mate-pair applications in parallel, in a 2×251 bp run for each bar-coded library. The reads were assembled de novo into a contig using the Spades assembler software v1.0 combined with GapFiller, which was used to enhance the assembly [21,22]. Protein-coding regions were predicted using GeneMark [23]. tRNA were searched using ARAGORN software [24].

The phylogenetic trees were built by protein sequence alignment using the MUSCLE program with the default parameters [25]. Phylogeny reconstructions were performed for Kaumoebavirus and other *Megavirales* members by the maximum likelihood method using FastTree [26], with the default parameters (JTT evolutionary model, discrete gamma model with 20 rate categories), based on the conserved genes.

A comparative analysis of the Kaumoebavirus genome and those of the *Asfarviridae* and Faustoviruses genomes was generated by creating a reference database of all the protein sequences. COGtriangles [27] and OrthoMCL [28] clustering algorithms were used to create protein clusters. Pangenome and core genes were defined using GET_HOMOLOGUES [29] with the following parameters: 75% as minimum coverage, 30% as minimum identity in BLASTP pairwise alignments and 1×10^{-5} as maximum E-value.

2.4. Accession Number

The complete genome sequence of Kaumoebavirus was submitted to GenBank and assigned the accession number KX552040.

3. Results

3.1. Replication Cycle

At an early stage (two hours PI), we observed the entry of Kaumoebavirus particles through phagocytosis into the *V. vermiformis* cytoplasm (Figure 1A). After phagocytosis, at three hours PI, the viral particles were mostly located as clumps of two to four particles within vacuoles (Figure 1B). At four hours PI, the "eclipse" phase was obvious, as no viruses could be observed within the cytoplasm or only few visible vacuoles appeared empty, which is likely due to the release of the viral DNA. Virus factories could be observed in the cytoplasm and close to the nucleus at six hours PI. Newly-formed mature virions of Kaumoebavirus were observed from eight to 16 h PI accumulating within the virus factory (Figure 1C). The nuclear morphology of the host cell did not change during the Kaumoebavirus replication cycle. During the final stages of morphogenesis (20 h PI), viral particles accumulated massively within the virus factory (Figure 1D) before being released from each infected cell via lysis.
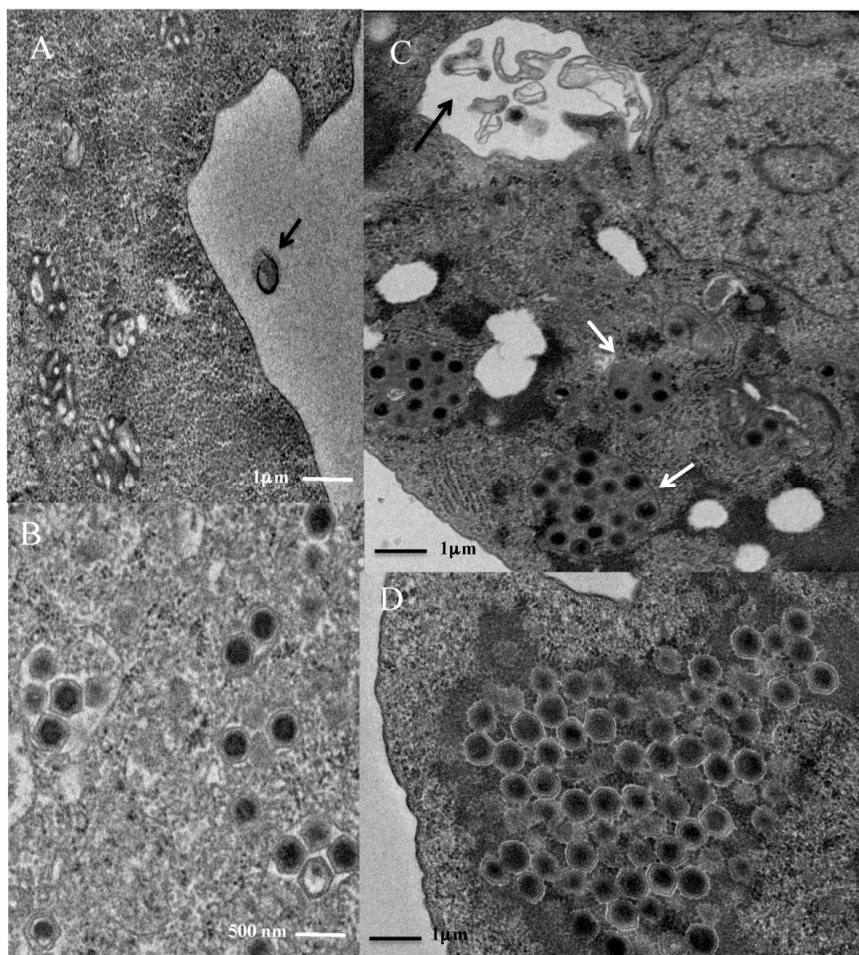
Figure 1. Observation of Kaumoebavirus in *Vermamoeba vermiformis* at selected times of development. Kaumoebavirus particles (black arrow) are phagocytized (**A**) then observed in the cytoplasm of *V. vermiformis* mostly packaged in clumps of two to four particles (**B**). After the eclipse phase, new virions may be observed within the virus factory (**C**,**D**). During the early stage of its formation, the virus factory of Kaumoebavirus is not round-shaped but poly-lobed. During the microtome cut, some slice planes showed newly formed viruses in these lobes as clusters of virions (white arrows). At this stage, remnants of phagocytized particles may be observed (black arrow) (**C**). At a late stage, the virus factory loses its poly-lobed aspect for a large round vacuole aspect and appears completely filled with Kaumoebavirus mature particles (**D**).

