

The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses

Jonatas Santos Abrahao, Rodrigo Araujo, Philippe Colson, Bernard La Scola

▶ To cite this version:

Jonatas Santos Abrahao, Rodrigo Araujo, Philippe Colson, Bernard La Scola. The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. PLoS Genetics, 2017, 13 (2), 10.1371/journal.pgen.1006532. hal-01496184

HAL Id: hal-01496184

https://hal.science/hal-01496184

Submitted on 7 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



REVIEW

The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses

Jônatas Santos Abrahão^{1,2©}, Rodrigo Araújo^{2©}, Philippe Colson¹, Bernard La Scola¹*

- 1 Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE) UM63 CNRS 7278 IRD 198 INSERM U1095, Aix-Marseille Univ., 27 boulevard Jean Moulin, Faculté de Médecine, Marseille, France, 2 Instituto de Ciências Biológicas, Departamento de Microbiologia, Laboratório de Vírus, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
- These authors contributed equally to this work.
- * bernard.la-scola@univ-amu.fr

Abstract

The giant mimiviruses challenged the well-established concept of viruses, blurring the roots of the tree of life, mainly due to their genetic content. Along with other nucleo-cytoplasmic large DNA viruses, they compose a new proposed order—named *Megavirales*—whose origin and evolution generate heated debate in the scientific community. The presence of an arsenal of genes not widespread in the virosphere related to important steps of the translational process, including transfer RNAs, aminoacyl-tRNA synthetases, and translation factors for peptide synthesis, constitutes an important element of this debate. In this review, we highlight the main findings to date about the translational machinery of the mimiviruses and compare their distribution along the distinct members of the family *Mimiviridae*. Furthermore, we discuss how the presence and/or absence of the translation-related genes among mimiviruses raises important insights to boost the debate on their origin and evolutionary history.



OPEN ACCESS

Citation: Abrahão JS, Araújo R, Colson P, La Scola B (2017) The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. PLoS Genet 13(2): e1006532. doi:10.1371/journal.pgen.1006532

Editor: Jonathan Filée, CNRS, FRANCE

Published: February 16, 2017

Copyright: © 2017 Abrahão et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors received funding from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), CAPES-COFECUB, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and the Mediterranean Foundation Infection (FMI). JSA is a CNPq researcher. The funders had no role in the preparation of the article.

Competing interests: The authors have declared that no competing interests exist.

Introduction

In 1957, a time when genetics and cellular biology were in their preliminary stages, André Lwoff proposed a modern concept of viruses based on a set of features that, directly or indirectly, emphasized that viruses are defined by "negative plesiomorphic or apomorphic nonnatural characteristics" [1]. In the following years, many research fields evolved—including molecular biology and virology—and several different viruses were discovered, presenting some characteristics which had never been seen before among living organisms. However, even with such progress, most of the features raised by Lwoff have still been well supported if we consider the universe of viral species catalogued by the International Committee on Taxonomy of Viruses (ICTV) [2].

However, the discovery of the giant viruses blurred some of those well-established concepts, surprising the scientific community by their size and genetic content [3]. Although they still fit in some of Lwoff's proposed non-natural features (the same is also true for some intracellular bacteria), giant viruses present an astonishing arsenal of genes not widespread in the



virosphere, some of them related to important steps of the translational process, including transfer RNAs (tRNAs), aminoacyl–tRNA synthetases (aaRS), and translation factors for peptide synthesis [4–6]. Recent studies have shown that some of those genes can be related to the improvement of viral fitness, despite the presence of related genes in their hosts' genomes [7]. The analysis of these intriguing translation-related gene sets has raised interesting theories about the "lifestyle" of giant viruses' ancestors. In this review, we highlight the main features about the diversity, function, and putative origin of mimivirus translation-related genes.

Mimiviridae: A giant virus family with exceptional genetic content

The first giant virus was isolated from a water sample of a cooling tower in Bradford, England, replicating in the protist host *Acanthamoeba polyphaga*. At the time of its isolation, the new microorganism was considered to be a gram-positive bacterium, and its viral nature was only established after transmission electron microscopy analysis, which led to its label *Acanthamoeba polyphaga mimivirus* (APMV) [8]. Since then, new mimivirus-like viruses have been isolated in different parts of the world. These viruses constitute the new family *Mimiviridae*, which was included in the recently proposed order *Megavirales* [9,10].

The mimiviruses present several unusual features, both genetic and structural (for details about the viral structure, see [11]). The genome of these viruses consists of a single linear dsDNA molecule, is A+T rich (reaching up to 1,259 Kb), and may have approximately 1,000 open reading frames with a coding density higher than 90% [4,12]. The mimiviruses are phylogenetically divided into two groups (I and II), the first, which comprises mimiviruses that infect *Acanthamoeba*, being subdivided into three lineages (A, B, and C). Lineage A comprises APMV [8], Mamavirus [13], Samba virus [14], Niemeyer virus [15], and many others. Lineage B is mainly represented by Moumouvirus [16] and lineage C by Megavirus chilensis (MCV) [6], Courdo11 virus [17], and LBA111 virus [18]. Group II comprises smaller *Mimiviridae* members distantly related to APMV and is represented by *Cafeteria roenbergensis* virus (CroV) [5] and some algae viruses, including *Phaeocystis globosa* virus (PgV) [19] and Organic Lake Phycodnaviruses (OLV) [20].

