



Faustoviruses: Comparative Genomics of New *Megavirales* Family Members

Samia Benamar¹, Dorine G. I. Reteno¹, Victor Bandaly², Noémie Labas¹, Didier Raoult^{1,2} and Bernard La Scola^{1,2*}

¹ UM63 CNRS 7278 IRD 198 INSERM U1095, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, Facultés de Médecine et de Pharmacie, Aix-Marseille University, Marseille, France, ² Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Institut Hospitalo-Universitaire Méditerranée Infection, Centre Hospitalo-Universitaire Timone, Assistance Publique - Hôpitaux de Marseille, Marseille, France

OPEN ACCESS

Edited by:

Gilbert Greub,
University of Lausanne, Switzerland

Reviewed by:

Encarna Martínez-Salas,
Consejo Superior de Investigaciones
Científicas, Spain
Shuang-yong Xu,
New England Biolabs, USA

*Correspondence:

Bernard La Scola
bernard.lascola@univ-amu.fr

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 09 October 2015

Accepted: 05 January 2016

Published: 05 February 2016

Citation:

Benamar S, Reteno DGI, Bandaly V,
Labas N, Raoult D and La Scola B
(2016) Faustoviruses: Comparative
Genomics of New *Megavirales* Family
Members. *Front. Microbiol.* 7:3.
doi: 10.3389/fmicb.2016.00003

An emerging interest for the giant virus discovery process, genome sequencing and analysis has allowed an expansion of the number of known *Megavirales* members. Using the protist *Vermamoeba* sp. as cell support, a new giant virus named Faustovirus has been isolated. In this study, we describe the genome sequences of nine Faustoviruses and build a genomic comparison in order to have a comprehensive overview of genomic composition and diversity among this new virus family. The average sequence length of these viruses is 467,592.44 bp (ranging from 455,803 to 491,024 bp), making them the fourth largest *Megavirales* genome after Mimiviruses, Pandoraviruses, and *Pithovirus sibericum*. Faustovirus genomes displayed an average G+C content of 37.14 % (ranging from 36.22 to 39.59%) which is close to the G+C content range of the *Asfarviridae* genomes (38%). The proportion of best matches and the phylogenetic analysis suggest a shared origin with *Asfarviridae* without belonging to the same family. The core-gene-based phylogeny of Faustoviruses study has identified four lineages. These results were confirmed by the analysis of amino acids and COGs category distribution. The diversity of the gene composition of these lineages is mainly explained by gene deletion or acquisition and some exceptions for gene duplications. The high proportion of best matches from Bacteria and *Phycodnaviridae* on the pan-genome and unique genes may be explained by an interaction occurring after the separation of the lineages. The Faustovirus core-genome appears to consolidate the surrounding of 207 genes whereas the pan-genome is described as an open pan-genome, its enrichment via the discovery of new Faustoviruses is required to better seize all the genomic diversity of this family.

Keywords: Faustovirus, giant viruses, pan-genome, phylogenetic, genomic comparison

INTRODUCTION

Viruses infecting the Eukaryota cell including the *Phycodnaviridae*, *Poxviridae*, *Iridoviridae*, *Ascoviridae*, *Asfarviridae*, *Mimiviridae*, and *Marseilleviridae* families are representative members of the *Megavirales* order families (Colson et al., 2012, 2013). Since the discovery of the *Acanthamoeba polyphaga mimivirus* (La Scola et al., 2003; Raoult et al., 2004; Iyer et al., 2006; Raoult and Forterre, 2008; Yutin et al., 2009, 2014; Yutin and Koonin, 2012), the first described protist-associated virus,

other giant viruses have been nearly all isolated from *Acanthamoeba castellanii* and *A. polyphaga* (Pagnier et al., 2013; Philippe et al., 2013; Legendre et al., 2014; Antwerpen et al., 2015; Scheid, 2015), probably neglecting a wide part of the giant virus world. Recently, Reteno et al. (2015) using the protist *Vermamoeba* sp. as cell support, have isolated a new giant virus named Faustovirus. It was proposed to be the first member of the eighth family in the *Megavirales* order. The Faustoviruses were shown to share several features with *Asfarviridae*. Phylogenetic analyses indicate that the evolutionary distance between Faustovirus and Asfarviruses is comparable to that between Pandoraviruses and Phycodnaviruses (Reteno et al., 2015). In this study, we describe the genome sequences of nine Faustoviruses and build a genomics comparison in order to have a comprehensive overview of genomic composition and diversity among this new virus family. Aspects of the pan-genome were described and discussed. A phylogenetic analysis with all the core-genes allowed us to identify four lineages. We observed that despite the divergent features, Faustoviruses show a convergent adaptation to their intracellular life.

Faustovirus Isolates

Isolation of Faustovirus strain E12 was previously described in detail (Reteno et al., 2015). The other eight strains analyzed herein were isolated using the same strategy in sewage samples. Faustovirus E23, E24 and E9 were isolated from Marseille sewage samples, France. Faustovirus D3, D6, D5a and D5b were isolated from Dakar sewage samples, Senegal. Faustovirus Liban was isolated from Tripoli El Mina sea water sample, Lebanon.

Genome Sequencing, Data Assembly and General Features Prediction

Faustovirus E12 being available yet (Reteno et al., 2015), the other eight genomes were sequenced using MiSeq Technology (Illumina, Inc., San Diego, CA, USA) by 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 contigs using Mira 3.4. SSPACE software v1.0 combined to GapFiller were also used to enhance the assembly (Boetzer et al., 2011; Nadalin et al., 2011), remaining gaps were closed using Sanger sequencing of PCR products. Coding DNA Sequences (CDSs) were predicted using Prodigal (Hyatt et al., 2009). Data were submitted to the EMBL database and was assigned Bio-projects numbers (E24: PRJNA279158, D5a: PRJNA279159, D5b: PRJNA279160, D3: PRJNA279161, D6: PRJNA279164, Liban: PRJNA279165, E9: PRJNA279166, E23: PRJNA279157).