3.2. Genomic Features

The Kaumoebavirus genome is a circular double stranded DNA (dsDNA) of 350,731 bp (Figure 2). The genome is devoid of internal repeats (Figure S1), except for a unique long inverted tandem repeat of 1407 bp length.
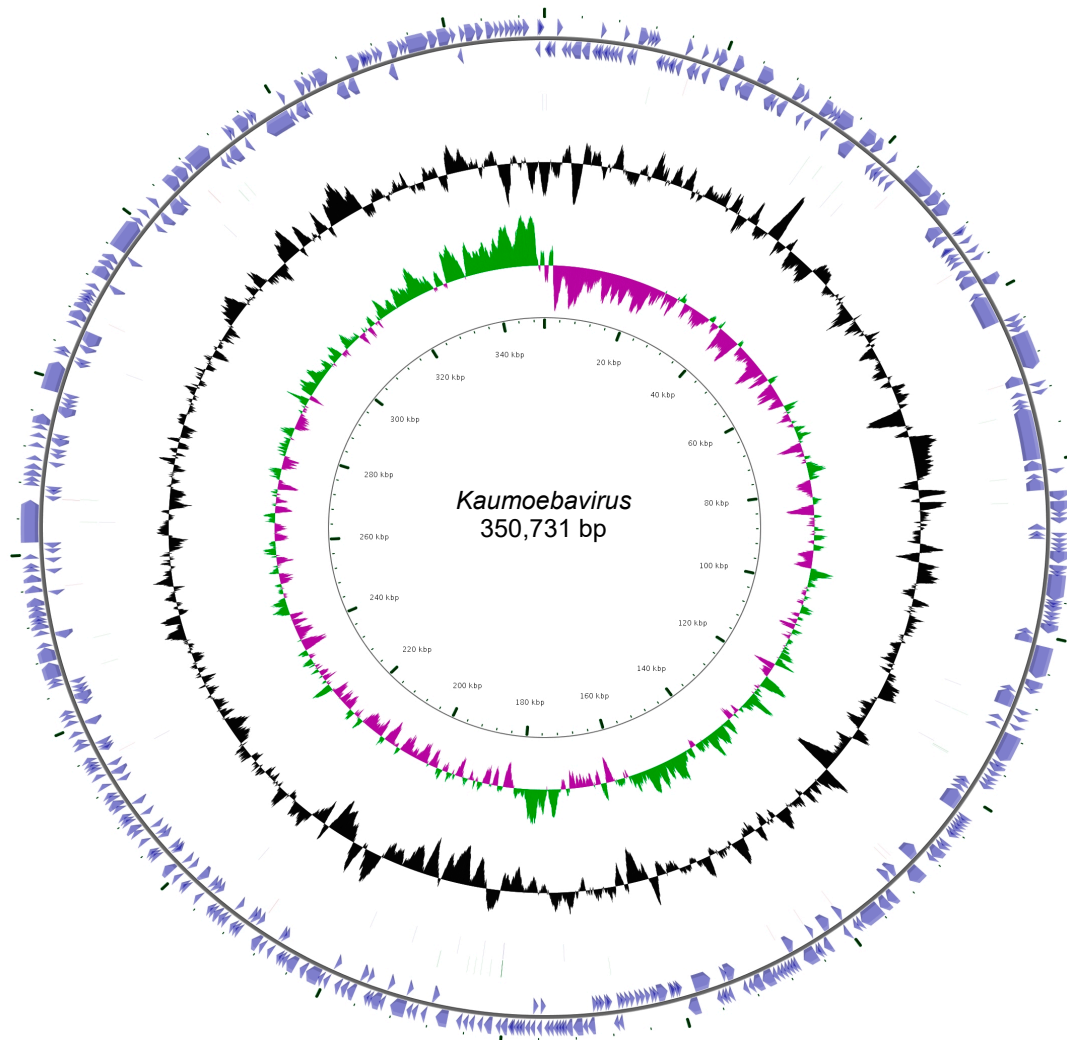