The genome of mimiviruses is impressive not only for its size but also (and mainly) for its genetic content, presenting many genes which have never previously been described for other viruses. The mimiviruses possess many genes codifying DNA repair enzymes and are the first viruses to code for topoisomerase type IA [4]. Moreover, mimiviruses have their own glycosylation apparatus, presenting glycosyltransferases that are involved in the biosynthesis of glycans and post-translational protein modifications [21,22]. Furthermore, and even more impressive, is the presence of genes related to the protein synthesis, such as aaRS, tRNAs, and translation factors, which are present in different amounts in several representatives of the family *Mimiviridae* (Table 1). Other giant viruses, such as Marseillevirus [23], Pandoravirus [24], Faustovirus [25], and Mollivirus [26], also have some of these components, but in much less abundance compared to the mimiviruses.

Aminoacyl-tRNA synthetases among viruses—Breaking barriers

The aaRS are key enzymes in gene translation, during which they catalyse the esterification of a specific amino acid to the 3'-end of its cognate tRNA, forming the aminoacyl-tRNAs [27,28]. There are 20 different aaRS, which are divided into two families named class I and class II [28]. The aaRS are present in a wide variety of different organisms from all domains of life and, until very recently, there were no descriptions of these enzymes in a virus, and they were thus considered trademarks of cellular organisms [29]. However, with the discovery of the mimi-viruses, this scenario has changed.



 Table 1. Giant viruses' translation-related genes.
 Representative isolates of each group or family.

Group/Viruses	Aminoacyl-tRNA synthetase	tRNA	Translation Factors
Mimivirus Lineage A			
APMV	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Mamavirus	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Lentille	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, eF-TU, eRF1
Hirudovirus	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
SMBV	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
OYTV	ArgRS (2x), CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
KROV	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine	IF4A, IF4E, eF-TU, eRF1
AMAV	CysRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
NYMV	ArgRS, CysRS (2x), MetRS (2x), TyrRS (2x)	Leucine (2x), Histidine, Cysteine	IF4A, IF4E, SUI1, eF-TU, eRF1
Terra2	ArgRS, CysRS, MetRS, TyrRS	Leucine (2x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Bombay	ArgRS, CysRS, MetRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU eRF1
Mimivirus Lineage B			
APMOUV	ArgRS (4x), CysRS, IleRS, MetRS, TyrRS	Leucine, Histidine, Cysteine	IF4E, SUI1, eF-TU, eRF1
Goulette	CysRS, MetRS	Leucine (3x), Histidine, Cysteine	IF4E, SUI1, eF-TU, eRF1
Monve	ArgRS (2x), AsnRS, CysRS, IleRS (2x), MetRS, TyrRS	Leucine, Histidine, Cysteine	IF4A, IF4E (2x), SUI1, eRF1
Mimivirus Lineage C			
MCV	ArgRS, AsnRS, CysRS, IleRS, MetRS, TrpRS, TyrRS	Leucine (2x), Tryptophan	IF4A, IF4E, SUI1, eF-TU eRF1
Terra1	ArgRS, CysRS, MetRS, TyrRS	Leucine, Tryptophan	IF4A, IF4E, SUI1, eF-TU eRF1
LBA111	ArgRS, AsnRS, CysRS, IleRS, MetRS, TrpRS, TyrRS	Leucine (2x), Histidine, Cysteine, Tryptophan	IF4A, IF4E, SUI1, eF-TU, eRF1
Courdo7	lleRS, TyrRS	Leucine (3x), Tryptophan	IF4A (2x), IF4E, SUI1, eRF1
Courdo11	ArgRS, AsnRS (2x), CysRS, IleRS, MetRS, TrpRS, TyrRS	Leucine (3x), Histidine, Cysteine, Tryptophan	IF4A (2x), IF4E, SUI1, eRF1
Mimivirus group II			
CroV	IIeRS	Leucine (9x), Serine (5x), Tyrosine, Asparagine, Lysine	IF4A, IF4E, SUI1
PgV	-	Leucine (3x), Asparagine (2x), Isoleucine, Arginine, Glutamine	IF4E
OLV	-	Leucine, Isoleucine, Tyrosine, Asparagine, Arginine	IF4E
Other giant viruses			
Marseillevirus	-	-	elF5, SUI1, EF1α, eRF1
Faustovirus E12	-	-	SUI1
Pandoravirus salinus	TyrRS, TrpRS	Proline, Methionine, Tryptophan	IF4E
Pandoravirus dulcis	TyrRS	Proline	IF4E
Pandoravirus inopinatum	-	Proline	IF4E
Mollivirus sibericum	-	Leucine, Methionine, Tyrosine	IF4E

doi:10.1371/journal.pgen.1006532.t001



A total of four aaRS were found in the APMV genome (Arginyl-RS, Cysteinyl-RS, Methionyl-RS, and Tyrosyl-RS), all of them classified as class I aaRS [4]. In the following years, other mimiviruses were discovered, and the number of mimiviral aaRS expanded. Moumouvirus presents three out of four aaRS described in APMV (ArgRS, CysRS, and TyrRS), plus two others (Asparaginyl-RS [class II] and Isoleucyl-RS [class I]) [16]; and remarkably, MCV presents all the aaRS found in APMV and Moumouvirus, plus another one (Tryptophanyl-RS) [6]. Therefore, to our knowledge, MCV displays the most diverse set of aaRS, i.e., seven different aaRS. It is noteworthy that several paralogs of aaRS can be found in the genome of some mimiviruses. Moumouvirus has four copies of ArgRS and, for this reason, it is the virus with the highest abundance of aaRS (eight); and Niemeyer virus, a new lineage A mimivirus, presents three sets of aaRS duplication (MetRS, CysRS, and TyrRS are duplicated) [15,16]. Regarding CroV (mimivirus group II), only IleRS is present, which is also found in mimiviruses of lineages B and C [5].