Comparative Genomics

A comparative analysis of the nine genomes was generated by creating a reference database of all the protein sequences. COGtriangles and OrthoMCL clustering algorithms were used to create protein clusters. Pangenome and core-genes was defined by using GET_HOMOLOGUES with the following parameters: 75% as minimum coverage, 30% as minimum identity in Blastp pairwise alignments and 1e-05 as maximum *E*-value.

Phylogenetic Tree Construction

Protein sequences were aligned using the MUSCLE program with default parameters; All the alignment were conserved even the columns containing a large fraction of gaps. A preliminary maximum-likelihood tree was constructed using the FastTree program with default parameters (JTT evolutionary model, discrete gamma model with 20 rate categories).

Genomics of Faustoviruses

The genome features of the nine Faustoviruses are summarized in **Table 1**. The average sequence length is 467,592 bp, making them the fourth largest *Megavirales* genome after Mimiviruses, Pandoraviruses and *P. sibericum* (Arslan et al., 2011; Philippe et al., 2013; Antwerpen et al., 2015; Scheid, 2015). These viruses have circular genomes except for Faustovirus Liban, for which genome could not be closed suggesting linear genome (**Figure 1**). Faustovirus D3 represents the strain with the smallest genome (with a size of 455,803 bp), while Faustovirus E9 possessed the largest genome (with a size of 491,024 bp). Faustovirus genomes displayed an average G+C content of 37.14% (ranging from 36.22 to 39.59%). No correlation was observed between the genome size and the GC content.

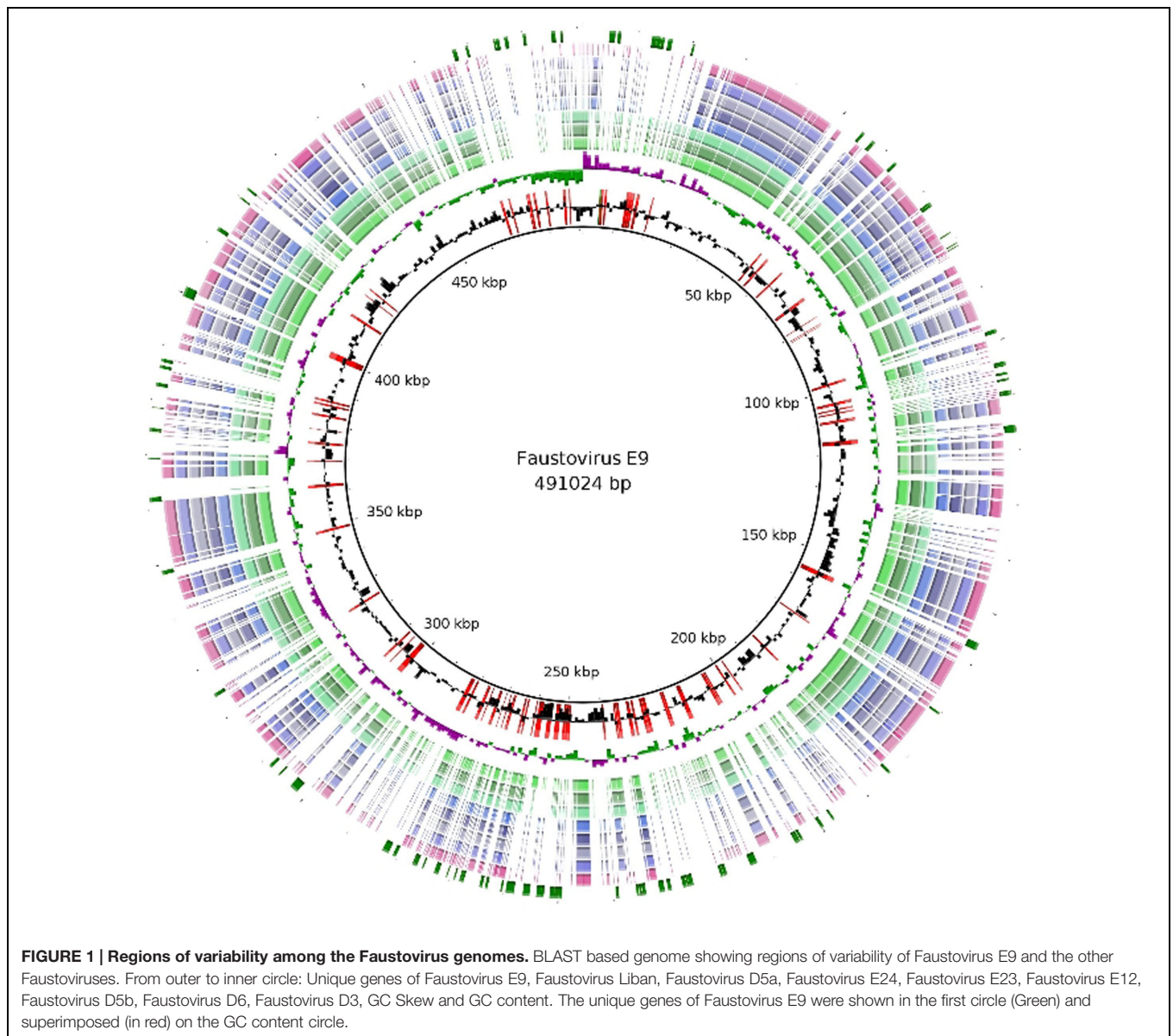
Pan-Genome and Core-Genome Analysis

In order to have a coherent comparative analysis, the same software for the prediction of the Open Reading Frame (ORF) was used. In this fashion, a comparable percent of DNA coding was obtained for each genome except for Faustovirus E12 (**Table 1**). In fact analysis of this last isolate benefited of proteomic study allowing deleting forty three ORFans (ORFs with no detectable homolog) and small genes (> 100 bp) whereas for other isolates only an *in silico* prediction of CDS was done.

