Codon Usage, Amino Acid Usage, Transfer RNA and Amino-Acyl-tRNA Synthetases in Mimiviruses

Philippe Colson^{a, b} Ghislain Fournous^{a, b} Seydina M. Diene^{a, b} Didier Raoult^{a–c}

^aURMITE UM3, CNRS 7278, IRD 198, INSERM U1905, Institut Hospitalo-Universitaire Méditerranée Infection, Facultés de Médecine et de Pharmacie, Aix-Marseille Université, and ^bPôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Centre Hospitalo-Universitaire Timone, Institut Hospitalo-Universitaire Méditerranée Infection, Marseille, France; ^cSpecial Infectious Agents Unit, King Abdulaziz University, Jeddah, Saudi Arabia

Key Words

Mimiviridae · Mimivirus · Codon usage · Codon bias · *Acanthamoeba* spp. · Amoeba · Marseillevirus · Marseilleviridae · Megavirales · Nucleocytoplasmic large DNA viruses

Abstract

Mimiviruses are giant viruses that infect phagocytic protists, including *Acanthamoebae* spp., which were discovered during the past decade. They are the current record holder among viruses for their large particle and genome sizes. One group is composed of three lineages, referred to as A, B and C, which include the vast majority of the Mimiviridae members. *Cafeteria roenbergensis* virus represents a second group, though the Mimiviridae family is still expanding. We analyzed the codon and amino acid usages in mimiviruses, as well as both the transfer RNA (tRNA) and amino acyl-tRNA synthetases. We confirmed that the codon and amino acid usages of these giant viruses are highly dissimilar to those in their amoebal host *Acanthamoeba castellanii* and are instead correlated with the high adenine and thymine (AT) content of Mimivirus genomes. We further describe that the set of tRNAs and amino acyl-tRNA synthetases in mimiviruses is globally not adapted to the codon and amino acid usages of

these viruses. Notwithstanding, Leu(TAA)tRNA, present in several Mimivirus genomes and in multiple copies in some viral genomes, may complement the amoebal tRNA pool and may contribute to accommodate the viral AT-rich codons. In addition, we found that the genes most highly expressed at the beginning of the Mimivirus replicative cycle have a nucleotide content more adapted to the codon usage in *A. castellanii*.

© 2013 S. Karger AG, Basel

Introduction

Acanthamoeba polyphaga mimivirus, discovered in 2003, is the first described giant virus to infect amoeba and is the founder of a new viral family named the Mimiviridae [1, 2]. Since its discovery, other mimiviruses have been described that infect *Acanthamoeba* spp. [3–6], and others, such as *Cafeteria roenbergensis*, have been described that infect green algae as well as heterokonts and haptophyta [7], which are two phagotrophic protists known to feed on bacteria [8]. Phylogeny reconstruction based on conserved genes has delineated two major groups (named I and II) within the family Mimiviridae [9, 10]. The first group is composed of three lineages, re-

ferred to as A, B and C, that include the vast majority of the Mimiviridae members, while *C. roenbergensis* virus (Crov) represents a second group [7], but the family Mimiviridae is still expanding [9, 10]. In addition, other giant, though smaller, amoebal viruses, including Marseillevirus and closely related viruses, have been discovered since 2008 and compose a second proposed family, the 'Marseilleviridae' [11–13].

Overall, mimiviruses and marseilleviruses appear to be common inhabitants of our biosphere, especially in water, and a growing body of evidence indicates that amoebal mimiviruses can be present in humans and can cause pneumonia [14–19]. *Acanthamoebae* spp., the hosts for these viruses, are free-living amoebae that are present worldwide in fresh and marine water and in soil [20, 21]. They can ingest any particle larger than 0.5 µm, including bacteria, fungi and giant viruses, some of which can survive and multiply within their amoebal host and are thus called amoeba-resisting microorganisms (ARMs) [22–24]. Multiple ARMs can live sympatrically within the same amoeba [11, 25], and it is therefore strongly suspected that these protists are hot spots for gene exchange between ARM genomes, including the mimiviruses and marseilleviruses, and are therefore genitors of mosaic genomes [23, 26].

