

Identification of giant Mimivirus protein functions using RNA interference

Haitham Sobhy¹, Bernard La Scola^{1,2}, Isabelle Pagnier¹, Didier Raoult^{1,2} and Philippe Colson^{1,2*}

¹ URMITE UM63 CNRS 7278 IRD 198 INSERM U1095, Faculté de Médecine, Aix-Marseille University, Marseille, France,

² IHU Méditerranée Infection, Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Centre Hospitalo-Universitaire Timone, Assistance Publique-Hôpitaux de Marseille, Marseille, France

OPEN ACCESS

Edited by:

Bruno Pozzetto,
University Jean Monnet
of Saint-Etienne, France

Reviewed by:

Anton G. Kutikhin,
Research Institute for Complex Issues
of Cardiovascular Diseases, Russia
Mart Krupovic,
Institut Pasteur, France

*Correspondence:

Philippe Colson,
URMITE UM63 CNRS 7278 IRD 198
INSERM U1095, Faculté
de Médecine, Aix-Marseille University,
27 Boulevard Jean Moulin,
13385 Marseille Cedex 05, France
philippe.colson@univ-amu.fr

Specialty section:

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

Received: 16 February 2015

Paper pending published:
21 March 2015

Accepted: 07 April 2015

Published: 28 April 2015

Citation:

Sobhy H, La Scola B, Pagnier I,
Raoult D and Colson P (2015)
Identification of giant Mimivirus protein
functions using RNA interference.
Front. Microbiol. 6:345.
doi: 10.3389/fmicb.2015.00345

Genomic analysis of giant viruses, such as Mimivirus, has revealed that more than half of the putative genes have no known functions (ORFans). We knocked down Mimivirus genes using short interfering RNA as a proof of concept to determine the functions of giant virus ORFans. As fibers are easy to observe, we targeted a gene encoding a protein absent in a Mimivirus mutant devoid of fibers as well as three genes encoding products identified in a protein concentrate of fibers, including one ORFan and one gene of unknown function. We found that knocking down these four genes was associated with depletion or modification of the fibers. Our strategy of silencing ORFan genes in giant viruses opens a way to identify its complete gene repertoire and may clarify the role of these genes, differentiating between junk DNA and truly used genes. Using this strategy, we were able to annotate four proteins in Mimivirus and 30 homologous proteins in other giant viruses. In addition, we were able to annotate >500 proteins from cellular organisms and 100 from metagenomic databases.

Keywords: Mimivirus, giant virus, Megavirales, fiber, short interfering RNA, RNA interference, nucleocytoplasmic large DNA virus

Introduction

Acanthamoeba polyphaga mimivirus was the first member discovered of the viral family *Mimiviridae*, which encompasses viruses that infect *Acanthamoeba* sp. (La Scola et al., 2003; Raoult et al., 2004). Subsequently, dozens of Mimivirus relatives have been isolated from environmental samples and, more recently, from humans (La Scola et al., 2008; Boyer et al., 2009; Fischer et al., 2010; Arslan et al., 2011; Yoosuf et al., 2012; Saadi et al., 2013a,b). Other viruses that infect protozoa were also subsequently discovered, including marseilleviruses (Boyer et al., 2009; Boughalmi et al., 2013a,b; Aherfi et al., 2014), pandoraviruses (Philippe et al., 2013) and *Pithovirus sibericum* (Legendre et al., 2014). Mimiviruses have been linked, along with the marseilleviruses, to the nucleocytoplasmic large DNA viruses (NCLDVs), which were recently proposed to be unified into a new viral order named the “Megavirales” (Colson et al., 2013). These giant viruses have raised considerable interest in the field of evolutionary biology because of their unexpectedly large size, as well as the fact that they contain genes encoding functions previously believed to be in the domain of cellular organisms, such as aminoacyl-tRNA synthetases or translations factors. They have challenged the definition of a virus (Raoult et al., 2004; Moreira and Brochier-Armanet, 2008; Raoult and Forterre, 2008; Forterre, 2010).

The genomes of *Megavirales* members contain a large number of predicted genes annotated either as hypothetical proteins or ORFan genes, i.e., genes without homologs in sequence databases (Raoult et al., 2004; Boyer et al., 2010). For example, genes encoding hypothetical proteins occupy more than 50% of the Mimivirus genome. The functions of these genes are not known. To date, the functions of only a few Mimivirus proteins have been studied experimentally, including amino-acyl-tRNA synthetases (Abergel et al., 2005, 2007) and proteins involved in sugar biosynthesis (Piacente et al., 2012). Thus, the large majority of Mimivirus genes have no known function and make up a ‘functional dark matter.’