A clustering analysis of the orthologous group of proteins based on COG triangles and OrthoMCL clustering algorithms, using the nine Faustovirus genomes, revealed the presence of 872 clusters. An overview of the Faustoviruses' pan-genome appears not to have been reached, as shown in **Figure 2A**. The average number of new genes added by each new genome was

TABLE 1 | General features of nine complete Faustovirus genomes.

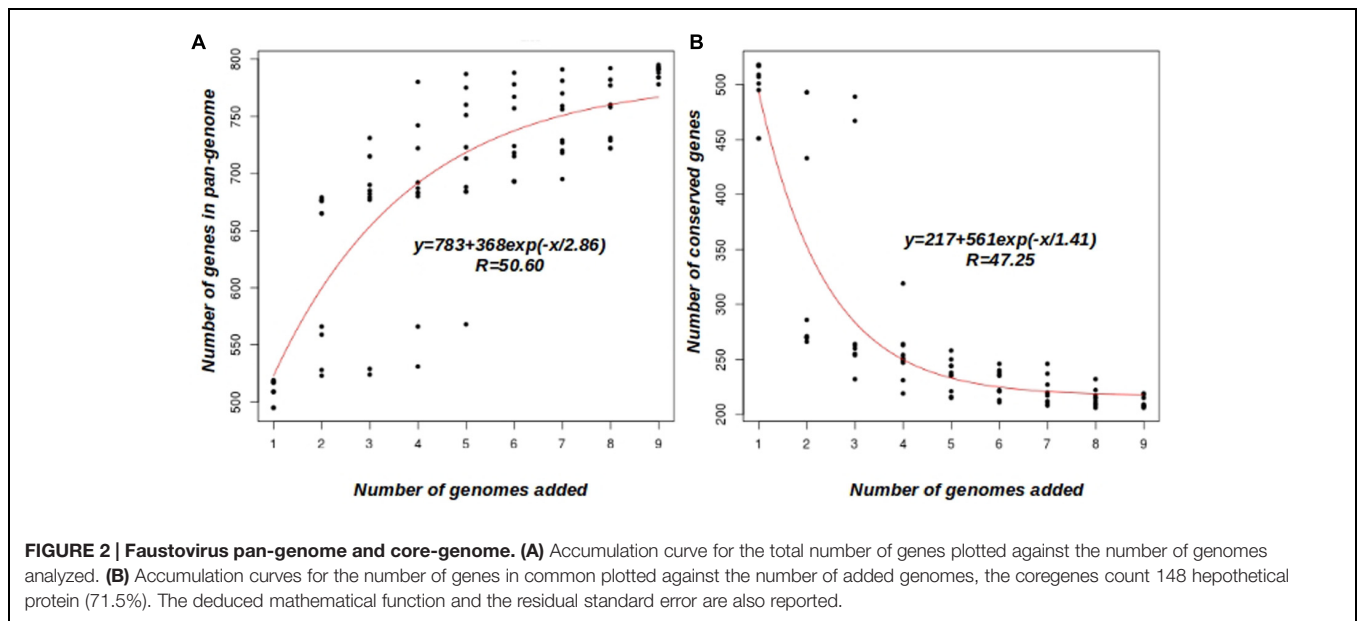
	Faustovirus E12	Faustovirus E23	Faustovirus E24	Faustovirus D5a	Faustovirus D5b	Faustovirus D3	Faustovirus D6	Faustovirus E9	Faustovirus Liban
Genome length (bp)	466,265	465,956	466,012	466,051	464,523	455,803	462,011	491,024	470,687
GC content (%)	36.22	36.22	36.22	36.21	37.66	37.76	37.67	39.59	36.74
Number of genes	457	519	518	517	507	495	509	511	518
Hypothetical proteins	303	457	456	454	433	430	438	449	467



30, and this number decreased to 19 after the addition of five genomes. According to the deduced mathematical function, the Faustovirus pan-genome probably surpasses 1000 genes; this finding suggests that the Faustoviruses' pan-genome is still open. On the other hand, the core-genome appears to reach a plateau at around 207 genes (Figure 2B). The curve indicates it will remain relatively constant, even if many more genomes are added. The number of core-genes in each genome is constant, except for Faustovirus Liban (five genes are in two copies per cluster), indicating no duplicated genes and an absence of paralogs. In the remaining clusters, 13 other duplications were observed (5 in Faustovirus E9 and 8 in Faustovirus Liban). These findings were confirmed by the Faustovirus genome pairwise dot plot (Supplementary Figure S1). Indeed, the dot plot showed a visible duplication in Faustovirus Liban. The set of 207 genes identified as core-genes comprises mainly orthologous groups of genes

(COGs) of the NCVOGs, which are present in at least two viral families (Yutin and Koonin, 2009; Yutin et al., 2009). Indeed, 78 core-genes (37.68% of the core-genes) have a significant blast result against the NCVOG database (E -value $< 10e-3$) and there are 148 hypothetical protein in the coregenes (71.5%).