Mimiviruses and marseilleviruses have been linked to the nucleocytoplasmic large DNA viruses, a monophyletic group of large DNA viruses that infect animals and unicellular eukaryotes and include poxviruses, asfarviruses, iridoviruses, ascoviruses and phycodnaviruses [2, 27–29]. These viruses were recently proposed to be reclassified within a new order named the 'Megavirales' [9]. One of the most remarkable and intriguing features revealed by the analyses of the gene content of mimiviruses was the presence of several genes encoding proteins involved in the translation apparatus, including amino acyl-transfer (t)RNA synthetases and elongation factors [2], which suggested that these viruses may not completely rely on the host translation machinery for their replication. The presence of these genes and of others that encode proteins involved in nucleotide biosynthesis (e.g. ribonucleotide reductase), DNA replication and repair (e.g. DNA polymerase B-family) and transcription (e.g. DNA-dependent RNA polymerase II) enabled phylogenetic and phyletic studies encompassing proteins from bacterial, archaeal, eukaryotic and viral lineages. Strikingly, these analyses suggested that mimiviruses and the other 'Megavirales' members compose a fourth domain of life [30], an issue that is debated but supported by recent findings [31–33].

Moreover, the genomes of mimiviruses were found to encode tRNAs, as previously observed in a few other large DNA viruses, including phycodnaviruses, baculoviruses, herpesviruses, myoviruses and siphoviruses [34, 35]. Remarkably, a study of evolutionary patterns in the sequences and structures of tRNAs from archaea, bacteria, eukaryotes and viruses indicated that viruses compose an ancient lineage that originated early in the archaeal lineage, before eukaryotes and bacteria [36]. The genomes of mimiviruses are also particularly rich in adenine and thymine (AT; approximately 75% of the nucleotide content) [2, 9]. In addition, it was reported that Mimivirus and other large DNA viruses display different amino acid compositions in their proteome compared with cellular organisms. Notably, in the proteomes of these viruses, it was shown that Asp is preferred over Glu, and Tyr is preferred over Phe, while the opposite is observed in most eukaryotic genomes [37]. Likewise, codon usage has been noted to be highly dissimilar for Mimivirus and its amoebal host [2]. In addition, in Mimivirus, codon usage bias, which corresponds to the unequal frequency of synonymous codons over coding DNA within genomes [38], was described as influenced by mutational pressure and translational selection, while amino acid usage was proposed to be guided by factors including aromaticity and cysteine content [39].

We analyzed codon and amino acid usages as well as tRNAs and amino acyl-tRNA synthetase distributions in the mimiviruses and in *Acanthamoeba castellanii*, their eukaryotic host.

Materials and Methods

Nucleotide sequences of the genes predicted to encode proteins from the genomes of giant viruses that infect amoebae were either available in our laboratory or were downloaded from the NCBI GenBank nucleotide sequence database. One or two representative viruses were chosen for the three lineages, A (Mimivirus; NC_014649 [2, 40]), B (Moumouvirus; NC_020104 [41]) and C (*Megavirus chilensis*; JN258408 [5] and LBA111; NC_020232 [16]), of mimiviruses associated with amoebae [9, 42]. More distantly related mimiviruses for which the genome was analyzed included Crov (NC_014637) [7] as well as Organic Lake phycodnavirus 2 (HQ704803) [43] and *Phaeocystis globosa* virus 12T (HQ634147) [44], which were recently reclassified in the family Mimiviridae [10].

Nucleotide composition, codon usage, amino acid usage, effective number of codons, relative synonymous codon usage (RSCU), codon adaptation index (CAI) and the relative deoptimization codon index (RDCI) were calculated using the CAIcal server (<http://genomes.urv.es/CAIcal/>) [45, 46]. Codon and amino acid usages are expressed as percentages and reflect the contribution made by each codon or amino acid, respectively.

RSCU assesses the frequency of use of a particular codon relative to the expected frequency of usage of this codon in the absence of codon usage bias [45]. RSCU values range from 0, if the codon is absent, through 1, if there is no codon usage bias, to 6, if a single codon is used from a six-codon family. CAI measures the synonymous codon usage bias by determining the level of similarity between the synonymous codon usage of a gene and the synonymous codon frequencies of a reference set [47, 48]. CAI values range from 0 to 1 when a gene encodes each amino acid using the most frequently used synonymous codon in the reference set of codon usage. This parameter has been used to compare codon usages, to assess viral gene adaptation to hosts and to predict levels of gene expression [47–50]. RCDI measures codon deoptimization by comparing the similarity in codon usage of a given gene to that of a reference genome [51, 52]. Furthermore, from a genetic point of view, estimation of RCDI may provide insight into the degree of coevolution between hosts and viruses. A low RCDI might indicate a high adaptation to a host. Moreover, a high RCDI might also indicate that some genes are expressed in latency phases or even that the virus might have a low replication rate. Transfer RNA sequences were detected in amoebae-associated giant viruses and in *A. castellanii* using the tRNA-Scan-SE website [53] and the Aragorn software [54], and manually checked using nucleotide BLAST searches against the viral genomes with formerly identified tRNAs [55]. The codon usage table for the *A. castellanii* genome was recalculated from the 14,012 mRNA coding DNA sequence, CDS, available from the NCBI GenBank nucleotide sequence database using the CAIcal server (<http://genomes.urv.es/CAIcal/>) [45, 46, 56]. The codon usages for the *A. castellanii* mitochondrion and for *Homo sapiens* were taken from the codon usage database (available at <http://www.kazuka.or.jp/codon/>).