The distribution of aaRS follows a clear pattern of diversity in family *Mimiviridae*: viral genomes with a wealthy aaRS gene set (e.g., MCV) contain all types of aaRS found in the rest of the family [6,13,15,16]. This suggests gradual gene loss throughout the evolution of different mimivirus groups (Fig 1A). The evolutionary pressures related to the conservation or loss of aaRS might be linked to specific environmental pressures to which each mimivirus group and/ or lineage was submitted after mimiviral species radiation. This scenario would be in accordance with previous phylogenetic and phylogenomic works, which suggested that giant viruses originated from a more complex organism and evolved by genomic reduction [30,31]. However, other studies suggested that some mimivirus aaRS were acquired by horizontal gene transfer (HGT) [32,33]. From this perspective, the mimiviruses should have originated from smaller organisms and evolved mainly by HGT events, being considered "gene pickpockets" [32,33,61] (Fig 2). It is important to highlight, however, that the phylogeny of aaRS is quite complex and sometimes violates the expected pattern of canonical domains of life [29]. The methods of tree construction, alignment, and hits sampling could explain conflicting results observed in different studies (S1-S7 Files). Nevertheless, in Fig 2, we present eight mimivirus aaRS-based trees constructed by maximum likelihood method (very similar results were observed for trees constructed by the neighbor joining method). Considering about 100 of the best hits obtained in GenBank related to each aaRS, we observed that phylogenic reconstructions suggested that all mimiviral genes but TyrRS clustered together as an independent group, with bootstrap values >90 both for viral and cellular taxa in most of the trees (Fig 2). However, both theories regarding mimivirus origin—that mimiviruses either originated from a more complex ancestor or that they originated from a simple ancestor—are plausible, although increasing evidence points to the former hypothesis [4,10,30,31,34].

The analysis of APMV transcriptome revealed that its four aaRS are expressed during the replication cycle of the virus [35]. Furthermore, experimental data have demonstrated that some mimiviruses' aaRS are indeed functional [4,36,37]. Protein structural and functional studies confirmed that APMV, MetRS, and TyrRS act as genuine enzymes. It was shown that TyrRS is a homodimer similar to other class I aaRS described so far, but in contrast to what is found in cellular organisms, the viral TyrRS seems to recognize only two bases in tRNA [37]. Although mimivirus amoebal host encodes aaRS, the conservation of aaRS in the mimiviruses' genomes seems to be associated with an increased viral fitness [38]. The expression of APMV aaRS transcripts can be modulated according to the nutritional status of its host: if mimivirus infection takes place in amoebas cultivated on starvation conditions, a higher mRNA expression of aaRS transcripts is observed. This indicates interplay between nutrient availability sensing of amoeba and the stimulation of the mimivirus aaRS genes as a mechanism related to the circumvention of starvation and maintenance of viral replication in usual levels.

It is intriguing that, even if mimiviruses' aaRS seem to be true proteins and related to the improvement of viral fitness, there is a trend of aaRS repertoire loss in the taxon. An accordion-like evolution model was recently proposed, suggesting such a pathway as the natural history of many mimivirus genes [39]. Indeed, it is possible to see the gain and loss of a given aaRS and tRNAs (e.g., ArgRS and Leu tRNA), but an overview of aaRS and tRNAs through mimiviruses suggests reduction of aaRS and tRNA classes' intra- and inter-lineages (Fig 1A). Considering that the occurrence of sequential HGT events involving mimiviruses' aaRS seems uncertain, this scenario may indicate that the mimiviruses' ancestor presented a more complete set of aaRS. Also, we can hypothesize that the loss of those genes might be a consequence of viral genome reduction and specialization to a given cell environment and to a more restricted host-range lifestyle (*Acanthamoeba*) (Fig 2). It is noteworthy that this gradual gene loss was evidenced and accelerated experimentally, causing the loss of TrpRS (and other genes) in APMV [40]. Another important piece of this puzzle is the lack of correspondence

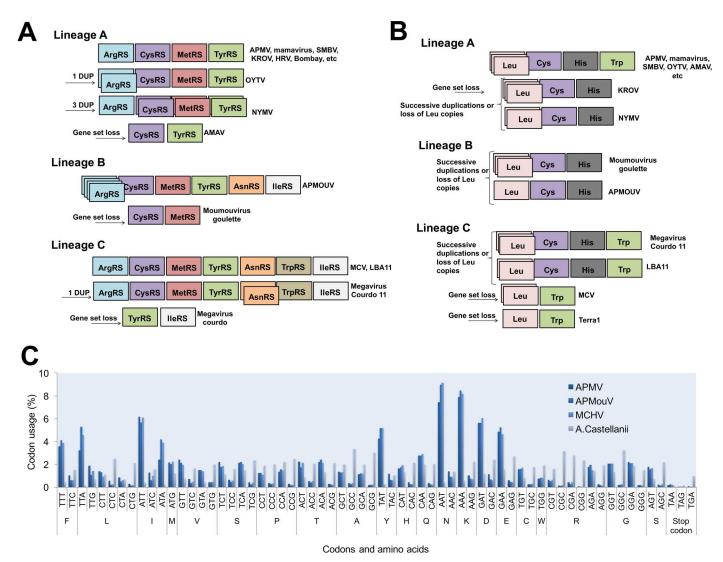


Fig 1. Schematic view of mimiviruses' aaRS (A) and tRNA (B) and codon/amino acid usage (C). The distribution of these genes follows a clear pattern of diversity among *Mimiviridae*: viral genomes with a wealthy aaRS and tRNA gene set (e.g., some lineage C isolates) contain all types of aaRS/tRNAs found in the rest of the family. This indicates a gradual inter- or intra-lineages gene loss throughout the evolution of different mimiviruses. The codon and amino acid usage analysis shows a clear difference between mimiviruses and *Acanthamoeba castellanii* patterns.