The core-genome displays a larger number of genes with detectable homologs in the NCBI bank (Supplementary Figure S4) than the pan-genome (44.92 and 21.33% respectively). For almost half (53.76%) of the Faustovirus core-genes with detectable homologs, best matches were with genes from viruses (*Asfarviridae*, *Iridoviridae*, *Malacoherpesviridae*, *Marseilleviridae*, *Mimiviridae*, *Myoviridae*, and *Phycodnaviridae*). One hundred and fourteen genes (55.07% of the core-genes) are specific to the Faustoviruses and are not found in any other organism sequence, while 16 genes (7.72% of the core-genes) are specific to the Faustoviruses and the



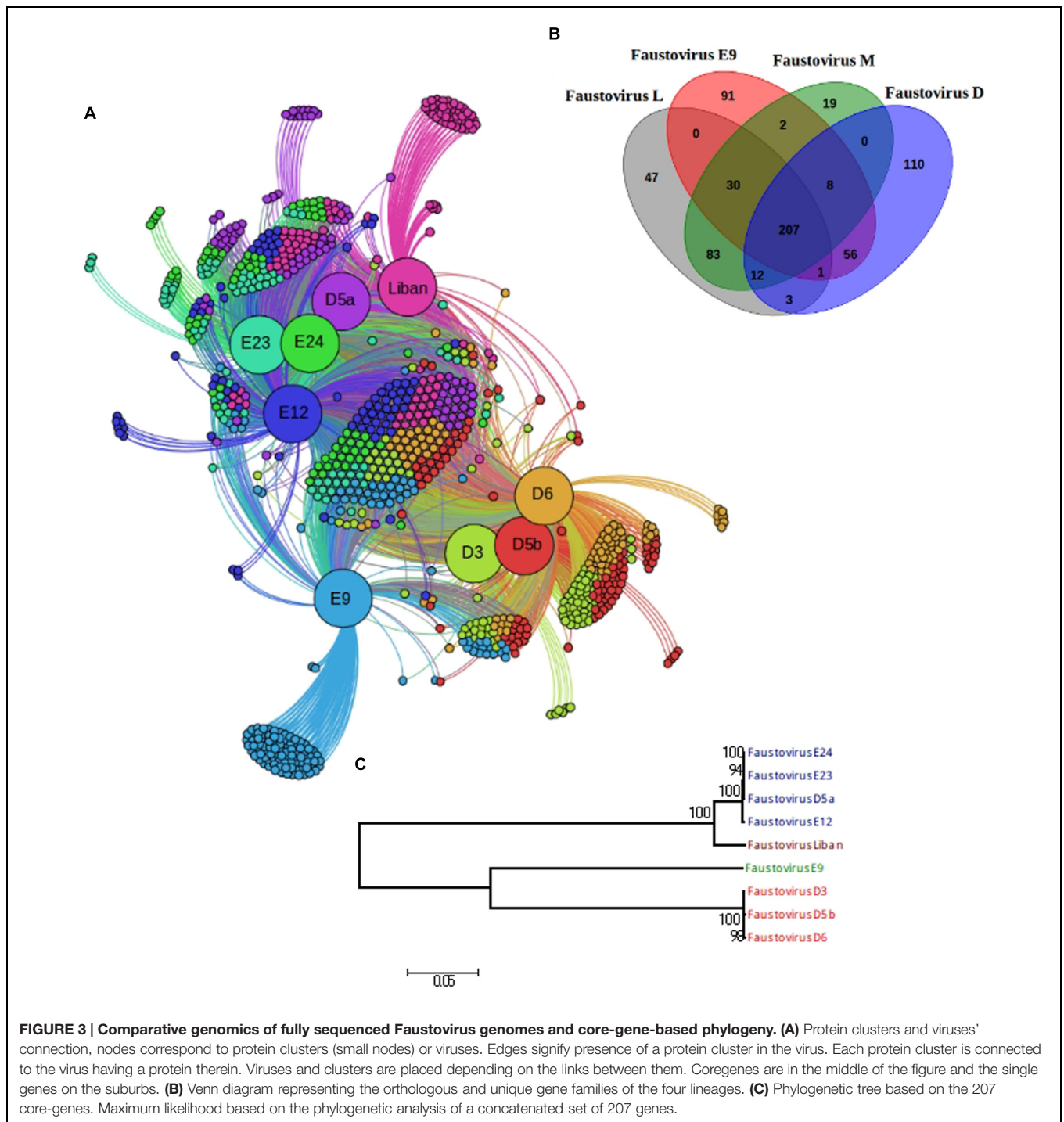
Asfarviridae (Supplementary Tables S1 and S2). The proportion of best matches from *Asfarviridae* is greater in the core-genome (74%) than in the pan-genome (61.25%), while a contrary effect was observed with the *Phycodnaviridae* more abundant in the pan-genome (20%) than in the core-genome (8%). No best match with genes from *Malacoherpesviridae* and *Marseilleviridae* was observed in the core-genome while 3 best matches were noticed in the pan-genome (1 from the *Malacoherpesviridae* and two from *Marseilleviridae*). On the whole, the pan-genome has fewer good matches with genes from viruses (43.01%) than the core-genome (Supplementary Figure S4). In contrast, we observed more best matches from Bacteria in the pan-genome (30.10%) than in the core-genome (22.58%). Best matches with genes from Eukaryota are in a similar proportion in the pan-genome (23.11%) and the core-genome (21.50%). Among the best matches from Eukaryota, five genes specific to *Viridiplantae* and several others found only in plants and Fungi were observed. Six best matches were with genes from *Phytophthora* (*Phytophthora parasitica*: 4, *Phytophthora sojae*: 1, *Phytophthora infestans* T30-4: 1) and were distributed equitably on the core and pan-genome (Supplementary Table S1). Seven best matches were with genes from Archaea including two with a known function (transcription factor S-II-related protein and putative DNA-directed RNA polymerase subunit D) identified in the core-genome.