The transcription profile of the first generation of Mimivirus corresponded to the transcript abundances previously determined by Legendre et al. [40]. For each gene, the total number of normalized read counts encompassing the different time points of the replication cycle and those determined at T = 0 h (T0) and 12 h (T12) of the viral replication cycle were used [40]. Phylogeny reconstruction of tRNA was performed by sequence alignment using the muscle software [57] and curated using gblocks [58]; then, the maximum likelihood method was used in the Mega v.5 software [59].

Plots were built in Microsoft Excel. Statistical analysis of the data was performed using a corrected χ^2 test or Fischer's exact test for comparison of proportions, while comparisons of means were performed with OpenEpi Epidemiologic Calculators v.2.3.1 (www.OpenEpi.com). p values <0.05 were considered to be statistically significant.

Results

The mean guanine and cytosine (GC)% for Mimivirus genes range from 23.4% for Crov to 33.5% for *P. globosa* virus 12T (fig. 1). When combining together all 3,571 genes from Mimivirus, Moumouvirus, *M. chliensis* and Crov, the mean GC% $\pm$ standard deviation (SD) was 26.3% $\pm$ 5.0. A set of 150 genes (4.2%) displayed a GC% 2

Fig. 1. GC% content of mimiviruses and *A. castellanii* genes.

SD (36.3) above the mean, whereas only 27 genes (0.7%) displayed a GC% 2 SD (16.3) below the mean value (online suppl. fig. S1; for all online suppl. material, see www.karger.com/doi/10.1159/000354557). In the first set of 150 genes, 115 were predicted to encode hypothetical proteins, whereas genes encoding several collagen-like proteins as well as capsid proteins and a putative regulator of chromosome condensation were among the 35 remaining genes. A total of 15 Mimivirus genes were among the top 20 most expressed as determined previously by transcriptomics [40], and 5 of these 15 genes have products that were identified in an earlier proteomic study as being incorporated into Mimivirus particles [60]. In the second set of 27 genes, all but 1 gene (a Crov gene encoding a putative N-acetyl transferase) were predicted to encode hypothetical proteins and belong to Crov.

Codon usages are very similar between studied mimiviruses and very dissimilar between mimiviruses and *A. castellanii* (fig. 2a, 3; online suppl. fig. S2–S3). The two most used codons in the mimiviruses are the AT-rich codons AAA and AAT, which encode lysine and asparagine, respectively (fig. 2a, 3; online suppl. fig. S2). In the set of mimiviruses studied here, these codons compose between 6.3 and 9.6% of all codons, with a mean value $\pm$ SD of 1.6% $\pm$ 1.8 (range 0–9.6; table 1). As a comparison,

Fig. 2. Codon (a) and amino acid (b) usages for genes of mimiviruses including Mimivirus, Moumouvirus, *M. chilensis*, LBA111, Crov, Organic Lake phycodnavirus 2 and *P. globosa* 12T.

AAA and AAT are the 2nd (4.9%) and 30th (1.4%) most frequently used codons in Marseillevirus, and the 46th (0.6%) and 49th (0.5%) most frequently used codons in *A. castellanii* (fig. 2a, 3). Conversely, the most frequently used codons in *A. castellanii* are GAG (5.2%), GCC (4.9%) and AAG (4.7%), which encode glutamic acid, alanine and lysine, and are the 41st (0.6%), 49th (0.4%) and 34th (1.1%) most frequently used codons in Mimivirus. Amino acid usages of predicted proteins in mimiviruses correlate with the AT-rich genomes of these viruses, and they are very similar among studied mimiviruses (fig. 2b, 4a, b). Indeed, the most frequently encountered amino acids in Mimivirus proteins are those encoded by AT-rich codons, namely Ile (ATT), Lys (AAA), Asn (AAT) and Leu (TTA), in decreasing order of frequency in Mimivirus. The Mimivirus amino acid usage is considerably different from the usage in *A. castellanii*, the host for these largest

viruses, whereas it tends to be less dissimilar between Marseillevirus, another amoebae-associated giant virus, and *A. castellanii* (fig. 4a, b).