doi:10.1371/journal.pgen.1006532.g001



between mimivirus and *Acanthamoeba* codon/amino acid usages [10]. The requirements of mimiviruses for gene translation are quite different than those of *Acanthamoeba*. However, despite the genomes of mimiviruses of lineages A, B, and C presenting important differences and dissimilarities, the codon and amino acid usages are very similar among the members [10] (Fig 1). But, remarkably, the aaRS gene set present in the known mimiviruses does not match with the viral (or host) codon/amino acid usage demands, providing further evidence that a more complex piece of this puzzle has yet to be found.

Transfer RNAs in giant viruses—Expanding the translational apparatus

Similar to aaRS, tRNAs are essential molecules for the process of gene translation, being responsible for transporting an amino acid to a template complementary sequence in the molecule of messenger RNA, where the ribosome will further translate the genetic information [41,42]. The tRNAs are largely diffused among the cellular organisms, being the most abundant type of nucleic acid in the cells and constituting up to 10% of all cellular RNAs [39]. Sequences of tRNAs have already been described in some dsDNA viruses, such as members of the family *Myoviridae* [43], *Herpesviridae* [44], and *Phycodnaviridae* [45]. With the discovery of the mimiviruses, the viral tRNA repertoire increased [4].

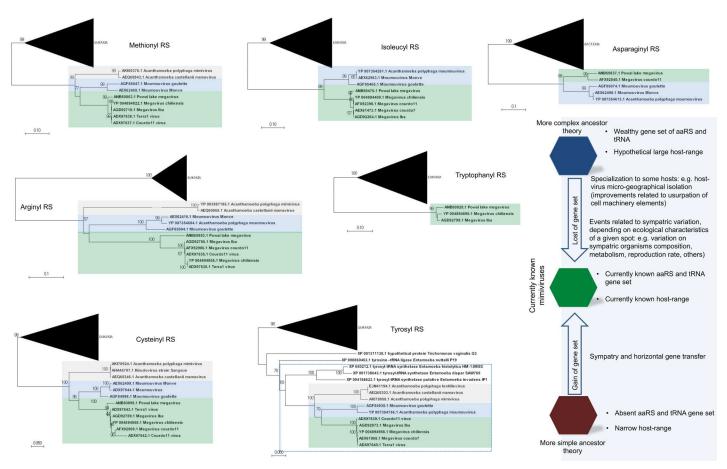


Fig 2. Phylogeny reconstruction of mimiviruses' aaRS. The unrooted trees were generated using MEGA 7 with the maximum likelihood method based on all aaRS found in mimiviruses. The trees were obtained after the alignment of the 100 best hits found in Genbank after BLASTing Megavirus chilensis aaRS predicted aa against all databases. Bacteria or Eukarya taxon, when present, were condensed from the outermost branch presenting bootstrap value >90. Mimiviruses of lineages A (grey), B (blue), and C (green) are highlighted. In all trees but the TyrRS tree, mimiviruses do not cluster inside cellular organism's branches. The bottom–right shows some evolutionary scenarios related to mimivirus evolution that considers their hosts and translation-related genetic data set.

doi:10.1371/journal.pgen.1006532.g002

APMV presents six sequences related to four different tRNAs: leucine (2x TAA and TTG), histidine (CAC), cysteine (TGC), and tryptophan (TGG) [4]. The same tRNAs were found in other mimiviruses of lineage A isolated in Brazil, such as Samba virus, Amazonia virus, Oyster virus, and Kroon virus (except Trp-tRNA) [46], as well as in mimivirus Terra2 [47]. Similar to Kroon virus, Niemeyer virus also has all tRNAs that are found in APMV but Trp-tRNA [15]. It is possible that the evolutionary history of these two viruses differs from that of the other representatives of lineage A, indicating a possible loss of this gene over time. In lineage B, Moumouvirus has sequences related to leucine (TTA), histidine (CAC), and cysteine (TGC), all present in the viruses from lineage A [16]. Considering lineage C, MCV also has three sequences related to tRNA, but only to leucine (TTA and TTG) and tryptophan (TGG) tRNAs [6]. The same tRNAs were identified in the genome of mimivirus Terra1 [47]. By contrast, the mimivirus LBA111 presents all four types of tRNAs found in mimiviruses of lineage A [18]. Similar to viruses from lineage A, it is possible that the ancestor of lineage C viruses had a more complete set of tRNA that was lost during the evolution. If we consider a common ancestor for all group I representatives of the *Mimiviridae* family, the same scenario is reasonable (Fig 1).