Faustovirus Classification

In order to build a robust phylogenetic analysis we used the 207 core-genes, comprising in total 78,429 amino acid positions, to distinguish the different Faustovirus lineages (Figure 3C). The phylogenetic tree shows height robustness. Indeed, the nodes were supported by 100% of the bootstrap iterations, with a slightly lower supported proportion (94–98%) observed in two nodes. Based on this phylogenetic construction, four lineages were identified: lineage M comprises the founding

member Faustovirus E12 with E24, E23 and D5a; lineage D is formed by Faustovirus D3, D6 and D5b; lineage L contains only Faustovirus Liban and E9. Lineage D is exclusively formed of viruses identified in the same sewage environment (Senegal), like lineages L and E9 (Lebanon and French respectively). In contrast, lineage M is formed of three viruses isolated in France and one in Senegal.

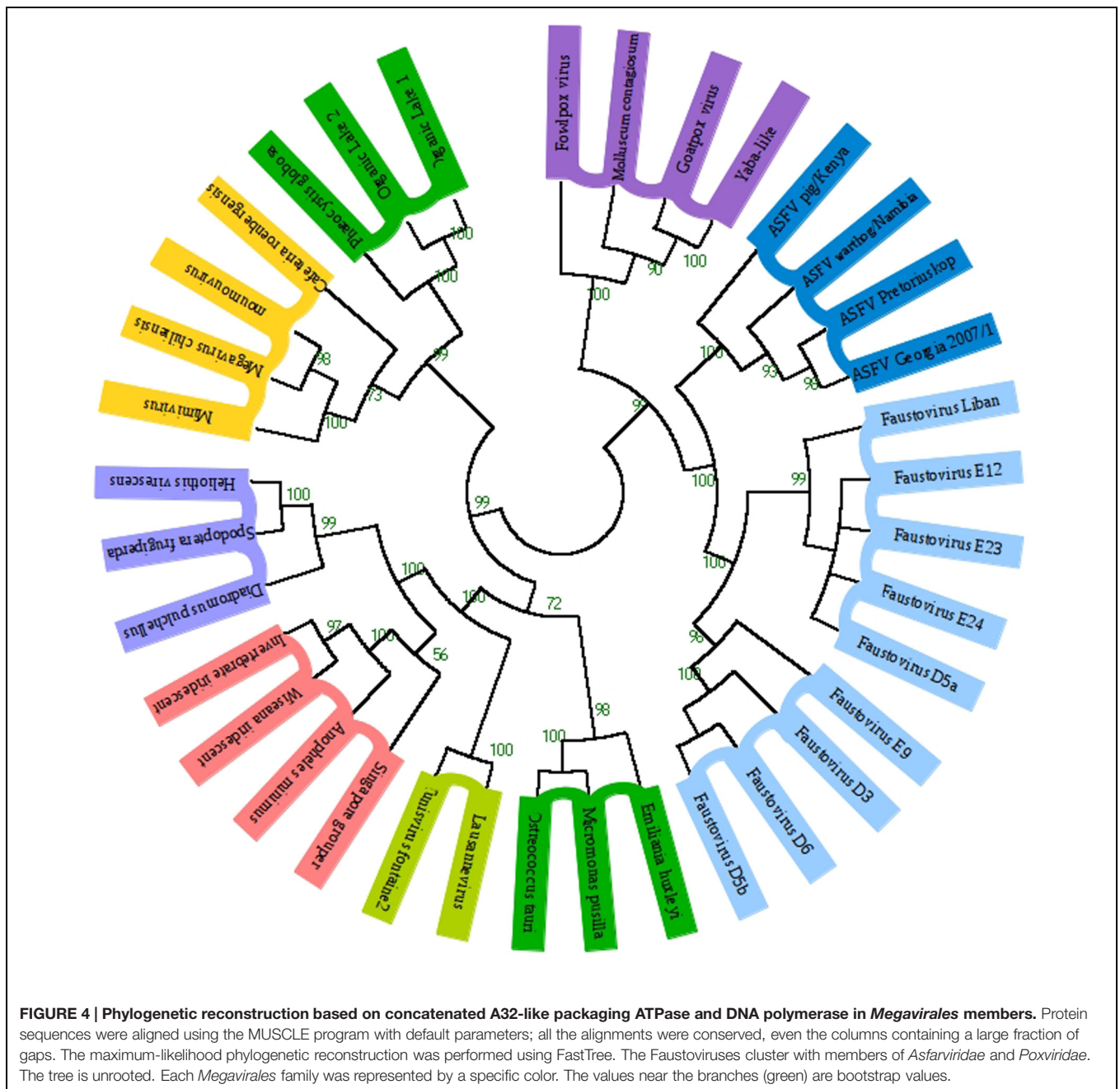
To determine which *Megavirales* family the Faustovirus is most closely related to, we concatenated the A32-like packaging ATPase and DNA polymerase of representative members of *Megavirales* families. Protein sequences were aligned using the MUSCLE program with default parameters. All the alignments were conserved, even the columns containing a large fraction of gaps. The phylogenetic tree maximum-likelihood phylogenetic reconstruction (Figure 4) shows that Faustovirus members are close but different from *Asfarviridae*. These findings are in correlation with the phylogenetic studies conducted by Reteno et al. (2015) This suggests that Faustoviruses represent a new family in the order of *Megavirales*. The gene repertoire of the common ancestor of Faustoviruses and *Asfarviridae* (named Lucasvirus) was determined by the selection of their core-genes to which we added the common genes of each one with 59 viruses of the *Megavirales* order. Over 40% of the *Asfarviridae* pan-genome takes part in the constitution of the Lucasvirus' repertoire against only 28.32% of the Faustovirus pan-genome. Supplementary Figure S5 shows the taxonomy of the best matches from the blastp result. We observed that the Lucasvirus and the *Asfarviridae* have fewer good matches with genes from Bacteria (10.97 and 14.86% respectively) than Faustoviruses (30.10%). On the other hand, the percentage of matches with genes from Eukaryota is much higher in Lucasvirus and *Asfarviridae* (40.54 and 40.24% respectively) than in the Faustovirus (23.11%). We observed that the percentage of matches with genes from Viruses and Archaea is almost identical in the three viruses.