The Mimivirus capsid, which is approximately 500 nm in size, is covered by a dense layer of fibers. These viral fibers are approximately 125–140 nm in length and approximately 1.4 nm in diameter and consist of a soft shaft and a globular shaped head (Xiao et al., 2009; Klose et al., 2010; Kuznetsov et al., 2010). Clusters of 3–4 fibers were found to be linked via a disk shaped base. They are highly glycosylated, antigenic, and resistant to protease and collagenase treatment (Xiao et al., 2009; Boyer et al., 2011). A putative GMC-type oxidoreductase (R135), and two hypothetical proteins (L725, which is the product of an ORFan, and L829) were identified in purified fibers by gel electrophoresis coupled with matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS; Boyer

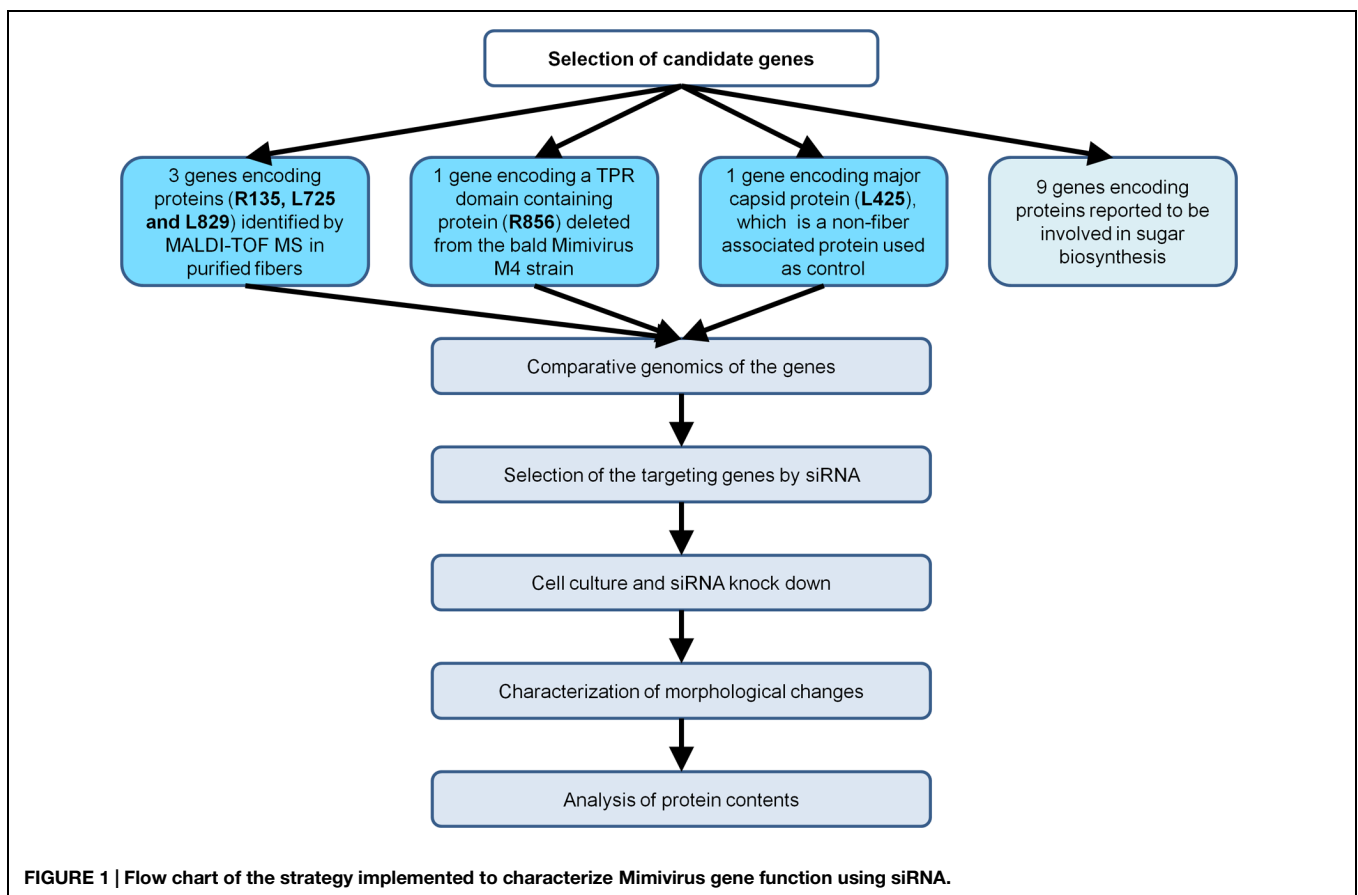
et al., 2011). Sub-culturing Mimivirus 150 times on germ-free amoebae led to the emergence of a mutant “M4” strain lacking fibers, and with a genome reduced by 16% and missing 150 genes (Boyer et al., 2011). Comparative proteomics of M4 and the original Mimivirus strain showed a deletion of the R135 and L829 proteins, as well as of the R856 protein (Boyer et al., 2011), which belongs to the group of tetratricopeptide repeat (TPR) containing proteins previously involved in virus–host interactions (Jeshtadi et al., 2010). In addition, nine proteins have been proposed to be involved in sugar biosynthesis and fiber formation (Piacente et al., 2012; Supplementary Table S1).

Here, we aimed to apply RNA interference (RNAi) to the identification of the function of Mimivirus proteins. We targeted four Mimivirus genes associated with fiber formation, as fibers can be easily observed by electron microscopy (Figure 1).

Materials and Methods

Cell Culture, siRNA, and Identification of Morphological Changes Targeted Genes and siRNA

We targeted the Mimivirus genes R135, L725, L829, and R856 using short interfering RNA (siRNA). These genes were either identified in purified fibers or deleted in the M4 strain (Boyer



et al., 2011; **Figure 1**). Regarding the negative controls, we used Mimivirus virions in absence of treatment with siRNAs, and also targeted the L425 gene, which is known to be expressed and encodes the major capsid protein that composes the Mimivirus shell located just beneath the fiber layer. The aim of this control was to ensure for the absence of possible non-specific, artefactual effect of siRNAs or lipid carrier on fibers.

