

Evolution of the genus *Mimivirus* based on translation protein homology and its implication in the tree of life

V.M. Marcelino, M.V.P.C. Espinola, V. Serrano-Solis and S.T. Farias

Laboratório de Genética Evolutiva Paulo Leminsk,
Departamento de Biologia Molecular, Universidade Federal da Paraíba,
João Pessoa, PB, Brasil

Corresponding authors: V.M. Marcelino / S.T. Farias
E-mail: victor-montenegro@hotmail.com / stfarias@yahoo.com.br

Genet. Mol. Res. 16 (3): gmr16039784

Received July 24, 2017

Accepted August 29, 2017

Published September 27, 2017

DOI <http://dx.doi.org/10.4238/gmr16039784>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. The natural history of mimiviruses (i.e., viruses that are members of the *Mimivirus* genus) is a challenge for modern biology. A new domain of life to include these organisms has been proposed from analysis of gene conservation. We analyzed the evolutionary relationship of proteins involved in the translation system, and our data show that mimiviruses are a sister group of Eukarya. New data about the origins of Eukarya, in which Eukarya appears as a branch derived from the Archaea domain, were discussed, and we suggest that the mimiviruses emerged from the initial population that gave origin to Eukarya and that, in this way are not part of a new domain of life.

Key words: Mimiviridae; *Mimivirus*; Evolution; Translation proteins

INTRODUCTION

Mimiviruses (i.e., viruses that are members of the *Mimivirus* genus) bring a new light to the studies of viruses, bringing up questions about their origin and evolution. Due to the fact that these viruses have proteins similar to those of Bacteria, Archaea, and Eukarya, they have been placed in the phylogenetic tree of life as a new branch or sometimes as a fourth domain of life (taking into account here the proteins that are unique to them) (Raoult and Forterre, 2008; Boyer et al., 2010; Forterre, 2010a,b; Colson et al., 2011; Clavarie, 2013; Sharma et al., 2014). This idea of placing them in the phylogenetic tree of life breaks the paradigm that viruses are not considered living beings and, as such, breaks the current definition of life, which is closely related to the definition of cellularity itself. This model received quite a lot of criticism because placing these viruses along with the other domains in the tree of life not only influences the overall classification of organisms but also reflects the traditional definition of viruses as “non-living beings”.

The viral family Mimiviridae is currently divided into three genera, with the genus *Mimivirus* being divided into three lineages: A (which includes *Acanthamoeba polyphaga mimivirus*, *Samba virus*, *Hirudovirus*, and *Terra2*), B (which includes *Moumouvirus*), and C (which includes megaviruses such as *Courdo7*, *Courdo II*, *Courdo 5*, *LBAIII*, *Megavirus chilensis*, and *Terra1*). The majority of Mimiviridae appears to be from the genus *Mimivirus*. The second subfamily is *Cafeteria* and includes the type species *Cafeteria roenbergensis* virus BV-PW1. The Klosneuvirinae have been proposed as a third genus and are divided into four lineages: *Klosneuvirus*, *Indivirus*, *Catovirus*, and *Hokovirus*. These lineages were discovered through analysis of low-complexity metagenomes from a wastewater treatment plant and apparently have all 20 AARS (aminoacyl-tRNA synthetase) normally present in living beings. Because they have some unique characteristics, it has been proposed that they should be classified as a new subfamily (provisionally denoted as Klosneuvirinae) in the family Mimiviridae (Schulz et al., 2017).

These organisms show unique features that place themselves at the edge of the cellular and the viral world (Raoult et al., 2004). Besides that, they do not go through many modifications on their genomes (Filée, 2015). Thus, it is most likely to find great levels of gene and protein preservation in their members and, as such, this makes developing concise phylogenies for this group very viable. Currently, there are two discussion fronts regarding the evolution of mimiviruses: one states that *Mimivirus* and other nucleocytoplasmic large DNA viruses (NCLDV) acquired their giant genome, as well as their genes that are similar to those of cells, due to multiple events of horizontal gene transfer (HGT) (Moreira and Brochier-Armanet, 2008; Williams et al., 2011), as viruses are already known to take part in such processes. The other hypothesis states that these giant viruses are originated from ancient cellular organisms that went through genome reduction and acquired a viral lifestyle, placing these viruses in a new branch of the tree of life. This idea is supported by genes that are similar to those found in the other three domains of life (Nasir et al., 2012).