Analysis of Amino Acids and COG Category Distribution

Amino acid composition is very similar among the studied Faustoviruses; however, we noted some differences between members from different lineages (Figure 5B). Indeed, we observed that the Faustoviruses of lineage D and E9 are more enriched in alanine, aspartic acid and arginine than other lineages (Supplementary Figure S2). On the other hand, E9

is less enriched in lysine and leucine than other viruses. The viruses of lineage M have a greater portion of asparagine, lysine, serine and valine than other lineages. Some amino acids are considerably more prevalent in all Faustoviruses such as two hydrophobic residues leucine and isoleucine (8.12 and 8.07% average composition respectively). In contrast, tryptophan, cysteine, histidine, methionine, and glutamine (1.04, 1.98, 2.25, 2.63, and 2.87% average composition respectively) are less



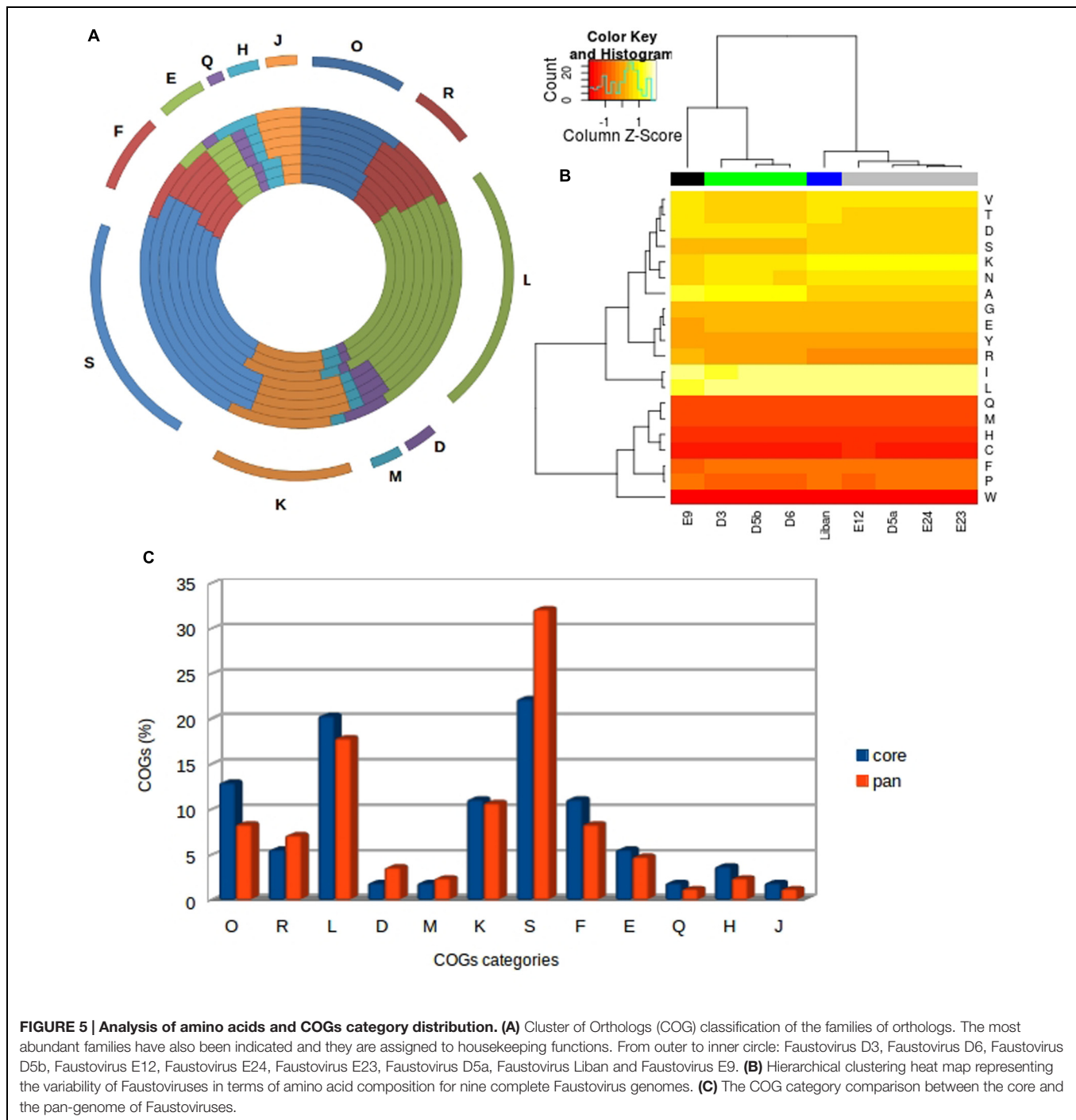
prevalent. Clustering based on the amino acid composition shows the same topology as the core-gene-based phylogeny.

An average of 15.21% of all Faustovirus genes has an ortholog identified by the Cluster of Orthologous Group (COG) classification (Figure 5A). Only 12 among the 25 COG categories were assigned to these genes and over half of them are predicted to be involved in two categories: L (replication, recombination and repair) and S (function unknown). The distribution of COG categories is different among the Faustovirus lineages with the exception of the Q category (secondary metabolite biosynthesis, transport and catabolism), represented by an amino-oxidase protein in all Faustovirus genomes. The

viruses of the same lineage have an identical category distribution excluding lineage D, which has a small difference with the L (DNA replication, recombination and repair) COG category. As observed in the amino acid composition, clustering based on COG category distribution shows the same topology as the core-Gene-based phylogeny (Supplementary Figure S3).