Cell Culture and RNAi

A culture of *A. polyphaga* in 10 ml of PYG medium (5e5 amoeba/mL) was seeded for 24–48 h. Then, 100 μ l of Lipofectamine RNAiMAX (Invitrogen, USA) and 0.25 μ g of duplex siRNA (designed and purchased from Invitrogen; for sequences, see Supplementary Table S2) were used according to the manufacturer's instructions. To improve the siRNA specificity, we used duplex siRNA and we checked for specific and non-specific appariements by performing BLASTn searches against Mimivirus genes and GenBank. One ml of Mimivirus preparation ($\approx 1e^6$ viruses) was added to the culture, and incubated for 11 h at 32°C. *A. polyphaga* were harvested by centrifugation (500 g for 10 min) and analyzed by electron microscopy. For protein analysis, four 10-ml flasks of Mimivirus culture incubated for 24 h at 32°C were used. Cells were completely lysed. Then, the culture medium was centrifuged at 500 g for 10 min, and the supernatant was filtered through a 1.2- μ m filter to eliminate *A. polyphaga* cell debris. The Mimivirus pellet, obtained by centrifugation of the medium at 12,000 g for 15 min, was washed twice with phosphate buffer serum (PBS), and the purified viruses were used in further investigations. For the negative control gene, the same experimental procedures were applied to the L425 gene encoding capsid protein.

Electron Microscopy

The preparation of samples for electron microscopy was previously described (Boyer et al., 2011). Briefly, samples were fixed with glutaraldehyde (2.5%) and cacodylate buffer (0.1 M), cut into 70-nm sections using an ultramicrotome (UC7; Leica), collected on 400-mesh nickel grids with formvar carbon, and stained for electron microscopy (FCF-400-Ni, Electron Microscopy Sciences). The samples were then viewed with a Philips electron microscope (Morgagni 268D) at 80 keV. Cross sections of all pictures that were selected for the analysis were positioned at the middle of the virions and characterized by a dense cluster (black mass).

Analysis of Protein Content of Knocked Down Viruses

Antibodies Preparation

Fibers were purified from the virus as previously described for vaccinia virus (Jensen et al., 1996). Previous analysis of the fiber by 2D-gel coupled with MALDI-TOF MS (Boyer et al., 2011) revealed that three proteins (R135, L725, and L829) were associated with Mimivirus fibers. For anti-L725 antibodies, L725 protein fused with thioredoxin was expressed in *Escherichia coli* and purified using ÄKTA avant 25 (GE Healthcare, USA). Purified Mimivirus virions, fibers, and L725 protein were injected into

mice to obtain anti-L725 polyclonal antibodies, as previously described (Boyer et al., 2011).

Immunogold Labeling

Grids were immersed in NH_4Cl (50 mM) diluted in PBS three times for 5 min, washed in PBS for 5 min, and then immersed twice in blocking buffer (1% normal goat serum (NGS), 1% bovine serum albumin (BSA), and 0.2% Tween 20 diluted in PBS; 2×10 min). The grids were incubated with anti-fiber polyclonal antibody that was diluted 1:100 in blocking buffer overnight at 4°C. After four 10-min washes, the grids were incubated for 90 min in biotin (Beckman Coulter, USA) that was diluted 1:100 in blocking buffer. Then, the grids were washed with 0.1% BSA-PBS (4×5 min) and then in 0.01% cold water fish skin (CWFS) gelatin-PBS (3×10 min), and incubated with streptavidin (labeled by 10-nm gold nano-particles; Aurion, The Netherlands) that was diluted 1:100 in 0.01% CWFS gelatin-PBS for 90 min and washed with PBS. After incubating with PBS-glutaraldehyde 2.5% for 15 min, the grids were washed with PBS (2×10 min) and distilled water for 10 min. Finally, the grids were contrasted by adding uranyl acetate for 20 min, immersed in water 60 times, and analyzed using electron microscopy. The number of gold particles that were bound to fibers in each image was counted. The experiments included several steps that were performed on successive days. For each step, we used Mimivirus treated without siRNA with the same experimental conditions as the siRNA-treated Mimivirus as negative control.

Proteomic Analysis

All proteomic analysis (sample preparation, 1D and 2D gel electrophoresis, silver staining, and western blotting) was performed as previously described (Azza et al., 2009). Briefly, Mimivirus was solubilized in 40 mM Tris-HCl, pH 7.5, supplemented with 2% (wt/vol) sodium dodecyl sulfate (SDS; Sigma-Aldrich) and 60 mM dithiothreitol (DTT), followed by 5 min of heating at 95°C. The insoluble fraction was removed by centrifugation (12,000x g, 4°C, 10 min), and soluble proteins were precipitated using a PlusOne 2-D Clean-Up kit (GE Healthcare, USA) to remove SDS. The final pellet was re-suspended in solubilization buffer [7 M urea, 2 M thiourea, 4% (wt/vol) 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS)] and stored at -80°C until use. The protein concentration was measured by Bradford assay (Bio-Rad, USA). Immobiline DryStrips (13 cm, pH 3–10; GE Healthcare) were rehydrated overnight using 250 μ l rehydration buffer [8 M urea, 2% (w/v) CHAPS, 60 mM DTT, 0.5% (v/v) IPG buffer (GE Healthcare)] containing 20 μ g of solubilized Mimivirus proteins and isoelectric focusing (IEF) was carried out according to the manufacturer's protocol (IPGphor II, GE Healthcare). Before the second dimension electrophoresis was performed, strips were equilibrated twice in 5 ml equilibration buffer [30% (v/v) glycerol, 3% (w/v) SDS, 6 M urea, 50 mM Tris-HCl, bromophenol blue, pH 8.8] for 15 min. This buffer was supplemented with 65 mM DTT for the first equilibration and with 100 mM iodoacetamide for the second one. The strips were then embedded in 0.5% agarose and the proteins resolved