Filée (2015) brings up a new hypothesis of the evolution of NCLDVs: the genomic accordion hypothesis, where giant viruses evolved through a balanced process of gene acquisition (gene duplication mostly, but also through HGT being compensated by gene loss). The hypothesis also states that the rate of the expansion of the genome has decreased before the diversification of Mimiviridae and, as such, the ancestors of these viral families were already giant viruses, indicating that the large size of the genome emerged early during the evolution of the NCLDVs (Filée, 2015).

The delineation of the fourth domain of life that supports the monophyly and common ancestry of these giant viruses was allowed based on phylogenetic and phyletic analysis characterized by the presence or absence of patterns of informational genes (e.g., DNA biosynthesis and processing, DNA replication and repair, recombination, and transcription) shared between Eukarya, Bacteria, Archaea, and the NCLDV (Boyer et al., 2010).

Because of these unique characteristics that differentiate mimiviruses from all other viruses, their large diversity, and since their fairly recent discovery, they have been receiving great attention from the scientific community. In this study, we try to clarify the existence of the fourth domain of in the life tree based on mimiviruses' translation proteins, as well as to clarify the phylogeny of the *Mimivirus* genus and their possible relationship with other living beings, as well as their possible position on the tree of life.

MATERIAL AND METHODS

Obtainment of proteins

In order to achieve this, we used eight translation proteins: Arg-tRNA synthetase (YP_003987185.1), Tyr-tRNA synthetase (YP_003986615.1), Cys-tRNA synthetase (YP_003986656.1), Met-tRNA synthetase (YP_003987159.1), elongation factor eF-Tu (YP_003987141.1), initiation factor SUI1 (YP_003986971.1), release factor eRF1 (YP_003987255.1), and tRNA (5-uracil-) methyltransferase (YP_003986909.1), because they have been shown to be remarkably conserved in the following organisms used in this study: *Acanthamoeba polyphaga mimivirus* (NC_014649.1), *Acanthamoeba castellanii mamavirus* strain Hal-V (JF801956.1), *Hirudovirus* strain Sangsue (KF493731.1), *Samba virus* (KF959826), *Terra2 virus* (NC_023639.1), *Mimivirus Bombay* (KU761889.1), *Mimivirus fauteuil* (LN871163.1), *Lentillevirus* (AFYC1000005.1), *Acanthamoeba polyphaga moumouvirus* (NC_020104.1), *Moumouvirus goulette* (KC008572.1), *Monve virus* (JN885994-JN886001), *Megavirus chiliensis* (NC_016072.1), *Megavirus lba* (JX975216.1), *Courdo 11 virus* (JX975216.1), *Courdo7 virus* (JN885990-JN885993), and *Terra1 virus* (NC_023640.1). These protein sequences were obtained from the online database of the National Center for Biotechnology Information (NCBI) and used the BLAST tool (Moreno-Hagelsieb and Latimer, 2008) to compare them to the homologous sequences of the three domains of life and those of other mimiviruses. For each protein, we obtained 100 sequences for each domain of life, plus the mimiviruses' sequences.

Alignment and best evolutionary model

For the alignment of these sequences, the software MAFFT v.7 software (Kato and Standley, 2016) was used. After that, we searched for the best evolutionary model using the ProtTest 3 software (Darriba et al., 2011) and we found that the best evolutionary model according to AIC was the LG+I+G model.

Phylogenetic trees

The phylogenetic trees were constructed by maximum likelihood using PhyML in the T-REX web server (Makarenkov, 2001; Boc et al., 2012), using the best evolutionary model estimated by ProtTest 3 (Darriba et al., 2011).

RESULTS AND DISCUSSION

The evolutionary history of NCLDV is a very significant challenge for modern biology. Gene comparisons with essential biological processes are good tool to analyze the evolutionary history of organisms. The use of molecules involved in translation processes has aided in the understanding of the biological diversity in the last 40 years.

The emergence of a translation system represents an important biological step during the origin and organization of life. All cellular organisms have a very complex translation system, this step being the last process and the heart of the informational flux in the cell.

Although viruses lack in their genomes ribosomes and protein sequences involved in the translation process, they can use the cellular ribosomes and all of the proteins of the cellular translational machinery; this characteristic may indicate a common origin of the information in the evolutionary history of life. These organisms present many proteins involved in processes that until the moment were observed only in cellular organisms, and among them are the translation proteins. The presence of translation proteins in NCLDVs opens new hypotheses about its origin and the phylogenetic relationship with other domains of life.