In the last years, the family *Mimiviridae* has also expanded the group distantly related to amoeba mimiviruses [48,49]. The analysis of the genome of CroV revealed an even higher range of tRNA-like sequences that had already been identified in their counterparts of group I, with a total of 22 sequences coding for five different tRNAs: leucine (9x TTA), serine (5x TCG), lysine (3x AAA), tyrosine (AAC), and asparagine (AAC), adding some new components in the tRNA set of mimiviruses [5]. Among the algae-infecting mimiviruses, PgV has eight sequences related to tRNA in its genome, coding for leucine (2x TTA e TTG), asparagine (2x AAC), isoleucine (ATA), arginine (AGA), and glutamine (CAA) [19]. OLV presents five tRNAs: leucine (TTG), isoleucine (ATA), tyrosine (TAC), asparagine (AAC), and arginine (AGA) [20] (Table 1). The great diversity of tRNAs coded by mimiviruses of group II is intriguing. Just like for mimiviruses of group I, it is possible that the common ancestor had a more complete set of these molecules that was lost over time. Taking into account the fact that those viruses infect different known hosts (microflagellates and algae), they likely had distinct evolutionary histories and have undergone different selective pressures, which might have contributed to the gain and loss of tRNA genes.

By analyzing the codon/amino acid usage of the mimiviruses and comparing it to the host usage, the hypothesis of multiple events of HGT become even less likely (Fig 2). Among the tRNAs encoded by mimiviruses, leucine (TTA) tRNA is the most common, being present in all of the viruses analyzed so far except for OLV (although it encodes for Leu[TTG]-tRNA), whereas in Acanthamoeba sp., it is one of the less frequently occurring tRNAs and is mainly encoded by CTG and CTC [10]. Leu(TAA)-tRNA, which is present in several mimivirus genomes and in multiple copies in some viral genomes, was hypothesized to complement the amoebal tRNA pool and may contribute to accommodating the viral AT-rich codons [10]. The mechanisms of gene expression in the beginning phase of the mimivirus replicative cycle may differ from the mechanisms for gene expression in later phases, and apart from viral RNA transcripts incorporated into mimivirus particles, mimivirus gene expression would first rely primarily on the amoebal machinery and then possibly become increasingly adapted to Mimivirus codon and amino acid usages [10]. In addition, 48% of all mimivirus tRNAs correspond to one of the 10 most frequently used codons in mimiviruses, while 84% of them correspond to one of the 10 least frequently used codons in their hosts [10]. Therefore, such differences suggest that the translational apparatus of the mimiviruses do not come from their currently known natural hosts, which supports the hypothesis that these viruses came from an unknown representative of the fourth TRUC of microbes [30,50,51], although other scenarios cannot be completely ruled out at this point [32,33,52].



Boosting the viral protein synthesis—Translation factors in mimiviruses

The translation of mRNA into proteins involves three major steps: initiation, elongation, and termination. For these steps to occur, some molecules, named translation factors (TFs), are required. Each step demands specific TFs that are essential for the protein synthesis to occur properly, generically known as initiation factors (IFs), elongation factors (EFs), and release factors (RFs) [53]. Each domain of life presents its own TFs that play similar roles during each step of the process [54]. The viruses do not possess such components, which is why they completely rely on the translational apparatus of their hosts. The discovery of the mimiviruses put this rule to the test, suggesting the existence of a possible fourth TRUC of life [7,50] and a "quasi-autonomous" nature of these giant viruses [55].

The genome of APMV presents sequences homologous to five TFs that are related to all three steps of translation: translation initiation factor 4E, translation initiation factor SUI1, translation initiation factor 4A, translation elongation factor eF-TU, and peptide chain release factor eRF1 [4]. The discovery of these genes, along with aaRS, was a milestone in virology since, until then, no virus was known to harbor sequences related to the translational apparatus, prompting an old debate: whether the viruses are living organisms and if they deserve a special place in the tree of life [4,56,57]. The same TFs were found in the genomes of several other mimiviruses of group I, and some of these genes are also present in the smaller mimiviruses, sequences homologous to IFs being found in the genomes of CroV [5], OLV [20], and PgV [19], but EFs and RFs are absent in these viruses (Table 1). The presence of TFs related to the initial step of translation in representatives of the whole family *Mimiviridae* suggests that these viruses have a weaker dependence on their hosts in the beginning of their replication cycle.

Until now, to the best of our knowledge, there have been no experimental studies about mimiviral initiation or elongation factors, which presents a gap in the biology of mimiviruses. However, work by Saini and Fischer (2007), based on 3-D models and an analysis of the conservation of functionally important residues and motifs, demonstrated that it is possible to derive functional attributes for six APMV ORFans, including initiation factor eIF4E [58]. Regarding the peptide chain RF present in mimiviruses, it was initially suggested it was a class-I RF with a sequence homologous to the RF found in eukaryotes and archaea [4]. The genes that encode eRF1 in APMV and MCV present two stop codons inside the coding region, and for the protein to be accurately synthesized, the viruses must be able to change the reading frame (frameshift recoding event), as well as bypass one of the stop codons (readthrough recoding event), which were considered exclusive features of bacteria [59]. Thus, the analysis of mimivirus RFs revealed a new type of TF that had never been seen before in any known group of organism.

This unique mixture of features of RFs from Eukarya, Archaea, and Bacteria found in mimiviruses raised further questions about their origin. The presence of a completely new type of RF makes its origin by successive events of HGT unlikely, thus supporting the hypothesis that the ancestral mimiviruses constituted a fourth TRUC of life [7,50]. If this scenario is true, we might consider that this ancestor presented a complete apparatus that has been gradually lost over successive speciation events. More studies about the TF of giant viruses will bring new insights about this issue, providing valuable clues to the intriguing mystery that is the origin and evolution of *Megavirales*.