by 10% SDS-PAGE (Protean II XL, Bio-Rad). Gels were stained either with silver or transferred onto nitrocellulose membranes for western blot analysis using anti-Mimivirus, anti-L725 or anti-fiber primary polyclonal antibodies. Then, the membrane was washed three times with PBS-Tween and probed for 2 h with horseradish peroxidase-conjugated goat anti-mouse secondary antibodies.

Comparative Genomics and Phylogenetic Tree Reconstruction

Protein sequences of mimiviruses were retrieved from the NCBI GenBank non-redundant protein sequence database (nr) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). BLASTp searches were performed with 0.01 as the *e*-value cutoff. The best hits were collected and aligned using ClustalW (Larkin et al., 2007). The multiple sequence alignments were trimmed by Gblock (Castresana, 2000). Phylogenetic tree reconstructions were performed using the Maximum likelihood method of the FastTree tool with default parameters (Price et al., 2010).

PCR Testing

The presence or absence of Mimivirus genes in the purified viral solution was determined by qPCR. The most conserved sites were identified, and universal primers and probes were designed using the Gemi tool (Sobhy and Colson, 2012; Supplementary Table S3). The 25 μ l-real-time PCR mixture contained 5 μ l of extracted DNA, 12.5 μ l qPCR Mastermix (Eurogentec, Belgium), 0.5 μ l of each primer (10 nmol/ μ l; Eurogentec), and 0.5 μ l probe (3 nmol/ μ l; Applied Biosystems UK). The PCR thermal cycling conditions were: a hold at 50°C for 2 min, a hold at 95°C for 5 min, and then 45 cycles of 30 s at 95°C then 1 min at 60°C.

Results

Consequences of Silencing Targeted Mimivirus Genes on Fiber Formation

We knocked down the Mimivirus genes encoding the R135, L725, L829, R856, and L425 proteins using siRNA. We then compared the fibers from the viruses produced under these conditions to those from control viruses produced in the absence of siRNA

or treated with siRNA targeting non-fiber associated proteins (L425), searching for any abnormal feature of the fibers, such as short, prone (procumbent), or non-stretched and curved fibers (Table 1; Figure 2; Supplementary Figures S1–S6). To measure length ratio for silenced versus control viruses, we selected 4–8 viruses that harbored ≥ 30 fibers and measured the lengths of the fibers in each condition (Table 2). To determine protein contribution in fiber formation, we counted the number of gold particle conjugated with anti-fiber antibodies, hence to Mimivirus fibers (Table 3; Figure 3).

Control Mimivirus Fibers

The average length of control fibers was 131 nm (Table 2; Figures 2A and 3A–C, Supplementary Figure S1). Silencing the gene encoding the L425 protein did not affect fiber length or topology (Figures 2B and 3D, Supplementary Figure S2).

Fibers after R856 Gene Silencing

Approximately 60% of si-R856 viruses (meaning viruses treated with an siRNA targeting the R856 gene) harbored abnormal or short fibers (Table 1; Figure 2C, Supplementary Figure S3). We observed that the average length of fibers from si-R856 viruses was 48 nm, which was 64% shorter than control viruses ($p < 1e-6$; Table 2). The number of gold particles bound to fibers was decreased by 81% after silencing R856 (Table 3; Figure 3).

Fibers after L725 Gene Silencing

Si-L725 viruses harbored approximately 50% abnormal curved fibers, which were 15% shorter than control fibers (Tables 1 and 2). The number of gold particles bound to fibers was decreased by 86% after silencing the L725 gene (Table 3; Figures 2D and 3E; Supplementary Figure S4).

Fibers after L829 Gene Silencing

The fiber length of si-L829 viruses was 92 nm, which was 30% shorter than the control virus fibers (Table 2). In addition, the fiber layer was sparse in si-L829 viruses and the gold particle count was decreased by 57% (Tables 1 and 3; Figures 2E and 3F; Supplementary Figure S5).

TABLE 1 | Number of viruses with fibers with normal or abnormal features.

Silenced gene	Total number of viruses observed	Abnormality %	Average abnormality %	Number of viruses with short fibers (%)	Number of viruses with sparse/curved fibers (%)
Control	91	0	0	0 (0)	0 (0)
si-L425 (control)	90	0	0	0 (0)	0 (0)
si-R135	123	16–54	35	20 (16)	47 (38)
si-L725	52	21–71	46	11 (21)	26 (50)
si-L829	90	13–49	31	12 (13)	32 (36)
si-R856	101	58–63	61	59 (58)	5 (5)

Normal fibers are those with classical shape and > 100 nm-long, dense and stretched fibers. Sparse/curved fibers are fibers that are not as thick or dense as the normal or non-stretched fibers. Short fibers are those with a length < 100 nm. The percentage of abnormality indicates the percentage of viruses with either short or non-stretched and sparse fibers. si-, virus after silencing the target gene. See also Supplementary Figures S1–S6.