In regards to amino acyl-tRNA synthetases, asparaginyl-tRNA synthetases are present in amoebae-associated mimiviruses of lineages B and C but are absent from genomes of amoebae-associated mimiviruses of lineage A and of some other distant mimiviruses including Crov (table 1). In addition, a tryptophanyl-tRNA synthetase is only present in amoebae-associated mimiviruses of lineage C. Overall, these amino acyl-tRNA synthetases do not match the amino acids most frequently found in Mimivirus-predicted proteins (fig. 4b). In addition, phylogeny reconstructions based on these amino acyl-tRNA synthetases show topologies that are similar to those based on core genes of mimiviruses and other members of the ‘Megavirales’, which delineates the three lineages

Fig. 3. Comparison of codon usage for Mimivirus and *A. castellanii* genes.

Fig. 4. Amino acid usage of Mimivirus (**a**, left) and Marseillevirus (**a**, right) compared to that of *A. castellanii*, and of Mimivirus (**b**, left), Moumouvirus (**b**, middle) and *M. chileensis* (**b**, right). Amino acyl-tRNA synthetase and tRNA encoded by the genomes of Mimivirus (**b**, left), Moumouvirus (**b**, middle) and *M. chileensis* (**b**, right) are indicated. Arrows indicate amino acyl-tRNA synthetases present only in some lineages of amoebae-associated mimiviruses.

[illegible]

TAT Y TAC	5.6	5.9	6.0	5.5	5.8	4.8	2.9	2.6	2.8	4.9	1.4	7.6	2.1	2.2	2.3	2.3	2.3	2.1	2.3	1.9	1.1	1.0	1.1	1.9	0.5	3.0	0.8	-	-	-	-	-	-	1	pr.	pr.	pr.	-	-			
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
CAT H CAC	2.1	2.1	2.3	2.2	2.3	1.9	1.9	2.7	2.6	2.2	0.3	2.7	1.6	0.8	0.8	0.8	0.8	0.7	0.9	0.7	0.7	1.0	0.9	0.8	0.1	1.0	0.6	1	1	-	-	-	-	-	-	-	-	-	-			
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
CAA Q CAG	3.1	2.9	3.1	3.1	3.2	3.5	2.8	3.4	4.0	4.7	3.3	0.4	4.0	2.5	0.8	0.7	0.8	0.8	0.8	0.9	0.7	0.8	1.0	1.2	0.8	0.1	1.0	0.6	-	-	-	-	-	-	-	-	-	-	-	-		
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
AAT N AAC	8.6	9.7	9.7	9.8	9.8	8.1	8.2	3.6	3.2	3.6	7.6	2.7	13.0	2.2	2.7	3.1	3.1	3.1	3.1	2.6	2.6	1.1	1.0	1.1	2.4	0.9	4.1	0.7	-	-	-	-	-	-	-	1	pr.	pr.	pr.	-	-	
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
AAA K AAG	8.9	9.4	8.9	8.9	10.6	9.9	9.7	8.8	5.3	5.6	8.9	1.6	12.1	5.7	1.7	1.8	1.7	1.7	2.0	1.9	1.8	1.7	1.0	1.1	1.7	0.3	2.3	1.1	-	-	-	-	-	-	-	-	-	-	-	-	-	
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
GAT D GAC	6.5	6.1	6.4	6.4	5.4	5.6	6.1	4.3	5.5	4.7	5.7	0.7	7.1	4.4	1.2	1.1	1.2	1.2	1.0	1.0	1.1	0.8	1.0	0.9	1.0	0.1	1.3	0.8	-	-	-	-	-	-	-	-	-	-	-	-		
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
GAA E GAG	5.4	5.7	5.0	5.0	5.8	6.9	6.1	9.2	6.7	6.9	6.4	1.3	9.0	3.7	0.8	0.8	0.7	0.7	0.9	1.0	0.9	1.4	1.0	1.0	0.9	0.2	1.3	0.6	-	-	-	-	-	-	-	-	-	-	-	-	-	
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
TGT C TGC	2.0	2.0	2.2	2.1	1.4	1.8	1.8	2.3	1.6	2.3	1.9	0.3	2.5	1.3	1.3	1.3	1.4	1.3	0.9	1.1	1.1	1.4	1.0	1.5	1.2	0.2	1.6	0.8	-	-	-	-	-	-	-	1	pr.	pr.	pr.	-	-	
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
CGT R CGC	3.3	3.0	3.0	3.0	2.3	3.0	3.4	6.1	6.2	5.7	3.8	1.5	6.8	0.8	0.5	0.5	0.5	0.5	0.4	0.5	0.5	1.0	1.0	0.9	0.6	0.2	1.1	0.1	-	-	-	-	-	-	-	-	19	pr.	pr.	pr.	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
CGA CGG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
AGA AGG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
GGT G GGC	4.3	4.3	4.3	4.2	3.6	3.9	4.6	6.3	6.9	6.6	4.8	1.2	7.1	2.4	0.6	0.6	0.6	0.6	0.5	0.6	0.7	0.9	1.0	1.0	0.7	0.2	1.0	0.3	-	-	-	-	-	-	-	-	-	-	-	-		
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
GGA GGG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
ATG M TGG W	2.4	2.1	2.4	2.5	2.2	3.0	3.0	2.2	2.2	2.2	2.4	0.4	3.2	1.7	1.1	0.9	1.1	1.1	1.0	1.3	1.3	1.0	1.0	1.1	0.2	1.4	0.8	-	-	-	-	-	-	-	-	21	pr.	pr.	pr.	pr.	-	
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
TTA Stop	0.8	0.9	0.9	0.9	0.8	0.9	0.8	1.7	1.4	1.3	1.0	0.3	1.7	0.4	0.6	0.6	0.6	0.6	0.6	0.6	0.6	1.2	1.0	0.9	0.7	0.2	1.2	0.3	1	-	1	-	-	-	-	-	-	-	-	-		
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
A gray background for amino acid usage indicates that the number is above (+2 SD) or below (-2 SD) the mean value for all studied organisms. pr. = Present. ^a pseudo-rRNA.																																										