What comes next? Unraveling the giant viruses' origin and evolutionary history

Since the discovery of mimiviruses, many theories regarding their origin and evolutionary history have arisen. As soon as the first mimivirus was discovered and its genome analyzed,

authors began to hypothesize that this virus stands within the tree (or rhizome [60]) of life [4]. In the following years, the discovery of new giant viruses increased their known pangenome, which supported the initial theories and opened windows for new ones by suggesting that they originated from a fourth TRUC of life [34,50] and also that they probably coexisted with cellular ancestors, evolving mainly through a genome reductive pattern [30,31]. Nevertheless, these theories were readily contested, with some researchers arguing that giant viruses should not be placed in the tree of life and that they came from other small parasitic elements instead of from an extinct branch of life, which had evolved by several HGT events [32,33,56,61,].

Despite many theories, the origin and evolution of mimiviruses remains a breathtaking mystery. It is not yet possible to exclude any specific scenario of giant viruses' evolution (Fig 2). However, the discovery of new viruses with exceptional genetic content, boosted by the advancement of phylogenomic analysis, provided increasing evidence to support a model wherein the mimivirus ancestor was a more complex organism. Taking this into account, we might speculate that this ancestor possessed a more complete translational-related gene set, which had been constantly losing and gaining genes (mainly through duplication) in accordance with an accordion model of evolution [39]. The Mimiviridae ancestor would already be a giant virus, but it would be one with a more independent and generalist lifestyle, able to infect different types of host cells or even interact with them, allowing the gene flow between ancestral lineages. This could have contributed to continuous gene gain and loss over time, shaping the whole viral genome and leading to modern mimiviruses. Considering the new techniques for isolating and discovering giant viruses that are currently implemented [62], we expect to find new mimi- and other giant viruses with genomes that more resemble this ancestor, advancing our understanding of the origin and evolution of this lineage of complex viruses.

Supporting information

S1 File. Arginyl RS alignment.
(MAS)
S2 File. Asparaginyl RS alignment.
(MAS)
S3 File. Cisteinyl RS alignment.
(MAS)
S4 File. Isoleucyl RS alignment.
(MAS)
S5 File. Metionyl RS alignment.
(MAS)
S6 File. Tirosyl RS alignment.
(MAS)

S7 File. Tryptophanyl RS alignment.

References

(MAS)

- Lwoff A. The concept of virus. J Gen Microbiol. 1957; 17: 239–253. doi: 10.1099/00221287-17-2-239 PMID: 13481308
- 2. ICTV, 2016 -Last taxonomic release: 2015 release.



- Raoult D, La Scola B, Birtles R. The Discovery and Characterization of Mimivirus, the Largest Known Virus and Putative Pneumonia Agent. Clin Infect Dis. 2007; 45: 95–102. doi: 10.1086/518608 PMID: 17554709
- Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-Megabase Genome Sequence of Mimivirus. Science. 2004; 306: 1344–1350. doi: 10.1126/science.1101485 PMID: 15486256
- Fischer MG, Allen MJ, Wilson WH, Suttle CA. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci USA. 2010; 107: 19508–19513. doi: 10.1073/pnas. 1007615107 PMID: 20974979
- Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci USA. 2011; 108: 17486–17491. doi: 10.1073/pnas.1110889108 PMID: 21987820
- Colson P, de Lamballerie X, Fournous G, Raoult D. Reclassification of Giant Viruses Composing a Fourth Domain of Life in the New Order Megavirales. Intervirology. 2012; 55: 321–332. doi: 10.1159/ 000336562 PMID: 22508375
- La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M. A giant virus in amoebae. Science. 2003; 299: 2033. doi: 10.1126/science.1081867 PMID: 12663918
- Abrahão JS, Dornas FP, Silva LC, Almeida GM, Boratto PV, Colson P, et al. Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding discoveries. Virol J. 2014; 11: 120. doi: 10.1186/1743-422X-11-120 PMID: 24976356
- Colson P, Fournous G, Diene SM, Raoult D. Codon Usage, Amino Acid Usage, Transfer RNA and Amino-Acyl-tRNA Synthetases in Mimiviruses. Intervirology. 2013; 56: 364–375. doi: 10.1159/ 000354557 PMID: 24157883
- 11. Klose T, Rossmann MG. Structure of large dsDNA viruses. Biol Chem. 2014; 359: 711–719.
- Legendre M, Santini S, Rico A, Abergel C, Claverie JM. Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. Virol J. 2011; 8: 99. doi: 10.1186/1743-422X-8-99 PMID: 21375749
- Colson P, Yutin N, Shabalina SA, Robert C, Fournous G, La Scola B, et al. Viruses with More Than 1,000 Genes: Mamavirus, a New Acanthamoeba polyphaga mimivirus Strain, and Reannotation of Mimivirus Genes. Genome Biol Evol. 2011; 3: 737–742. doi: 10.1093/gbe/evr048 PMID: 21705471
- Campos RK, Boratto PV, Assis FL, Aguiar ER, Silva LC, Albarnaz JD, et al. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. Virol J. 2014; 11: 95. doi: 10.1186/1743-422X-11-95 PMID: 24886672
- Boratto PV, Arantes TS, Silva LC, Assis FL, Kroon EG, La Scola B, et al. Niemeyer Virus: A New Mimivirus Group A Isolate Harboring a Set of Duplicated Aminoacyl-tRNA Synthetase Genes. Front Microbiol. 2015; 6: 1256. doi: 10.3389/fmicb.2015.01256 PMID: 26635738
- 16. Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, 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. doi: 1093/gbe/evs109 PMID: 23221609
- 17. Saadi H, Pagnier I, Colson P, Cherif JK, Beji M, Boughalmi M, et al. First isolation of Mimivirus in a patient with pneumonia. Clin Infec Dis. 2013; 57: e127–134.
- Yoosuf N, Pagnier I, Fournous G, Robert C, La Scola B, Raoult D, et al. Complete genome sequence of Courdo11 virus, a member of the family Mimiviridae. Virus Genes. 2014; 48: 218–223. doi: 10.1007/ s11262-013-1016-x PMID: 24293219
- Santini S, Jeudy S, Bartoli J, Poirot O, Lescot M, Abergel C, et al. Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. Proc Natl Acad Sci USA. 2013; 110: 10800–10805. doi: 10.1073/pnas.1303251110 PMID: 23754393
- Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, et al. Virophage control of antarctic algal host–virus dynamics. Proc Natl Acad Sci USA. 2011; 108: 6163–6168. doi: 10.1073/pnas. 1018221108 PMID: 21444812
- Piacente F, Marin M, Molinaro A, De Castro C, Seltzer V, Salis A, et al. Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine). J Biol Chem. 2012; 287: 3009–3018. doi: 10.1074/jbc.M111.314559 PMID: 22157758
- Piacente F, Bernardi C, Marin M, Blanc G, Abergel C, Tonetti MG. Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNAvirus Mimivirus. Glycobiology. 2014; 24: 51–61. doi: 10.1093/glycob/cwt089 PMID: 24107487