Here, we analyzed 8 proteins involved in the translation process that was conserved in *Mimivirus* genomes. The original trees are in Supplementary material (**Figures S1-S8**). In Figure 1, we present the results for AARS to arginine (A), cysteine (B), methionine (C), and tyrosine (D). The data for ArgRS, MetRS, and TyrRS showed *Mimivirus* as more related to Eukarya; with Bacteria as the external group. The data to CysRS showed *Mimivirus* as the most ancestral protein, but still more related to Eukarya. Boyer et al., (2010), analyzed the AARS of *Mimivirus* and proteins involved in different steps of DNA processing and showed that *Mimivirus* rooted with Eukarya. From these data, they suggested that *Mimivirus* are a clade distinct of Eukarya and that these results can support the existence of a fourth domain of life (Boyer et al., 2010).

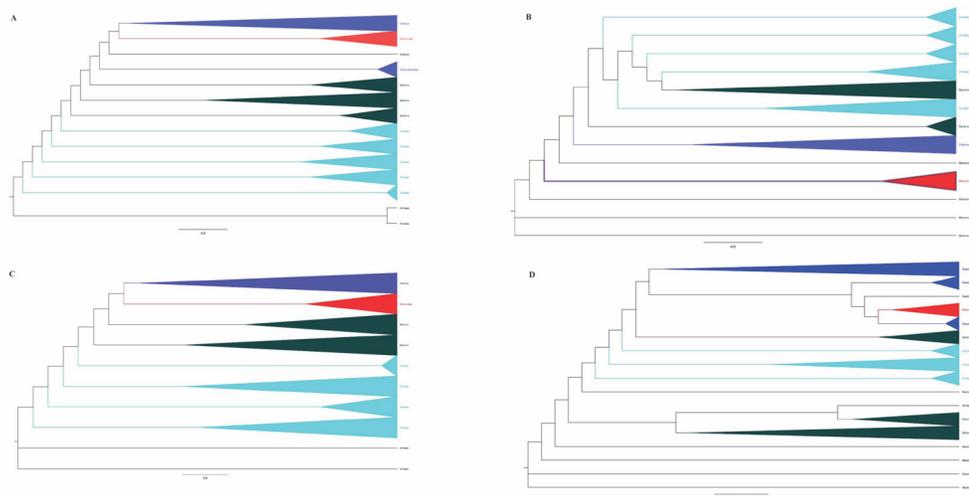


Figure 1. Collapsed phylogenetic tree to (A) arginyl-tRNA synthetase, (B) cysteinyl-tRNA synthetase, (C) methionyl-tRNA synthetase, and (D) tyrosyl-tRNA synthetase. Eukarya (blue), *Mimivirus* (red), Bacteria (green) and, Archaea (cyan).

In Figure 2, we present the phylogenetic analysis of elongation factor eF-TU, release factor eRF1, initial factor SUI1, and tRNA (5-uracil-) methyltransferase. The data showed that the elongation factor eF-TU, the release factor eRF1, and the initial factor SUI1 of *Mimivirus* are more related to Archaea and Eukarya, while tRNA (5-uracil-) methyltransferase is related to Eukarya, with Bacteria as the external group.