Figure 2. Circular representation of the Kaumoebavirus genome. The circles from the center to the outside show: GC skew (green/purple); GC content (black); open reading frames (ORFs) on the plus and minus strands (blue).

The Kaumoebavirus genome has a global GC content of 43.7% and a GC value of 44.5% when only coding sequences are considered. The GC content value of genes ranging from 28.8% to 60%, standard deviations (3.93) and quartiles (first quartile: 41.6, third quartile: 46.2) were calculated. Values exceeding the third quartile were classified as GC-rich sequences, while those inferior to the first quartile formed GC-poor sequences. Half of the Kaumoebavirus genes were classified as GC-poor (117) or GC-rich (118). More than 70% of these genes have no match against the non-redundant (NR) database. A relationship between GC content and gene length could be observed. The GC-rich genes are mainly the longest coding sequences (mean: 872 bp, first quartile: 306 bp, third quartile: 1089 bp), while the GC-poor genes are the shortest (mean: 326 bp, first quartile: 189 bp, third quartile: 580 bp).

We identified at total number of 465 proteins (amino acid numbers ranged from 113 aa to 6209 aa), representing an 86% coding density. ARAGORN software did not predict tRNA genes. The sequences

were analyzed using BLAST (with an E-value threshold of 10^{-2}) against the non-redundant protein sequence database. Of the Kaumoebavirus proteins which had homology in the databases, 74 (59%) were most similar to a virus protein, 19 (15%) to bacteria proteins, 31 (25%) to proteins of other eukaryotes, and two (2%) to Archaea proteins (Figure 3).
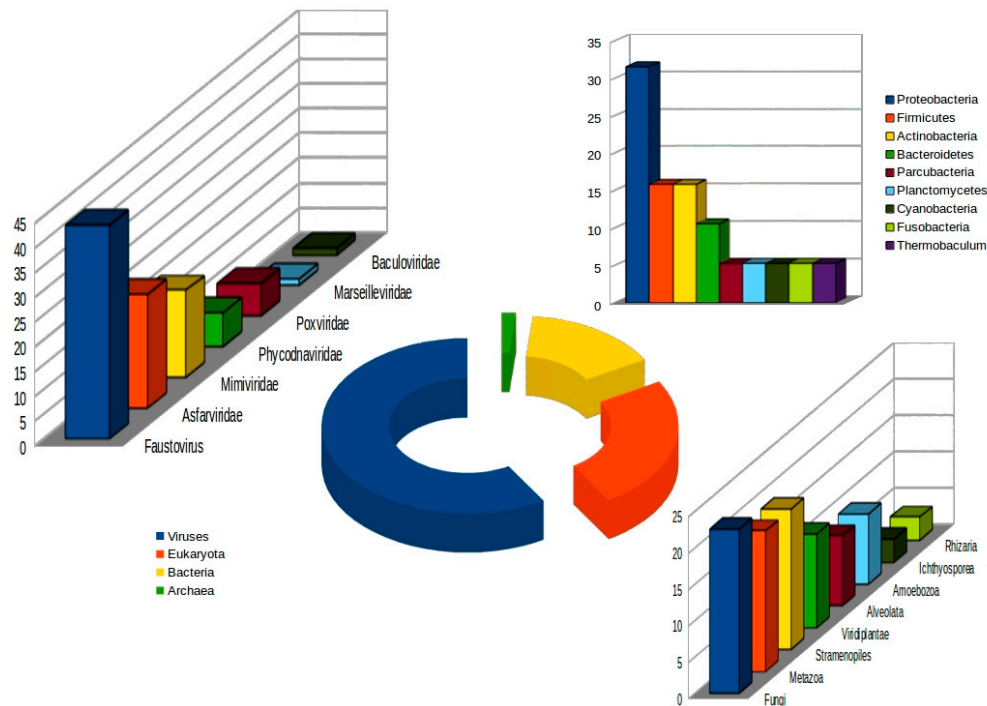