- 23. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, 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. doi: 10.1073/pnas.0911354106 PMID: 20007369
- Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, et al. Pandoraviruses: amoebae viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science. 2013; 341: 281–286. doi: 10.1126/science.1239181 PMID: 23869018
- Reteno DG, Benamar S, Khalil JB, Andreani J, Armstrong N, Klose T, et al. Faustovirus, an asfavirusrelated new lineage of giant viruses infecting amoebae. J Virol. 2015; 89: 6585–6594. doi: 10.1128/JVI. 00115-15 PMID: 25878099
- Legendre M, Lartigue A, Bertaux L, Jeudy S, Bartoli J, Lescot M, 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: E5327–5335. doi: 10.1073/pnas.1510795112 PMID: 26351664
- Ribas de Pouplana L, Schimmel P. Aminoacyl-tRNA synthetases: potential markers of genetic code development. Trends Biochem Sci. 2001; 26: 591–596. PMID: 11590011
- Li R, Macnamara LM, Leuchter JD, Alexander RW, Cho SS. MD Simulations of tRNA and AminoacyltRNA Synthetases: Dynamics, Folding, Binding, and Allostery. Int J Mol Sci. 2015; 16: 15872–15902. doi: 10.3390/ijms160715872 PMID: 26184179
- Fournier GP, Andam CP, Alm EJ, Gogarten JP. Molecular Evolution of Aminoacyl tRNA Synthetase Proteins in the Early History of Life. Orig Life Evol Biosph. 2011; 41: 621–632. doi: 10.1007/s11084-011-9261-2 PMID: 22200905
- 30. Nasir A, Kim KM, Caetano-Anollés G. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. BMC Evol Biol. 2012; 12: 156. doi: 10.1186/1471-2148-12-156 PMID: 22920653
- Nasir A, Caetano-Anollés G. A phylogenomic data-driven exploration of viral origins and evolution. Sci Adv. 2015; 1: e1500527. doi: 10.1126/sciadv.1500527 PMID: 26601271
- Yutin N, Wolf YI, Koonin EV. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology. 2014; 466–467: 38–52. doi: 10.1016/j.virol.2014.06.032 PMID: 25042053
- Koonin EV, Krupovic M, Yutin N. Evolution of double-strand DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. Ann N Y Acad Sci. 2015; 1341: 10–24. doi: 10.1111/nyas. 12728 PMID: 25727355
- Boyer M, Madoui MA, Gimenez G, La Scola B, Raoult D. Phylogenetic and Phyletic Studies of Informational Genes in Genomes Highlight Existence of a 4th Domain of Life Including Giant Viruses. PLoS ONE. 2010; 5: e15530. doi: 10.1371/journal.pone.0015530 PMID: 21151962
- **35.** Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D, et al. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. Genome Res. 2010; 20: 664–674. doi: 10.1101/gr.102582.109 PMID: 20360389
- 36. Abergel C, Chenivesse S, Byrne D, Suhre K, Arondel V, Claverie JM. Mimivirus TyrRS: preliminary structural and functional characterization of the first amino-acyl tRNA synthetase found in a virus. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2005; 61(Pt 2): 212–215. doi: 10.1107/S174430910500062X PMID: 16510997
- Abergel C, Rudinger-Thirion J, Giegé R, Claverie JM. Virus-Encoded Aminoacyl-tRNA Synthetases: Structural and Functional Characterization of Mimivirus TyrRS and MetRS. J Virol. 2007; 81: 12406–12417. doi: 10.1128/JVI.01107-07 PMID: 17855524
- Silva LC, Almeida GM, Assis FL, Albarnaz JD, Boratto PV, Dornas FP, et al. Modulation of the expression of mimivirus-encoded translation-related genes in response to nutrient availability during *Acantha-moeba castellanii* infection. Front Microbiol. 2015; 6: 539. doi: 10.3389/fmicb.2015.00539 PMID: 26082761
- Filée J. Genomic comparison of closely related giant viruses supports an accordion-like model of evolution. Front Microbiol. 2015; 6: 593. doi: 10.3389/fmicb.2015.00593 PMID: 26136734
- Boyer M, Azza S, Barrassi L, Klose T, Campocasso A, Pagnier I, et al. Mimivirus shows dramatic genome reduction after intraamoebal culture. Proc Natl Acad Sci USA. 108: 10296–10301. doi: 1073/pnas.1101118108 PMID: 21646533
- Giegé R. The early history of tRNA recognition by aminoacyl-tRNA synthetases. J Biosci. 2006; 31: 477–488. PMID: 17206068
- Barciszewska MZ, Perrigue PM, Barciszewski J. tRNA-the golden standard in molecular biology. Mol Biosyst. 2016; 12: 12–17. doi: 10.1039/c5mb00557d PMID: 26549858
- Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. Bacteriophage T4 Genome. Microbiol Mol Biol Rev. 2003; 67: 86–156. doi: 10.1128/MMBR.67.1.86-156.2003 PMID: 12626685