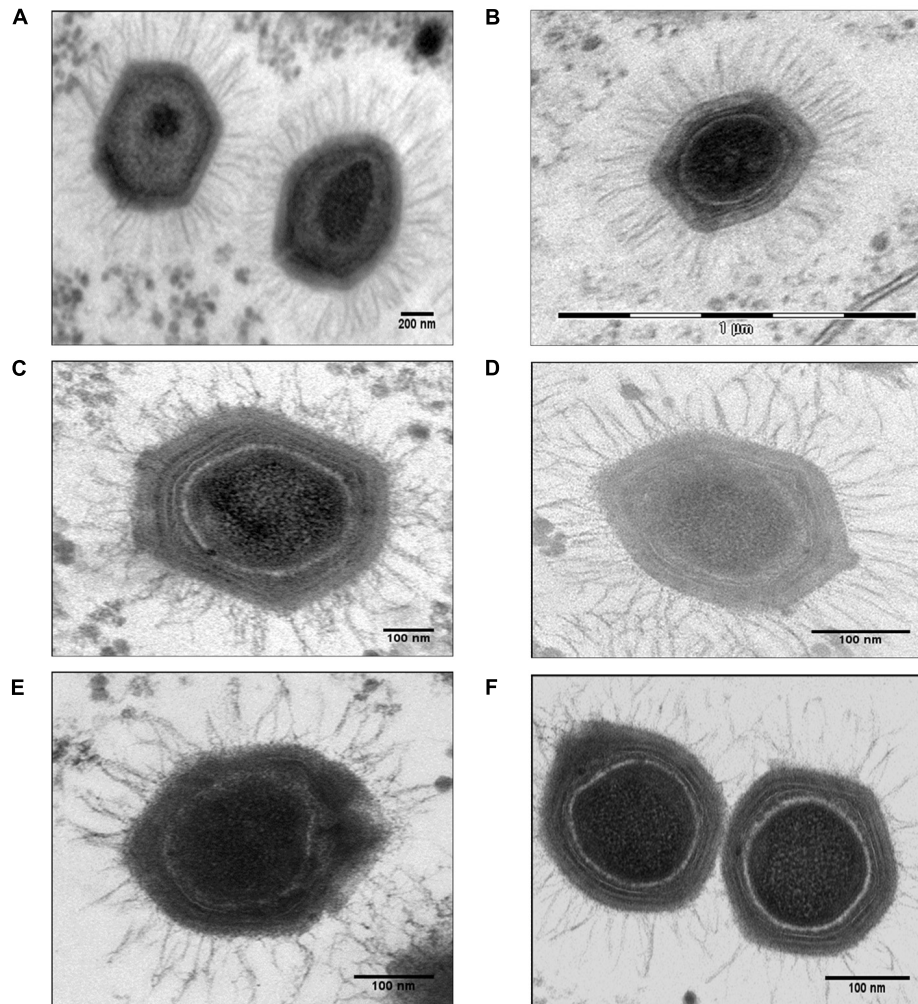


FIGURE 2 | Electron micrographs of control Mimivirus and Mimivirus after knocking down genes encoding fiber associated proteins (FAPs). (A) Control, (B) Control si-L425 Mimivirus (meaning Mimivirus treated with siRNA targeting the L425 gene); (C) si-R856

Mimivirus, (D) si-L725 Mimivirus, (E) si-L829 Mimivirus, and (F) si-R135 Mimivirus. The pictures were taken for virions within *Acanthamoeba polyphaga* host. See Supplementary Figures S1–S6 for additional figures.

TABLE 2 | Length of the Mimivirus fibers according to each siRNA experimental condition.

Virus	Number of viruses observed	Number of fibers	Mean length of fibers \pm standard deviation	Relative length of fibers compared to control (%; silenced / control fiber *100)	p-value ¹
Control	8	30	131.4 \pm 23.8	100.0	–
si-R135	6	34	115.0 \pm 27.5	87.5	0.0138
si-L725	4	31	110.8 \pm 26.0	84.4	0.0021
si-L829	7	33	91.8 \pm 25.0	69.9	<1e-6
si-R856	8	29	47.8 \pm 18.8	36.4	<1e-6

¹P-values were calculated using the ANOVA test (<http://www.openepi.com>). si-, virus after silencing the target gene. See also Supplementary Figures S1–S6.

Fibers after R135 Gene Silencing

The fiber length of the si-R135 viruses was 12% shorter than the control and 38% of these fibers were curved (Table 2; Figure 2F, Supplementary Figure S6).