Figure 3. Distribution of the best-matching homologs to Kaumoebavirus proteins (the diagram shows only the 59% of genes with significant BLAST hits). Best-matching homologous proteins were determined using BLASTP (E value < 10^{-3}) against the non-redundant (NR) database at the National Center for Biotechnology Information (NCBI).

For 59% of the Kaumoebavirus proteins with detectable homologs to virus proteins, the best matches were to proteins from other members of the *Megavirales*, mostly Faustoviruses (32 cases; 43%) and Asfarviruses (17 cases; 23%). The remaining best viral matches were to proteins from Mimiviruses, Phycodnaviruses, *Poxviridae*, Marseilleviruses and *Baculoviridae* in 13, 5, 5, 1 and 1 cases, respectively. Despite the fact that Kaumoebavirus has a significant number of homologs to *Megavirales* proteins, no detectable collinearity between the Kaumoebavirus and Faustovirus E9, Faustovirus Liban, or African swine fever virus E75 could be observed. This suggests that the basic structure of the Kaumoebavirus genome is original, despite the significant number of shared proteins (Supplementary Materials Figure S2). However, as compared with other *Megavirales* members, the gene content of Kaumoebavirus is globally more similar to those of Faustoviruses and *Asfarviridae* than other *Megavirales*. This observation is illustrated by the phylogenetic clustering of the concatenated Kaumoebavirus protein sequences of the A32-like packaging ATPase and the family B DNA polymerase (Figure 4).
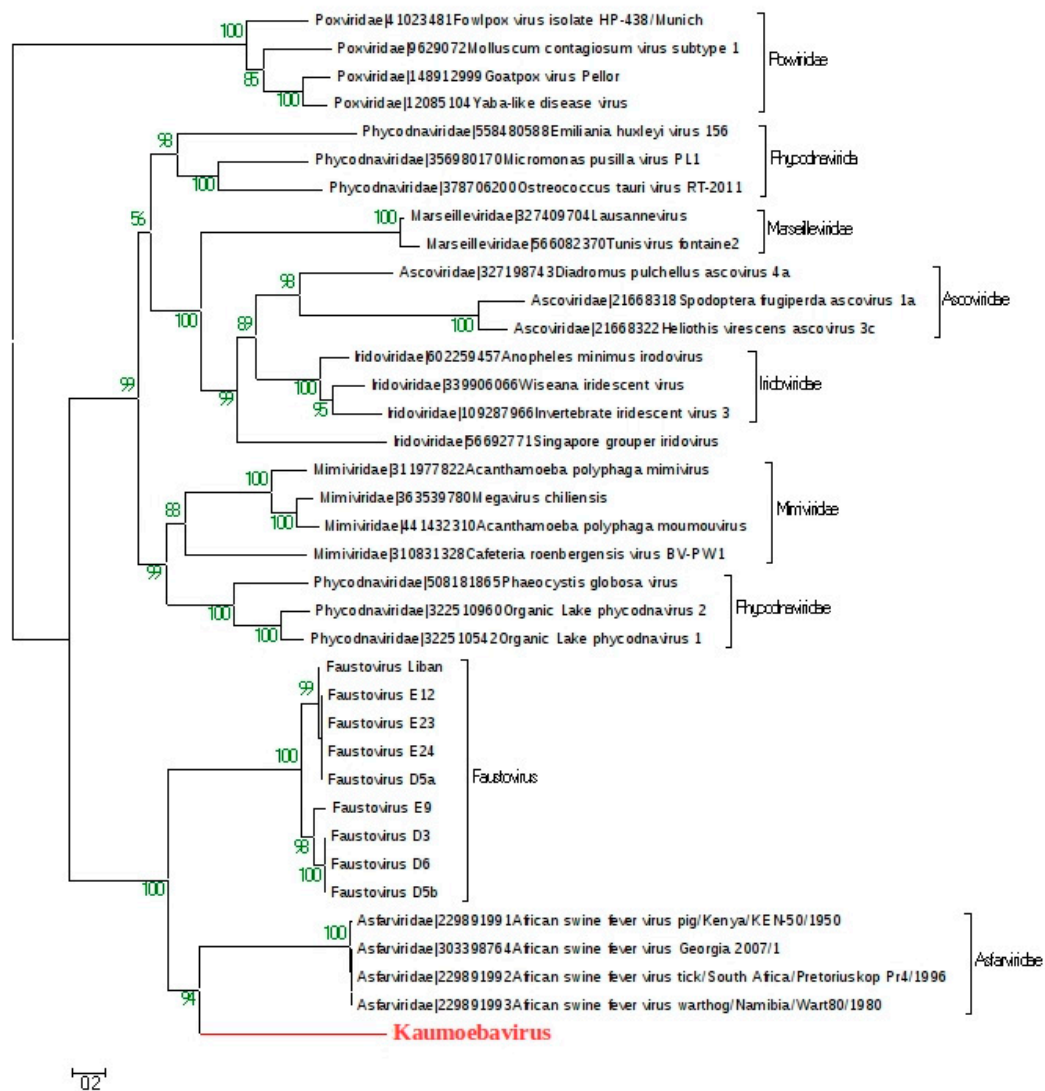
Figure 4. Phylogenetic reconstruction based on a concatenated A32-like packaging ATPase and the family B DNA polymerase. Phylogenetic analysis was performed using the maximum likelihood method based on protein sequences from Kaumoebavirus and representative members from the different families of *Megavirales* order.