A gray background for amino acid usage indicates that the number is above (+2 SD) or below (-2 SD) the mean value for all studied organisms. pr. = Present. ^a pseudo-tRNA.

Fig. 6. Correlation of Mimivirus transcript abundance of RDCI (**a**) and CAI (**b**) for Mimivirus genes in reference to the *A. castellanii* codon usage table.

A–C for mimiviruses of amoebae (fig. 5a; online suppl. fig. S4–S6).

As for tRNAs, overall, they are not necessarily the cognate for codons and amino acids that are the most abundant in mimiviruses (table 1). Amongst the mimiviruses, only Crov has a genome that encodes Lys(AAA)-tRNA, present in three copies amongst the 22 tRNA genes, and no Mimivirus genome encodes Asn(AAT)-tRNA, while Lys(AAA)- and Asn(AAT)-tRNA represent 0.8 and 0.2% of the *A. castellanii* tRNA pool, respectively. In contrast, the most frequently encountered tRNA in Mimivirus genomes is the Leu(TTA)-tRNA, which accounts for 17 of the 50 Mimivirus tRNA (34%), and TTA is between the 3rd (in Crov) and the 8th (in Mimivirus) most frequently used codon in proteins of mimiviruses. 9 Leu(TTA)-tRNA genes are particularly present in the Crov genome, which represent 45% of the Crov tRNA pool. In contrast, Leu(TTA)-tRNA is one of the less frequent tRNAs amongst those detected in the *A. castellanii* genome. Leucine is frequently found in both Mimivirus and *A. castellanii* proteins, but, in the first case, it is mostly encoded by TTA (anticodon TAA) and, in the second case, it is mostly encoded by CTG and CTC (fig. 2, 3). In addition, 24 Mimivirus tRNAs (48% of the total number) correspond to 1 of the 10 most frequently used codons in mimiviruses, and 20 of these 24 tRNAs correspond to one of the 10 less frequently used codons in *A. castellanii*. Be-

sides, the Mimivirus tRNA pool generally does not complement amino acyl-tRNA synthetases present in mimiviruses with the exception of Trp-tRNA in *M. chileensis* and Cys-tRNA in Mimivirus and Moumouvirus (fig. 4b). Notably, phylogeny reconstructions based on these Mimivirus tRNAs show in almost all cases topologies that delineate the previously defined lineages of mimiviruses of amoebae (fig. 5b; online suppl. fig. S7–S11).

We correlated the values of transcript abundance obtained for Mimivirus genes by transcriptomics [40] to those of CAI and RDCI determined with the *A. castellanii* codon usage as reference (fig. 6a, b). A total of 6 Mimivirus genes (38%) were identified as the most transcribed at T0 amongst the 16 genes with a CAI 2 SD (0.30) above the mean value for all genes. All of these 16 genes encode hypothetical proteins, 2 of which (R705, L725) were previously detected by proteomics inside Mimivirus particles [60], and 5 had been identified as being amongst the 20 most expressed genes. In addition, of the 47 with an RDCI in the lowest 5th percentile (<4.33), 30 genes (64%) were identified as the most transcribed at T0. These 30 gene products include an ankyrin-containing protein (L675), a VV A18-like helicase (L173), a choline dehydrogenase (L128) and two serine/threonine protein kinases (R826 and R831). Of note, the three latter proteins were predicted to be involved in lateral gene transfer with eukaryotes, the two serine/threonine protein kinases being pre-