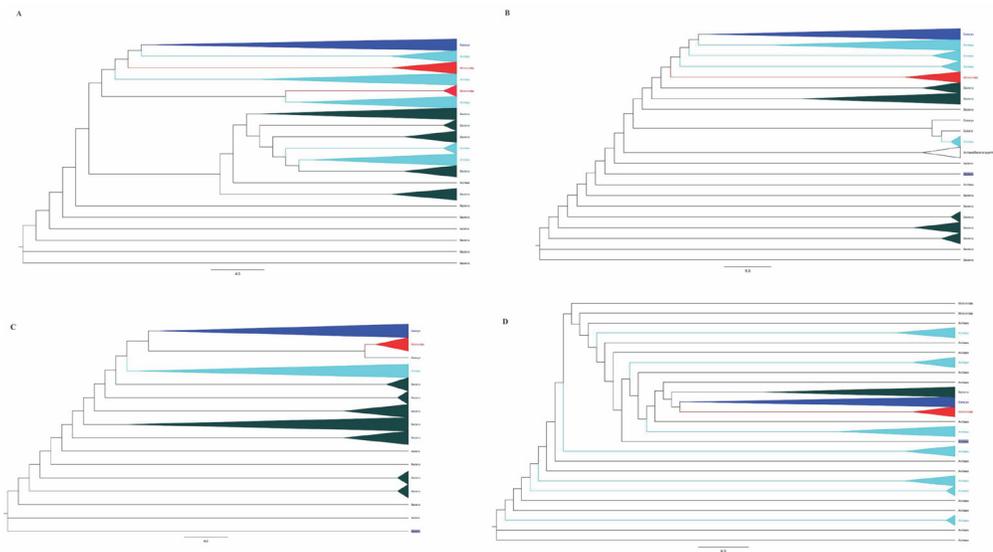


Figure 2. Collapsed phylogenetic tree to (A) elongation factor eF-TU, (B) release factor eRF1, (C) initial factor SUI1, and (D) tRNA (5-uracil-) methyltransferase. Eukarya (blue), *Mimivirus* (red), Bacteria (green), and Archaea (cyan).

According to Sharma et al. (2016), the NCLDVs, proposed as a new-order Megavirales (i.e., giant viruses) (Colson et al., 2013), which *Mimivirus* is part of, is monophyletic and is part of a new domain of life called TRUC (things resisting uncompleted classification) based on the phylogeny of RNA-dependent RNA polymerase (RNAP) and DNA polymerase. These genes, mainly RNAP subunits 1/2, represent valuable markers to classify new NCLDV members and uncharacterized microbes.

Nasir et al. (2016) argues that a chimerical origin of the Eukarya is very improbable, for there is a large difference between the Archaea, Eukarya, and Bacteria membranes that would require a transformation of an ether-bound lipid membrane in Archaea into an ester-bound lipid membrane in Bacteria and Archaea and there is no evolutionary reason for this. Besides, there is no genetic and morphological overlap between archeovirus and eukaryovirus (Nasir et al., 2016).

However, Raymann et al. (2015) investigated the relationship between Archaea and Eukaryotic organisms by applying a unique and alternative strategy for the use of Universal trees by using the model of evolution of complex proteins (Raymann et al., 2015). Their results strongly suggest that the appearance of Eukaryotes within the Archaea group is firmly connected to a root within Euryarchaeota. Furthermore, using a concatenated set of 16 ribosomal protein sequences from each 1011 genomes of organisms of all domains, Hug et al. (2016) constructed a tree of life with two domains, placing Eukaryotes branched within the Archaea domain,

specifically within the TACK superphylum and as a sister group to the Lokiarchaeota phylum (Hug et al., 2016). This approach classified those organisms based on the inheritance of their information systems as opposed to lipid or other cellular structures. Zaremba-Niedzwiedzka et al. (2017) suggest the new archaeal superphylum Asgard as ancestral to the first Eukarya population. Martin et al. (2015) reinforce the symbiotic origin to Eukarya from a host cell derived from an archaeal lineage and a bacterial endosymbiont.

As we can see from our results, all inner branches of *Mimivirus* and most of the clusters have good support with a high bootstrap value, generally ranging from values of about 70-100. In seven of the eight phylogenetic trees, the *Mimivirus* cluster diverges at the base of the Eukarya branch, showing a close relationship with this group.

Based on these results, the *Mimivirus* taxa can compose a sister group of Eukarya and, based on recent phylogeny studies regarding the position of the Eukarya domain in the tree of life (Hug et al. 2016), we, therefore, hypothesize that this family of viruses is, in fact, a new branch that can have derived from the primordial population that gave origin to the modern Eukarya, after the initial separation of the superphylum TACK and not indeed a new domain of life.

Conflicts of interest

All authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors thank CAPES for supporting V. Serrano-Solis and CNPq for supporting M.V.P.C Espinola.

REFERENCES

- Boc A, Diallo AB and Makarenkov V (2012). T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* 40: W573-9. <https://doi.org/10.1093/nar/gks485>
- Boyer M, Madoui MA, Gimenez G, La Scola B, et al. (2010). Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses. *PLoS One* 5: e15530. <https://doi.org/10.1371/journal.pone.0015530>
- Clavarié JM (2013). Giant virus in the sea: Extending the realm of Megaviridae to Viridiplantae. *Commun. Integr. Biol.* 6: e25685. <https://doi.org/10.4161/cib.25685>
- Colson P, Gimenez G, Boyer M, Fournous G, et al. (2011). The giant Cafeteria roenbergensis virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. *PLoS One* 6: e18935. <https://doi.org/10.1371/journal.pone.0018935>
- Colson P, De Lamballerie X, Yutin N, Asgari S, et al. (2013). “Megavirales”, a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch. Virol.* 158: 2517-2521. <https://doi.org/10.1007/s00705-013-1768-6>
- Darriba D, Taboada GL, Doallo R and Posada D (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164-1165. <https://doi.org/10.1093/bioinformatics/btr088>
- Filée J (2015). Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution. *Front. Microbiol.* 6: 593.
- Forterre P (2010a). Giant viruses: conflicts in revisiting the virus concept. *Intervirology* 53: 362-378. <https://doi.org/10.1159/000312921>
- Forterre P (2010b). Defining life: the virus viewpoint. *Orig. Life Evol. Biosph.* 40: 151-160. <https://doi.org/10.1007/s11084-010-9194-1>
- Hug LA, Baker BJ, Anantharaman K, Brown CT, et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1: 16048. <https://doi.org/10.1038/nmicrobiol.2016.48>
- Katoh K and Standley DM (2016). A simple method to control over-alignment in the MAFFT multiple sequence alignment