In addition, the phylogeny of several conserved genes shows the same result (Supplementary Materials Figures S3–S6). From an evolutionary perspective, Kaumoebavirus appears as an intermediate, close to Faustoviruses and *Asfarviridae*. The Kaumoebavirus genomic coding sequence undergoes splicing, just as Faustovirus. Indeed, the capsid protein-encoding gene of Faustovirus spans a 5 kbp region (180,327 bp–185,966 bp), including six open reading frames (ORFs), five of which are exons separated by putative introns (Figure 5).
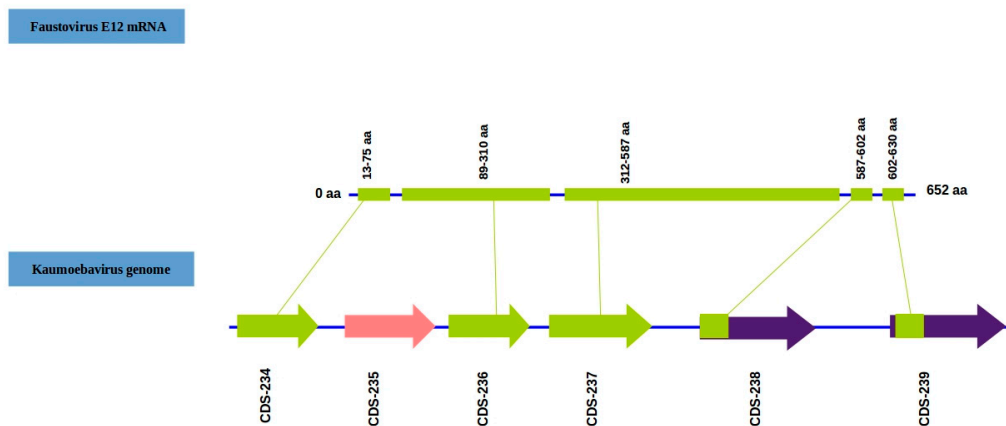
Figure 5. Overview of the genomic region in Kaumoebavirus that encodes the major capsid protein and the corresponding match in the Faustovirus E12 mRNA. The sequences coding for the major capsid protein are shown in green. A conserved nuclease found in the capsid region of both Faustoviruses and Kaumoebavirus is colored pink. Regions of Kaumoebavirus annotated as ORFs having a partial similarity with Faustovirus E12 mRNA are colored purple; only the parts with high similarity to Faustovirus E12 mRNA are colored green.