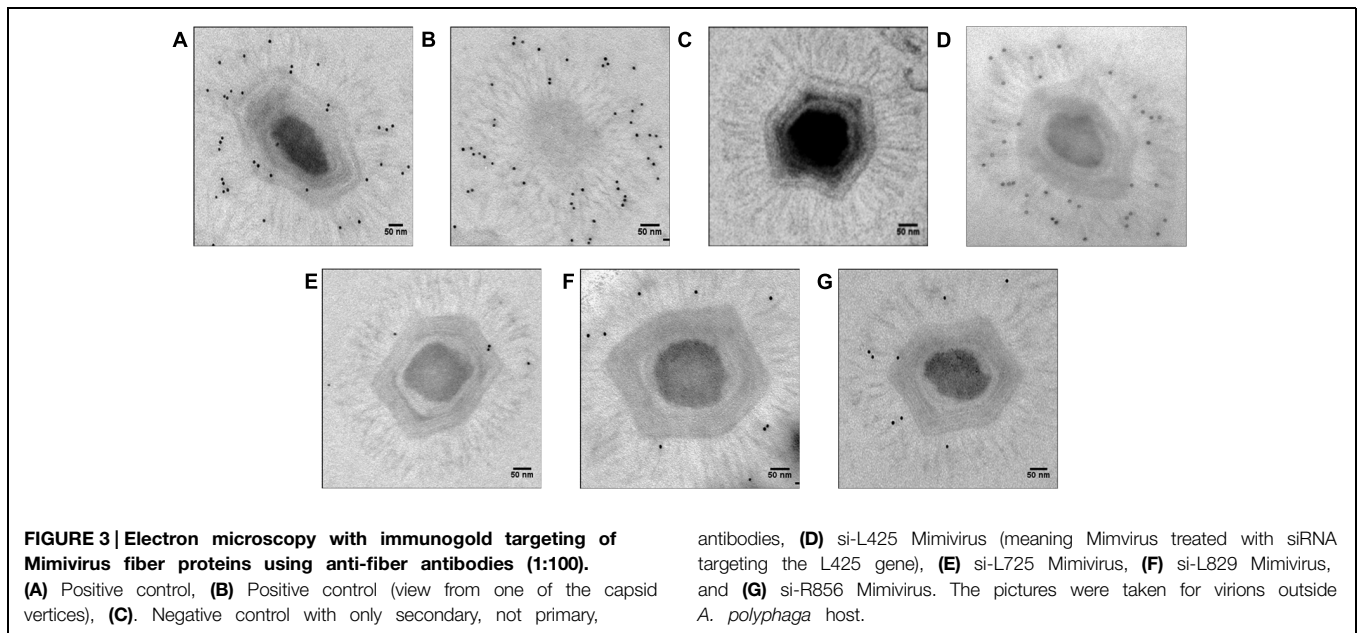
Additionally, Mimivirus is surrounded by fibers that are glycosylated, and usually there is a space surrounding the virion that

separates the viral particle from the surrounding intracytoplasmic content of amoebas. This space can be observed clearly for all control viruses but not for silenced viruses. These findings indicate that (i) the presence of short or sparse fibers after siRNA treatment was due to the siRNAs, and not to lipid carrier or suboptimal experimental conditions, and that these siRNAs were

TABLE 3 | Number of gold nanoparticles per virus in immunogold.

Type	Number of viruses observed	Total number of gold particles	Mean number of gold particles per virus \pm SD	Decrease in particles per virus (%)
Control 1 (no siRNA)	30	880	29 \pm 7	0
si-L829	48	611	13 \pm 6	57
si-R856	44	249	6 \pm 3	81
si-L725	31	129	4 \pm 3	86
Control 2 (no anti-fiber antibodies)	39	0	0 \pm 0	100

The percentage of reduction measures the drop or reduction in gold particles after knocking down each gene. si-, virus after silencing the target gene.



specific to their targeted genes as we observed different characteristics of the Mimivirus fibers and their reactivity with antibodies depending on which gene was targeted by siRNAs; (ii) the R135, L725, L829, and R856 proteins are either principal elements of Mimivirus fibers or play a key role during fiber biosynthesis and can be functionally annotated as fiber associated proteins (FAPs); and (iii) the L725 and R856 proteins are major contributors to fiber formation.

Consequences of Silencing Targeted Genes on Protein Content

Western blot analyses were performed to validate these results and revealed that the reactivity of antibodies to FAPs was reduced against viruses whose genes were silenced compared to control viruses (Figures 4A,B). Thus, intensity for three bands that might correspond to R135, L725, and L829 molecular masses were reduced in silenced viruses. To confirm these results, nitrocellulose membranes were incubated with anti-Mimivirus (Figures 4C,D) and anti-L725 antibodies (Figures 4E,F), which showed reduced reactivities against silenced viruses, although some differences in reactivities were minor. These results suggest that the R135, L725, L829, and R856 proteins might be associated with Mimivirus fiber formation and changes observed in the Mimivirus fiber layer might be due to depletion of these proteins

(i.e., post-transcriptional events). 2D-gel electrophoresis western blots performed for both si-L829 and si-R856 viruses also revealed a reduction in anti-fiber antibodies bound to the targeted FAPs compared to control viruses (Figure 5). Thus, knocking down the genes encoding R135, L725, L829, and R856 protein led to some decreases in the binding of anti-fiber antibodies to these proteins, which indicate that they may play a role in fiber formation. Taken together, TEM results, immunogold and proteomic analyses showed that R135, L725, L829, and R856 proteins can be functionally annotated as FAPs.

Comparative Genomics and Protein Re-Annotation

We propose here that the R135, L725, L829, and R856 proteins can be annotated as Mimivirus FAPs, and their names can be abbreviated as FAP1, FAP2, FAP3, and FAP4, respectively.