- 44. Virgin HW 4th, Latreille P, Wamsley P, Hallsworth K, Weck KE, Dal Canto AJ, et al. Complete Sequence and Genomic Analysis of Murine Gammaherpesvirus 68. J Virol. 1997. 71: 5894–5904. PMID: 9223479
- **45.** Van Etten JL, Meints RH. Giant viruses infecting algae. Annu Rev Microbiol. 1999; 53: 447–494. doi: 10.1146/annurev.micro.53.1.447 PMID: 10547698
- Assis FL, Bajrai L, Abrahao JS, Kroon EG, Dornas FP, Andrade KR, et al. Pan-Genome Analysis of Brazilian Lineage A Amoebal Mimiviruses. Viruses. 2015; 7: 3483–3499. doi: 10.3390/v7072782 PMID: 26131958
- 47. Yoosuf N, Pagnier I, Fournous G, Robert C, Raoult D, La Scola B, et al. Draft genome sequences of Terra1 and Terra2 viruses, new members of the family Mimiviridae isolated from soil. Virology. 2014; 452–453:125–132. doi: 10.1016/j.virol.2013.12.032 PMID: 24606689
- 48. Yutin N, Colson P, Raoult D, Koonin EV. Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. Virol J. 2013; 10: 106. doi: 10.1186/1743-422X-10-106 PMID: 23557328
- **49.** Claverie JM. Giant virus in the sea: Extending the realm of Megaviridae to Viridiplantae. Commum Integr Biol. 2013; 6: e25685.
- Raoult D. TRUC or the need for a new microbial classification. Intervirology. 2013; 56: 349–353. doi: 10.1159/000354269 PMID: 23867259
- Seligmann H, Raoult D. Unifying view of stem-loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses. Curr Opin Microbiol. 2016; 31: 1–8. doi: 10.1016/j. mib.2015.11.004 PMID: 26716728
- Moreira D, López-García P. Comment on "The 1.2-megabase genome sequence of Mimivirus". Science. 2005; 308: 1114. doi: 10.1126/science.1110820 PMID: 15905382
- Noble CG, Song H. Structural studies of elongation and release factors. Cell Mol Life Sci. 2008; 65: 1335–1346. doi: 10.1007/s00018-008-7495-6 PMID: 18213444
- Kyrpides NC, Woese CR. Universally conserved translation initiation factors. Proc Natl Acad Sci USA. 1998; 95: 224–228. PMID: 9419357
- Claverie JM, Abergel C. Mimivirus: the emerging paradox of quasi-autonomous viruses. Trends Genet. 2010; 26: 431–437. doi: 10.1016/j.tig.2010.07.003 PMID: 20696492
- Moreira D, Lópes-García P. Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol. 2009; 7: 306–311. doi: 10.1038/nrmicro2108 PMID: 19270719
- 57. Raoult D. There is no such thing as a tree of life (and of course viruses are out!). Nat Rev Microbiol. 2009. 7: 615.
- Saini HK, Fischer D. Structural and functional insights into Mimivirus ORFans. BMC Genomics. 2007;
 115. doi: 10.1186/1471-2164-8-115 PMID: 17490476
- Jeudy S, Abergel C, Claverie JM, Legendre M. Translation in Giant Viruses: A Unique Mixture of Bacterial and Eukaryotic Termination Schemes. PLoS Genet. 2012; 8: e1003122. doi: 10.1371/journal.pgen. 1003122 PMID: 23271980
- 60. Raoult D. The post-Darwinist rhizome of life. Lancet. 2010; 375: 104–105. doi: 10.1016/S0140-6736 (09)61958-9 PMID: 20109873
- Moreira D, Brochier-Armanet C. Giant viruses, giant chimeras: The multiple evolutionary histories of Mimivirus genes. BMC Evol Biol. 2008; 8: 12. doi: 10.1186/1471-2148-8-12 PMID: 18205905
- 62. Khalil JY, Andreani J, La Scola B. Updating strategies for isolating and discovering giant viruses. Curr Opin Microbiol. 2016; 31: 80–87. doi: 10.1016/j.mib.2016.03.004 PMID: 27039269