4. Discussion

Co-culture with *V. vermiformis* enabled the isolation of a new giant DNA virus from a sewage water sample obtained in the southern part of Jeddah. Although this virus, named Kaumoebavirus, replicates in *V. vermiformis*, its genome organization and gene content indicate that it does not belong to the recently described Faustoviruses [17,18]. The TEM images of Kaumoebavirus show penetration through phagocytosis, as described in previous studies for other giant viruses and newly formed virions within the virus factory (contrary to *Acanthamoeba polyphaga mimivirus* (APMV), for which the newly formed virions are observed only at the periphery of the virus factory) [1,6,17]. In comparison with Faustovirus, phagocytosis and internalization of the individual Kaumoebavirus viral particles by the amoeba occurred between two and three hours, the same as in the case of Faustovirus [17]. This is late, as compared to other giant viruses such as APMV, but likely due to the behavior of its amoeba host *V. vermiformis*. However, the eclipse phase of this virus occurred between four and five hours PI, which is shorter than for Faustovirus (around six hours PI). Like Faustovirus, the replication cycle of Kaumoebavirus lasts between 16 and 20 h [17].

The phylogenetic study clusters Kaumoebavirus as a distant relative of Faustoviruses and *Asfarviridae*. This suggests that all these viruses represent an independent cluster within the proposed *Megavirales* order, composed of the family *Asfarviridae* and two other putative families, "*Faustoviridae*", and Kaumoebavirus as a first isolate of a possible new family. Indeed, comparative genomics shows that the size of the core genome decreased dramatically when the Kaumoebavirus genome was included in the analysis, compared to the sizes estimated for the Faustoviruses (nine genomes) and the *Asfarviridae* (five genomes) genomes taken separately (Supplementary Materials Figure S7). However, its phylogenetic position is not yet clear. It is currently not possible to decide whether it has an ancestral or posterior relationship with *Asfarviridae* and Faustoviruses. The discovery of other viruses in this cluster could provide a more robust and intact phylogeny. Our characterization of Kaumoebavirus shows a dsDNA circular genome of 350 kb. Kaumoebavirus adds a potential new giant virus family with its 0.35 Mb genome. The Kaumoebavirus genome is smaller than that of Pandoraviruses, *Pithovirus sibericum*, *Mollivirus sibericum*, Faustoviruses and Mimiviruses. The range of GC content of the NCLDVs family members ranges from 17% (Entomopoxvirinae) to 64% (Parapoxvirus) [30]. Kaumoebavirus has a 43.7% GC content and the *Marseilleviridae* GC content's value is the closest one to this virus. In addition to the phylogenetic link between Faustovirus and Kaumoebavirus, the latter seems to present the same dispersed structure of the capsid [31]. Indeed, the Kaumoebavirus capsid protein gene

is split into five exons. The majority of Kaumoebavirus-encoded proteins (73%) do not have homologs, even in metagenomics databases such as environmental protein sequences database (env_nr). Of the Kaumoebavirus proteins presenting homology in the databases, they were more similar to eukaryotes (25%) than to bacteria proteins (15%), which is not surprising for a virus that infects eukaryotes.

Supplementary Materials: The following are available online at www.mdpi.com/1999-4915/8/11/278/s1, Figure S1. No large-scale repeated regions were observed in the Kaumoebavirus genome. A: Pairwise sequence alignment of the Kaumoebavirus against itself (dot plot). B: Pairwise sequence alignment of the concatenated peptide sequences of Kaumoebavirus proteins. Figure S2. Concatenated proteins pairwise sequence alignment (dot plot) of Kaumoebavirus against itself and the fully sequenced Faustovirus E9, Faustovirus Liban and African swine fever virus strain E75. The Kaumoebavirus retains no observed collinearity with the other viruses. Figure S3. Phylogenetic reconstruction performed for Kaumoebavirus and other *Megavirales* members. Phylogenetic reconstruction was performed using the maximum likelihood method and was based on A32-like packaging ATPase. Figure S4. Phylogenetic reconstruction performed for Kaumoebavirus and other *Megavirales* members. Phylogenetic reconstruction was performed using the maximum likelihood method and based on D5-like helicase. Figure S5. Phylogenetic reconstruction performed for Kaumoebavirus and other *Megavirales* members. Phylogenetic reconstruction was performed using the maximum likelihood method and based on DNA topoisomerase. Figure S6. Phylogenetic reconstruction performed for Kaumoebavirus and other *Megavirales* members. Phylogenetic reconstruction was performed using the maximum likelihood method and based on the transcription termination factor. Figure S7. Comparative genomics of Kaumoebavirus and representatives of Faustoviruses and *Asfarviridae* family members.