dicted to be of amoebal origin [61, 62]. In addition, 5 of these 30 genes were among the 20 most expressed genes. Next, we considered the proportions of Mimivirus genes with a transcript abundance above the median value (69) calculated for all genes, within a set of 50 genes with the lowest RDCI (mean 3.52 ± 0.77 ; range 1.53–4.35) indicating they were the genes most adapted to *A. castellanii*. Among genes with a low RDCI, the proportion of highly transcribed genes was statistically significantly higher at T0 than at T12 in the Mimivirus replicative cycle (28/50 vs. 4/50; $p < 1e-3$). In addition, a significantly higher proportion of genes had an RDCI below the median value (7.1) amongst the 50 most transcribed Mimivirus genes when the RDCI was measured at T0 compared to T12 of the replicative cycle [48/50 (96%) vs. 23/50 (46%); $p < 1e-3$]. Additionally, the mean RDCI values were significantly lower for the most expressed genes at T0 than for the most expressed genes at T12 [4.60 ± 1.49 (2.15 ± 8.94) vs. 7.62 ± 2.65 (3.18–16.85); $p < 1e-3$]. Finally, the CAI determined for Mimivirus genes in reference to the codon usage tables of the *A. castellanii* genome, the *A. castellanii* mitochondrial genome, the *Dictyostelium discoideum* genome (a soil amoeba) and the *H. sapiens* genome indicates that the Mimivirus codon usage is less adapted to that inferred from the *A. castellanii* genome (online suppl. fig. S12).

Discussion

The codon usage in mimiviruses is biased by the high AT content of the genomes of these viruses. In addition, the most frequently encoded amino acids in Mimivirus proteins generally correspond to the AT-rich composition of the Mimivirus genomes. There are some exceptions, such as leucine, which is encoded by both AT-rich and GC-rich codons. Moreover, the amino acid composition of Mimivirus-predicted proteins appears not to be one that is favored by the tRNA pool of mimiviruses, but rather corresponds to the AT-rich composition of the genomes of mimiviruses. Thus, in a majority of cases, the tRNAs of mimiviruses do not predispose these viruses to a better adaptation to their high AT content, nor do they correspond to the codons frequently used by these viruses but rarely used by the amoebal genes. An exception is Leu(TAA)-tRNA encoded by the genome of mimiviruses, which is present in up to 9 copies in the Crov genome and is concurrently very rare in *A. castellanii*. The latter pattern observed for Leu(TAA)-tRNA distribution among the mimiviruses and *A. castellanii* is similar to that de-

scribed in phages and their bacterial hosts in a previous study [63]. It was observed that tRNAs present in phages tended to simultaneously fit codons that are highly used in phage genes and that are rarely used in host genes, which led to the hypothesis that a selective recruitment of tRNAs could compensate for differences in codon usages in phages and their hosts. Such a pattern has also been recently shown for cyanophages of the T4-like Myoviridae family that infect oceanic *Prochlorococcus* and *Synechococcus* hosts [64]. Thus, these cyanophages have genomes that can be more AT-rich than that of their host, but they concurrently have their own specific set of tRNAs that complements that of their host to accommodate the viral AT-rich codons. Overall, previous findings suggest that *Acanthamoebae* spp. may not be the natural hosts for mimiviruses.

The phylogeny based on the tRNAs encoded by the genomes of mimiviruses shows the same topology as that observed for the phylogeny based on Megavirales core genes. These observations are congruent with previous results obtained by Sun and Caetano-Anolles [36] using a comprehensive database of tRNAs, which suggested that the sequence and structure of these tRNA molecules are considerably conserved and may carry evolutionary signatures. Moreover, these data are congruent with the previous description of 3 lineages amongst amoeba-associated mimiviruses [9, 42]. In addition, we detected a set of Mimivirus genes for which the nucleotide composition and codon usage bias differ significantly from that of the majority of genes, which suggests they might have been involved in lateral gene transfer events. An interesting finding was the correlation revealed between the Mimivirus gene transcription levels measured at T0 of the replicative cycle in a previous experiment and the adaptation of codon usage of these genes to the amoebal host. This observation suggests that 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. Thus, apart from viral RNA transcripts incorporated into Mimivirus particles [65], Mimivirus gene expression may first rely primarily on the amoebal machinery and, then, may become increasingly adapted to Mimivirus codon and amino acid usages.

Disclosure Statement

No potential conflict of interest or financial disclosure for any authors.