Searching for sequence homology of the FAPs with proteins from giant virus *Megavirales* members and other organisms from public sequence databases revealed that these proteins are conserved in most of the giant viruses, but do not share sequence homology with any fiber or spike protein encoded by any virus, including adenoviruses (Tables 4 and 5; Supplementary Table S1; Figure S7). FAP2 (L725) is an ORFan only present in mimiviruses. FAP4 (R856) contains seven TPR domains and

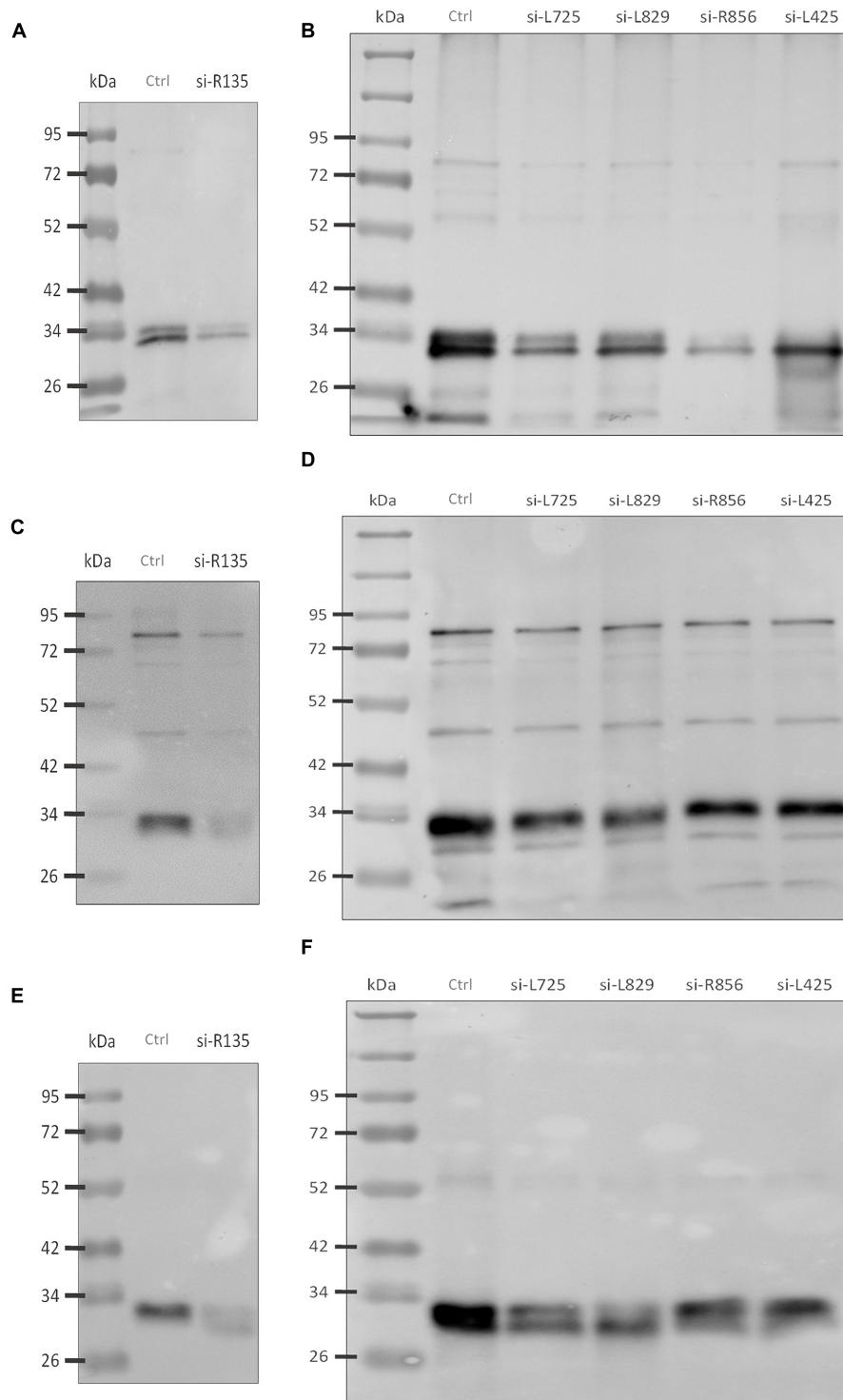
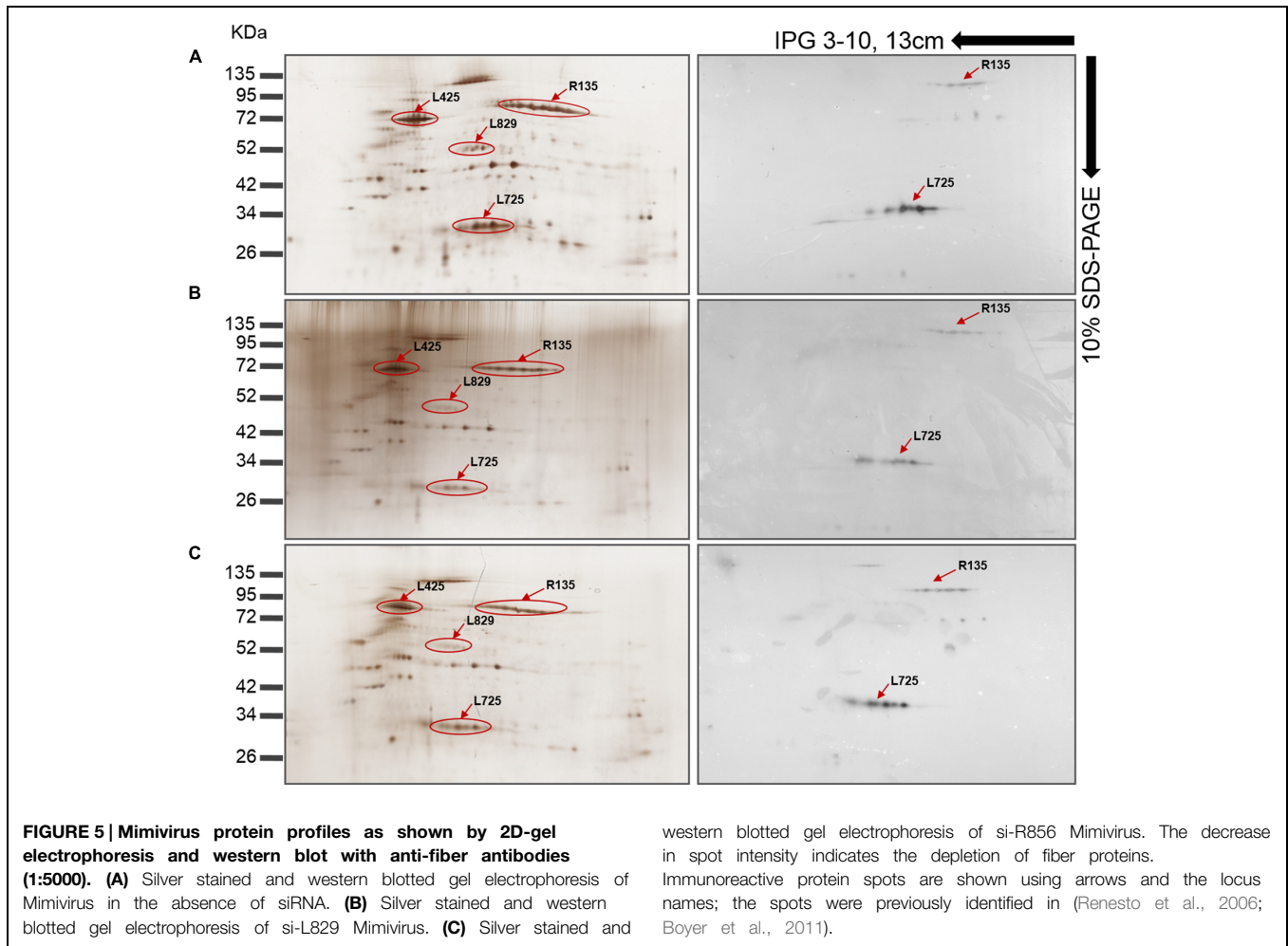