Author Contributions: L.H.B., C.R., performed the experiments; L.H.B., S.B. and B.L. analyzed the data and wrote the paper; E.I.A., A.L., D.R. contributed to the final version of the manuscript, and read and approved it.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. La Scola, B.; Audic, S.; Robert, C.; Jungang, L.; De Lamballerie, X.; Drancourt, M.; Birtles, R.; Claverie, J.M.; Raoult, D. A giant virus in amoebae. *Science* **2003**, *299*, 2033. [CrossRef] [PubMed]
2. Raoult, D.; Audic, S.; Robert, C.; Abergel, C.; Renesto, P.; Ogata, H.; La Scola, B.; Suzan, M.; Claverie, J.M. The 1.2-megabase genome sequence of Mimivirus. *Science* **2004**, *306*, 1344–1350. [CrossRef] [PubMed]
3. La Scola, B.; Desnues, C.; Pagnier, I.; Robert, C.; Barrassi, L.; Fournous, G.; Merchat, M.; Suzan-Monti, M.; Forterre, P.; Koonin, E.; et al. The virophage as a unique parasite of the giant Mimivirus. *Nature* **2008**, *455*, 100–104. [CrossRef] [PubMed]
4. Boyer, M.; Yutin, N.; Pagnier, I.; Barrassi, L.; Fournous, G.; Espinosa, L.; Robert, C.; Azza, S.; Sun, S.; Rossmann, M.G.; et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21848–21853. [CrossRef] [PubMed]
5. Fischer, M.G.; Allen, M.; Wilson, W.H.; Suttle, C.A. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19508–19513. [CrossRef] [PubMed]
6. Arslan, D.; Legendre, M.; Seltzer, V.; Abergel, C.; Claverie, J.M. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17486–17491. [CrossRef] [PubMed]
7. Yoosuf, N.; Yutin, N.; Colson, P.; Shabalina, S.A.; Pagnier, I.; Robert, C.; Azza, S.; Klose, T.; Wong, J.; Rossmann, M.G.; et al. Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. *Genome Biol. Evol.* **2012**, *4*, 1324–1330. [CrossRef] [PubMed]
8. Saadi, H.; Pagnier, I.; Colson, P.; Cherif, J.K.; Beji, M.; Boughalmi, M.; Azza, S.; Armstrong, N.; Robert, C.; Fournous, G.; et al. First isolation of Mimivirus in a patient with pneumonia. *Clin. Infect. Dis.* **2013**, *57*, e127–e134. [CrossRef] [PubMed]
9. Saadi, H.; Reteno, D.G.; Colson, P.; Aherfi, S.; Minodier, P.; Pagnier, I.; Raoult, D.; La Scola, B. Shan virus: A new mimivirus isolated from the stool of a Tunisian patient with pneumonia. *Intervirology* **2013**, *56*, 424–429. [CrossRef] [PubMed]

10. Boughalmi, M.; Pagnier, I.; Aherfi, S.; Colson, P.; Raoult, D.; La Scola, B. First isolation of a Marseillevirus in the Diptera Syrphidae *Eristalis tenax*. *Intervirology* **2013**, *56*, 386–394. [CrossRef] [PubMed]

11. Boughalmi, M.; Saadi, H.; Pagnier, I.; Colson, P.; Fournous, G.; Raoult, D.; La Scola, B. High-throughput isolation of giant viruses of the *Mimiviridae* and *Marseilleviridae* families in the Tunisian environment. *Environ. Microbiol.* **2013**, *15*, 2000–2007. [CrossRef] [PubMed]

12. Aherfi, S.; Boughalmi, M.; Pagnier, I.; Fournous, G.; La Scola, B.; Raoult, D.; Colson, P. Complete genome sequence of tunisvirus, a new member of the proposed family *Marseilleviridae*. *Arch. Virol.* **2014**, *159*, 2349–2358. [CrossRef] [PubMed]