References

- La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie JM, Raoult D: A giant virus in amoebae. *Science* 2003;299:2033.
- Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM: The 1.2-megabase genome sequence of Mimivirus. *Science* 2004;306:1344–1350.
- La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, Merchat M, Suzan-Monti M, Forterre P, Koonin E, Raoult D: The virophage as a unique parasite of the giant Mimivirus. *Nature* 2008;455:100–104.
- La Scola B, Campocasso A, N'Dong R, Fournous G, Barrassi L, Flaudrops C, Raoult D: Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. *Intervirology* 2010;53:344–353.
- 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.
- Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, Azza S, Klose T, Wong J, Rossmann MG, La Scola B, Raoult D, Koonin EV: Related giant viruses in distant locations and different habitats: *Acanthamoeba polyphaga mouloumouvirus* represents a third lineage of the Mimiviridae that is close to the Megavirus lineage. *Genome Biol Evol* 2012;4:1324–1330.
- 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.
- Colson P, Gimenez G, Boyer M, Fournous G, Raoult D: The giant *Cafeteria roenbergensis* virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. *PLoS One* 2011;6:e18935.
- 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.
- 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.
- Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG, Suzan-Monti M, La Scola B, Koonin EV, Raoult D: 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.
- Thomas V, Bertelli C, Collyn F, Casson N, Telenti A, Goesmann A, Croxatto A, Greub G: Lausannevirus, a giant amoebal virus encoding histone doublets. *Environ Microbiol* 2011;13:1454–1466.
- Colson P, Pagnier I, Yoosuf N, Fournous G, La Scola B, Raoult D: Marseilleviridae, a new family of giant viruses infecting amoebae. *Arch Virol* 2013;158:915–920.
- La Scola B, Marrie TJ, Auffray JP, Raoult D: Mimivirus in pneumonia patients. *Emerg Infect Dis* 2005;11:449–452.
- Cohen G, Hoffart L, La Scola B, Raoult D, Drancourt M: Ameba-associated Keratitis, France. *Emerg Infect Dis* 2011;17:1306–1308.
- Saadi H, Pagnier I, Colson P, Kanoun Cherif J, Beji M, Boughalmi M, Azza S, Armstrong N, Robert C, Fournous G, La Scola B, Raoult D: First isolation of Mimivirus in a patient with pneumonia; 2013;57:e127–e134.
- Colson P, Fancello L, Gimenez G, Armougom F, Desnues C, Fournous G, Yoosuf N, Million M, La Scola B, Raoult D: Evidence of the megavirome in humans. *J Clin Virol* 2013;57:191–200.
- Popgeorgiev N, Boyer M, Fancello L, Monteil S, Robert C, Rivet R, Nappez C, Azza S, Chiaroni J, Raoult D, Desnues C: Giant Blood Marseillevirus recovered from asymptomatic blood donors. *J Infect Dis* 2013, E-pub ahead of print.
- Colson P, La Scola B, Raoult D: Giant viruses of amoebae as potential human pathogens. *Intervirology* 2013;56:376–385.
- Barker J, Brown MR: Trojan horses of the microbial world: protozoa and the survival of bacterial pathogens in the environment. *Microbiology* 1994;140:1253–1259.
- Rodriguez-Zaragoza S: Ecology of free-living amoebae. *Crit Rev Microbiol* 1994;20:225–241.
- Greub G, Raoult D: Microorganisms resistant to free-living amoebae. *Clin Microbiol Rev* 2004;17:413–433.
- Raoult D, Boyer M: Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 2010;53:321–329.
- Aderem A, Underhill DM: Mechanisms of phagocytosis in macrophages. *Annu Rev Immunol* 1999;17:593–623.
- Bertelli C, Greub G: Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. *Front Cell Infect Microbiol* 2012;2:110.
- Moliner C, Fournier PE, Raoult D: Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol Rev* 2010;34:281–294.
- Iyer LM, Aravind L, Koonin EV: Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol* 2001;75:11720–11734.
- Iyer LM, Balaji S, Koonin EV, Aravind L: Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res* 2006;117:156–184.
- Yutin N, Koonin EV: Hidden evolutionary complexity of nucleo-cytoplasmic large DNA viruses of eukaryotes. *Virol J* 2012;9:161.
- 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.
- Williams TA, Embley TM, Heinz E: Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. *PLoS One* 2011;6:e21080.
- Nasir A, Kim KM, Caetano-Anolles 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.
- Legendre M, Arslan D, Abergel C, Claverie JM: Genomics of Megavirus and the elusive fourth domain of Life. *Commun Integr Biol* 2012;5:102–106.
- Dreher TW: Viral tRNAs and tRNA-like structures. *Wiley Interdiscip Rev RNA* 2010;1:402–414.
- Chan PP, Lowe TM: GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 2009;37:D93–D97.
- Sun FJ, Caetano-Anolles G: Evolutionary patterns in the sequence and structure of transfer RNA: a window into early translation and the genetic code. *PLoS One* 2008;3:e2799.
- Mrazek J, Karlin S: Distinctive features of large complex virus genomes and proteomes. *Proc Natl Acad Sci USA* 2007;104:5127–5132.
- Plotkin JB, Kudla G: Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 2011;12:32–42.
- Sau K, Gupta SK, Sau S, Mandal SC, Ghosh TC: Factors influencing synonymous codon and amino acid usage biases in Mimivirus. *Biosystems* 2006;85:107–113.
- Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D, Lartigue A, Lescot M, Bernadac A, Poulain J, Abergel C, Claverie JM: mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res* 2010;20:664–674.
- Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, Azza S, Klose T, Wong J, Rossmann MG, La Scola B, Raoult D, Koonin EV: Related giant viruses in distant locations and different habitats: *Acanthamoeba polyphaga mouloumouvirus* represents a third lineage of the Mimiviridae that is close to the Megavirus lineage. *Genome Biol Evol* 2012;4:1324–1330.
- Desnues C, La Scola B, Yutin N, Fournous G, Robert C, Azza S, Jardot P, Monteil S, Campocasso A, Koonin EV, Raoult D: Provirophages and transpovirons as the diverse mobilome of giant viruses. *Proc Natl Acad Sci USA* 2012;109:18078–18083.
- Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Gibson JA, Cavicchioli R: Virophage control of antarctic algal host-virus dynamics. *Proc Natl Acad Sci USA* 2011;108:6163–6168.