- program. *Bioinformatics* 32: 1933-1942. <https://doi.org/10.1093/bioinformatics/btw108>
- Makarencov V (2001). T-REX: reconstructing and visualizing phylogenetic trees and reticulation networks. *Bioinformatics* 17: 664-668. <https://doi.org/10.1093/bioinformatics/17.7.664>
- Martin WF, Garg S and Zimorski V (2015). Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370: 20140330. <https://doi.org/10.1098/rstb.2014.0330>
- Moreira D and Brochier-Armanet C (2008). Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. *BMC Evol. Biol.* 8: 12. <https://doi.org/10.1186/1471-2148-8-12>
- Moreno-Hagelsieb G and Latimer K (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24: 319-324. <https://doi.org/10.1093/bioinformatics/btm585>
- Nasir A, Kim KM and Caetano-Anolles G (2012). Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. *BMC Evol. Biol.* 12: 156. <https://doi.org/10.1186/1471-2148-12-156>
- Nasir A, Kim KM, Da Cunha V and Caetano-Anollés G (2016). Arguments Reinforcing the Three-Domain View of Diversified Cellular Life. *Archaea* 2016: 1851865. <https://doi.org/10.1155/2016/1851865>
- Raoult D and Forterre P (2008). Redefining viruses: lessons from Mimivirus. *Nat. Rev. Microbiol.* 6: 315-319. <https://doi.org/10.1038/nrmicro1858>
- Raoult D, Audic S, Robert C, Abergel C, et al. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science* 306: 1344-1350. <https://doi.org/10.1126/science.1101485>
- Raymann K, Brochier-Armanet C and Gribaldo S (2015). The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl. Acad. Sci. USA* 112: 6670-6675. <https://doi.org/10.1073/pnas.1420858112>
- Schulz F, Yutin N, Ivanova NN, Ortega DR, et al. (2017). Giant viruses with an expanded complement of translation system components. *Science* 356: 82-85. <https://doi.org/10.1126/science.aal4657>
- Sharma V, Colson P, Giorgi R, Pontarotti P, et al. (2014). DNA-dependent RNA polymerase detects hidden giant viruses in published databanks. *Genome Biol. Evol.* 6: 1603-1610. <https://doi.org/10.1093/gbe/evu128>
- Sharma V, Colson P, Pontarotti P and Raoult D (2016). Mimivirus inaugurated in the 21st century the beginning of a reclassification of viruses. *Curr. Opin. Microbiol.* 31: 16-24. <https://doi.org/10.1016/j.mib.2015.12.010>
- Williams TA, Embley TM and Heinz E (2011). Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. *PLoS One* 6: e21080. <https://doi.org/10.1371/journal.pone.0021080>
- Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541: 353-358. <https://doi.org/10.1038/nature21031>

Supplementary material

- [Figure S1](#). The original tree to Arginyl-tRNA synthetase for the three domains plus *Mimivirus*.
- [Figure S2](#). The original tree to Cisteinyl-tRNA synthetase for the three domains plus *Mimivirus*.
- [Figure S3](#). The original tree to Initiation factor SUI1 for the three domains plus *Mimivirus*.
- [Figure S4](#). The original tree to tRNA (5-uracil-) methylase for the three domains plus *Mimivirus*.
- [Figure S5](#). The original tree to Elongation factor eF-Tu for the three domains plus *Mimivirus*.
- [Figure S6](#). The original tree to Methionyl-tRNA synthetase for the three domains plus *Mimivirus*.
- [Figure S7](#). The original tree to Release factor eRF1 u for the three domains plus *Mimivirus*.
- [Figure S8](#). The original tree to Tyrosyl-tRNA synthetase for the three domains plus *Mimivirus*.