Faustovirus Unique Gene Set

The comparative analyses allowed the identification of an average of 66 unique genes per lineage, mostly represented by hypothetical protein. In fact, the COG classification showed that the majority of these genes have an unknown function. Lineage M



has the smallest number of unique genes; indeed, it shares over 94% of its genes with the other lineages. The viruses of lineage D, with their 110 unique genes (27.7% of their gene repertoire) show the greatest gene composition diversity. The number of unique genes identified in each virus is more reduced compared to that of the lineages (Figures 3A,B). The best matches of the lineages' unique genes with detectable homologs were mainly with genes from Bacteria (46.15%). The best matches with genes from viruses are represented by four families: *Marseilleviridae* (2.56%),

Mimiviridae (5.12%), *Asfarviridae* (2.56%), and *Phycodnaviridae* (12.82%), which has the greatest prevalence. Only unique genes of lineages D and E9 have a best match with genes from Archaea. To summarize, the unique genes are mostly composed by ORFans having no match (only 14% have a match while we observed almost 44.92% in the coregenes for example). The genes having a match, are composed by: hypothetical protein (79.48%), MORN variant repeat protein (7.69%), AAA family ATPase, DNA methyltransferase, HNH endonuclease, cell wall

anchor protein and the Phage integrase (Supplementary Table S3). The unique genes of the lineage D and E9 are composed only by hypothetical proteins and we found that they are richer in bacterial original genes compared to the coregenes and the pangenome.

CONCLUSION

An emerging interest for the giant virus discovery process, genome sequencing and analysis has allowed an expansion of the number of known *Megavirales* members (Boughalmi et al., 2013; Pagnier et al., 2013; Philippe et al., 2013; Saadi et al., 2013; Aherfi et al., 2014; Legendre et al., 2014; Antwerpen et al., 2015; Assis et al., 2015; Scheid, 2015). *Acanthamoeba castellanii* and *A. polyphaga* being the only cellular supports used to isolate these viruses, this approach may neglect a part of the giant virus world. Reteno et al. (2015) using another protist *Vermamoeba* sp. as cell support, have isolated a new giant virus: Faustovirus E12. In this study, we describe and analyze the genome sequences of nine Faustoviruses. The average sequence length of these viruses is 467,592.44 bp (minimum: 455,803 bp and maximum: 491,024 bp), making them the fourth largest *Megavirales* genome after Mimiviruses, Pandoraviruses and *P. sibericum* (Arslan et al., 2011; Philippe et al., 2013; Antwerpen et al., 2015; Scheid, 2015). Faustovirus genomes displayed an average G+C content of 37.14% (ranging from 36.22 to 39.59%), which is close to the G+C content range of the *Asfarviridae* genomes (38%) (Wan et al., 2013). The Faustoviruses' pan-genome is still open. Indeed, the clustering identified 872 protein clusters while the deduced mathematical function predicted that the Faustoviruses' pan-genome probably surpasses 1000 genes. On the other hand, the core-genome appears to consolidate the surrounding of 207 genes. The number of core-genes in each genome is constant, except for the duplication of five Faustovirus Liban genes. Research of duplicated genes in remaining clusters has highlighted five duplications in Faustovirus E9 and eight other duplications in Faustovirus Liban. There are no duplicated genes, which implies an absence of paralogs, for other Faustoviruses were observed. This finding suggests that the genomic diversity of Faustoviruses is essentially due to gene deletion or acquisition and timely to minor duplications. A comparison between the COG classification of the pan and core-genome showed that the last one contains a majority of genes involved in various housekeeping functions (Figure 5C). We identified especially those related to: (i) nucleotide, coenzyme and amino acid metabolism and associated transport activities; (ii) translation, ribosomal structure and biogenesis and transcription replication, recombination and repair. This is in agreement with the conserved developmental cycle and the intracellular lifestyle of these viruses. The core-genome displays a larger number of genes with detectable homologs than the pan-genome. The proportion of best matches from viruses in general and from *Asfarviridae* in particular is greater in the core-genome than in the pan-genome. On the basis of these

observations, we may assume that the Faustoviruses share a common origin with *Asfarviridae*. The phylogenetic analysis is simply a confirmation of that fact, indeed based on the phylogenetic tree the Faustoviruses are close to *Asfarviridae* but still too different to establish a same family. The matches with genes from Bacteria and Eukaryota were similar between Lucasvirus and *Asfarviridae* while we observed a significant gap between Lucasvirus and Faustovirus. These results suggest that Faustoviruses have a precocious divergence from Lucasvirus compared to *Asfarviridae*. The low percentage of Faustovirus pan-genomes included in the Lucasvirus gene repertoire indicates an enrichment of Faustovirus with genes from horizontal gene transfer (HGT).

We observed more best matches from Bacteria and *Phycodnaviridae* in the pan-genome than in the core-genome. Their proportion is significantly higher in the unique genes. This may be due to an interaction occurring after the separation of the lineages. The core-gene-based phylogeny of Faustoviruses study identified four lineages and these results were confirmed by the analysis of amino acids and COGs category distribution. Indeed, lineages M and L share a huge number of genes (9.86% of the pan-genome) as well as lineages E9 and D (6.42%) which refer to the two main clusters of the core-gene-based phylogeny. The reproducibility of the clustering in four lineages suggests their earlier separation. The study of the unique gene distribution shows greater prevalence when we focus on lineages than on an individual virus, suggesting that the inter-lineage diversity is much greater than the intra-lineage diversity.