- 44 Brussaard CP, Short SM, Frederickson CM, Suttle CA: Isolation and phylogenetic analysis of novel viruses infecting the phytoplankton *Phaeocystis globosa* (Prymnesiophyceae). *Appl Environ Microbiol* 2004;70:3700–3705.
- 45 Puigbo P, Bravo IG, Garcia-Vallve S: CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct* 2008;3:38.
- 46 McInerney JO: GCUA: general codon usage analysis. *Bioinformatics* 1998;14:372–373.
- 47 Puigbo P, Bravo IG, Garcia-Vallve S: E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* 2008;9:65.
- 48 Sharp PM, Li WH: The Codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 1987;15:1281–1295.
- 49 Limor-Waisberg K, Carmi A, Scherz A, Pilpel Y, Furman I: Specialization versus adaptation: two strategies employed by cyanophages to enhance their translation efficiencies. *Nucleic Acids Res* 2011;39:6016–6028.
- 50 Goetz RM, Fuglsang A: Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from *Escherichia coli*. *Biochem Biophys Res Commun* 2005;327:4–7.
- 51 Puigbo P, Aragones L, Garcia-Vallve S: RCDI/eRCDI: a web-server to estimate codon usage deoptimization. *BMC Res Notes* 2010;3:87.
- 52 Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E: Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol* 2006;80:9687–9696.
- 53 Schattner P, Brooks AN, Lowe TM: The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 2005;33:W686–W689.
- 54 Laslett D, Canback B: ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 2004;32:11–16.
- 55 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
- 56 Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, Bertelli C, Schilde C, Kianianmomeni A, Burglin TR, Frech C, Turcotte B, Kopec KO, Synnott JM, Choo C, Paponov I, Finkler A, Soon Heng TC, Hutchins AP, Weinmeier T, Rattei T, Chu JS, Gimenez G, Irimia M, Rigden DJ, Fitzpatrick DA, Lorenzo-Morales J, Bateman A, Chiu CH, Tang P, Hegemann P, Fromm H, Raoult D, Greub G, Miranda-Saavedra D, Chen N, Nash P, Ginger ML, Horn M, Schaap P, Caler L, Loftus B: Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol* 2013;14:R11.
- 57 Edgar RC: MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004;5:113.
- 58 Castresana J: Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000;17:540–552.
- 59 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28:2731–2739.
- 60 Renesto P, Abergel C, Decloquement P, Moinier D, Azza S, Ogata H, Fourquet P, Gorvel JP, Claverie JM: Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. *J Virol* 2006;80:11678–11685.
- 61 Moreira D, Brochier-Armanet C: Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. *BMC Evol Biol* 2008;8:12.
- 62 Filee J, Siguier P, Chandler M: I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *Trends Genet* 2007;23:10–15.
- 63 Bailly-Bechet M, Vergassola M, Rocha E: Causes for the intriguing presence of tRNAs in phages. *Genome Res* 2007;17:1486–1495.
- 64 Enav H, Beja O, Mandel-Gutfreund Y: Cyanophage tRNAs may have a role in cross-infectivity of oceanic *Prochlorococcus* and *Synechococcus* hosts. *ISME J* 2012;6:619–628.
- 65 Suzan-Monti M, La Scola B, Raoult D: Genomic and evolutionary aspects of Mimivirus. *Virus Res* 2006;117:145–155.