13. Philippe, N.; Legendre, M.; Doutre, G.; Coute, Y.; Poirot, O.; Lescot, M.; Arslan, D.; Seltzer, V.; Bertaux, L.; Bruley, C.; et al. Pandoraviruses: Amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* **2013**, *341*, 281–286. [CrossRef] [PubMed]

14. Legendre, M.; Bartoli, J.; Shmakova, L.; Jeudy, S.; Labadie, K.; Adrait, A.; Lescot, M.; Poirot, O.; Bertaux, L.; Bruley, C.; et al. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 4274–4279. [CrossRef] [PubMed]

15. Legendre, M.; Lartigue, A.; Bertaux, L.; Jeudy, S.; Bartoli, J.; Lescot, M.; Alempica, J.M.; Ramusb, C.; Bruleyb, C.; Labadiee, K.; et al. In-depth study of *Mollivirus sibericum*, a new 30,000-y-old giant virus infecting *Acanthamoeba*. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5327–E5335. [CrossRef] [PubMed]

16. Colson, P.; De Lamballerie, X.; Yutin, N.; Asgari, S.; Bigot, Y.; Bideshi, D.K.; Cheng, X.W.; Federici, B.A.; Van Etten, J.L.; Koonin, E.V.; et al. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch. Virol.* **2013**, *158*, 2517–2521. [CrossRef] [PubMed]

17. Reteno, D.G.; Benamar, S.; Bou, K.J.; Andreani, J.; Armstrong, N.; Klose, T.; Rossmann, M.; Colson, P.; Raoult, D.; La Scola, B. Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *J. Virol.* **2015**, *89*, 6585–6589. [CrossRef] [PubMed]

18. Benamar, S.; Reteno, D.G.; Labas, N.; Raoult, D.; La Scola, B.; Bandaly, V.; Martinez-Salas, E.; Shuang-Yong, X. Faustoviruses: Comparative genomics of new *Megavirales* family members. *Front. Microbiol.* **2016**, *9*, 1–9. [CrossRef] [PubMed]

19. Khalil, J.Y.B.; Andreani, J.; La Scola, B. Updating strategies for isolating and discovering giant viruses. *Curr. Opin. Microbiol.* **2016**, *31*, 80–87. [CrossRef] [PubMed]

20. La Scola, B.; Campocasso, A.; N'dong, R.; Fournous, G.; Barrassi, L.; Flaudrops, C.; Raoult, D. Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. *Intervirology* **2010**, *53*, 344–353. [CrossRef] [PubMed]

21. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]

22. Nadalin, F.; Vezzi, F.; Policriti, A. GapFiller: A de novo assembly approach to fill the gap within paired reads. *BMC Bioinform.* **2011**, *13* (Suppl. S14), S8. [CrossRef] [PubMed]

23. Besemer, J.; Lomsadze, A.; Borodovsky, M. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **2001**, *29*, 2607–2618. [CrossRef] [PubMed]

24. Laslett, D.; Canback, B. ARAGORN, a program for the detection of transfer RNA and transfer-messenger RNA genes in nucleotide sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. [CrossRef] [PubMed]

25. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]

26. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef] [PubMed]

27. Kristensen, D.M.; Kannan, L.; Coleman, M.K.; Wolf, Y.I.; Sorokin, A.; Koonin, E.V.; Mushegian, A. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* **2010**, *26*, 1481–1487. [CrossRef] [PubMed]

28. Li, L.; Stoeckert, C.J.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189. [CrossRef] [PubMed]

29. Contreras-Moreira, B.; Vinuesa, P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* **2013**, *79*, 7696–7701. [CrossRef] [PubMed]

30. Delhon, G.; Tulman, E.R.; Afonso, C.L.; Lu, Z.; De la Concha-Bermejillo, A.; Lehmkuhl, H.D.; Piccone, M.E.; Kutish, G.F.; Rock, D.L. Genomes of the Parapoxviruses orf virus and bovine papular stomatitis virus. *J. Virol.* **2004**, *78*, 168–177. [CrossRef] [PubMed]

31. Klose, T.; Reteno, D.G.; Benamar, S.; Hollerbach, A.; Colson, P.; La Scola, B.; Rossmann, M.G. Structure of faustovirus, a large dsDNA virus. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 6206–6211. [CrossRef] [PubMed]