In summary, whole genome sequencing and pan-genome analysis of Faustoviruses in this study revealed that the representative members of this family were separated into four lineages and the core-genome represents one quarter of their pan-genome. These findings point to a great diversity of gene composition mainly explained by gene deletion or acquisition and some exceptions for gene duplications (Faustovirus E9 and Faustovirus Liban). Regarding the separation on four lineages, we observed that the clustering was confirmed by several analyses; this would suggest their earlier separation. The high proportion of best matches from Bacteria and *Phycodnaviridae* on the pan-genome and unique genes may be due to an interaction occurring after the separation of the lineages. The Faustoviruses' pan-genome is described as an open pan-genome; its enrichment via the discovery of new Faustoviruses is required to better grasp the full genomic diversity of this family.

AUTHOR CONTRIBUTIONS

SB wrote the manuscript and did bioinformatic analysis, DGIR isolated Faustoviruses and did their presumptive identification, VB isolated Faustovirus, NL did genome sequencing, DR reviewed the manuscript and did several corrections, BL designed the study and wrote the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank the Institut Hospitalo-Universitaire Méditerranée Infection for their financial support in France.

REFERENCES

- Aherfi, S., La Scola, B., Pagnier, I., Raoult, D., and Colson, P. (2014). The expanding family *Marseilleviridae*. *Virology* 466–467, 27–37. doi: 10.1016/j.virol.2014.07.014
- Antwerpen, M. H., Georgi, E., Zoeller, L., Woelfel, R., Stoecker, K., and Scheid, P. (2015). Whole-genome sequencing of a *Pandoravirus* isolated from keratitis-inducing *Acanthamoeba*. *Genome Announc.* 3, e136–e115. doi: 10.1128/genomeA.00136-15
- Arslan, D., Legendre, M., Seltzer, V., Abergel, C., and Claverie, J.-M. (2011). Distant *Mimivirus* relative with a larger genome highlights the fundamental features of Megaviridae. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17486–17491. doi: 10.1073/pnas.1110889108
- Assis, F. L., Bajrai, L., Abrahao, J. S., Kroon, E. G., Dornas, F. P., Andrade, K. R., et al. (2015). Pan-genome analysis of Brazilian lineage A amoebal *Mimiviruses*. *Viruses* 2015, 3483–3499. doi: 10.3390/v7072782
- Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *J. Gerontol.* 27, 578–579.
- Boughalmi, M., Pagnier, I., Aherfi, S., Colson, P., Raoult, D., and La Scola, B. (2013). First isolation of a marseillevirus in the diptera syrphidae *eristalis tenax*. *Intervirology* 56, 386–394. doi: 10.1159/000354560
- Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D. K., et al. (2013). “Megavirales,” a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch. Virol.* 158, 2517–2521. doi: 10.1007/s00705-013-1768-6
- Colson, P., Lamballerie, X., Fournous, G., and Raoult, D. (2012). Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. *Intervirology* 55, 321–332. doi: 10.1159/000336562
- Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2009). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006). Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res.* 117, 156–184. doi: 10.1016/j.virusres.2006.01.009
- La Scola, B., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299, 2033. doi: 10.1126/science.1081867
- Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., et al. (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a *Pandoravirus* morphology. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4274–4279. doi: 10.1073/pnas.1320670111
- Nadalin, F., Vezzi, F., and Policriti, A. (2011). GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 13(Suppl. 14):S8. doi: 10.1186/1471-2105-13-S14-S8
- Pagnier, I., Reteno, D.-G. I., Saadi, H., Boughalmi, M., Gaia, M., Slimani, M., et al. (2013). A decade of improvements in Mimiviridae and Marseilleviridae isolation from amoeba. *Intervirology* 56, 354–363. doi: 10.1159/000354556
- Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., et al. (2013). *Pandoraviruses*: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341, 281–286. doi: 10.1126/science.1239181
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., et al. (2004). The 1.2-megabase genome sequence of mimivirus. *Science* 306, 1344–1350. doi: 10.1126/science.1101485
- Raoult, D., and Forterre, P. (2008). Redefining viruses: lessons from *Mimivirus*. *Nat. Rev. Microbiol.* 6, 315–319. doi: 10.1038/nrmicro1858
- Reteno, D. G., Benamar, S., Khalil, J. B., Andreani, J., Armstrong, N., Klose, T., et al. (2015). Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *J. Virol.* 89, 6585–6594. doi: 10.1128/JVI.00115-15
- Saadi, H., Pagnier, I., Colson, P., Cherif, J. K., Beji, M., Boughalmi, M., et al. (2013). First isolation of *Mimivirus* in a patient with pneumonia. *Clin. Infect. Dis.* 57, e127–e134. doi: 10.1093/cid/cit354
- Scheid, P. (2015). Viruses in close associations with free-living amoebae. *Parasitol. Res.* 114, 3959–3967. doi: 10.1007/s00436-015-4731-5
- Wan, X. F. J., Barnett, L., Cunningham, F., Chen, S., Yang, G., Nash, S., et al. (2013). Detection of African swine fever virus-like sequences in ponds in the Mississippi Delta through metagenomic sequencing. *Virus Genes* 46, 441–446. doi: 10.1007/s11262-013-0878-2
- Yutin, N., and Koonin, E. V. (2009). Evolution of DNA ligases of nucleocytoplasmic large DNA viruses of eukaryotes: a case of hidden complexity. *Biol. Direct* 4, 51. doi: 10.1186/1745-6150-4-51
- Yutin, N., and Koonin, E. V. (2012). Hidden evolutionary complexity of nucleocytoplasmic large DNA viruses of eukaryotes. *Virol. J.* 9, 161. doi: 10.1186/1743-422X-9-161
- Yutin, N., Wolf, Y. I., and Koonin, E. V. (2014). Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* 46, 38–52. doi: 10.1016/j.virol.2014.06.032
- Yutin, N., Wolf, Y. I., Raoult, D., and Koonin, E. V. (2009). Eukaryotic large nucleocytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* 17, 223. doi: 10.1186/1743-422X-6-223

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00003>

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Benamar, Reteno, Bandaly, Labas, Raoult and La Scola. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.