FIGURE 4 | Western blot analyses of fibers from silenced Mimivirus and control Mimivirus. (A) Reduction in anti-Mimivirus fiber antibodies (1:1000) binding to fiber proteins of si-R135 Mimivirus and **(B)** Reduction in anti-fiber antibodies (1:1000) binding to fiber proteins of si-L725, si-L829, si-R856, and si-L425 Mimivirus, **(C)** Reduction in anti-Mimivirus antibodies (1:5000) binding to si-R135 Mimivirus, and **(D)** Reduction in anti-Mimivirus antibodies (1:5000) binding to si-L725, si-L829, si-R856, and si-L425 Mimivirus,

(E) Reduction in anti-L725 antibodies (1:1000) binding to si-R135 Mimivirus, and **(F)** Reduction in anti-L725 antibodies (1:1000) binding to si-L725, si-L829, si-R856, and si-L425 Mimivirus. Figures correspond to different experiments. Ctrl indicates control Mimivirus; si- indicates the virus after silencing a target gene; molecular masses are indicated on the left; molecular masses for FAPs are as follows: L725: 27 kDa; R856: 40 kDa; L829: 50 kDa; and R135: 77 kDa.



shares sequence similarity with hypothetical proteins encoded by archaea, bacteria, choanoflagellida, ciliophora, metazoa (including rotifera, cnidaria, and hydra), and from metagenomes, but not with any protein encoded by any virus (Table 5, Supplementary Table S1; Supplementary Figures S8–S10). FAP1 (R135) shares homology with oxidoreductases and hypothetical proteins encoded by *Acanthamoeba*, metazoa, fungi, and bacteria, including proteobacteria, as well as *P. sibiricum*, and metagenomes. Finally, FAP3 (L829) is encoded by mimiviruses, marseilleviruses, *Pandoravirus* sp., and shares similarity with hypothetical proteins encoded by bacteria, and eukaryotes, including amoebozoa and fungi. Phylogenetic analyses indicate that FAP4 is widely distributed among environmental and aquatic species (Supplementary Figure S10). In addition, tree topologies suggest that FAP1 and FAP4 may have been subject to horizontal gene exchange with cellular organisms.

Taken together, these data indicate that Mimivirus FAPs are divergent from proteins that are encoded by other viruses, including *Megavirales* members other than giant viruses of amoeba, and might share a common ancestor or have been exchanged through horizontal gene transfer with proteins from cellular organisms. Moreover, with our siRNA-based strategy, we are able to functionally annotate 30 proteins from mimiviruses, as

well as re-annotate 108 proteins from metagenomic (dark matter) databases and approximately 1,000 hypothetical proteins archived in public sequence databases and encoded by archaea, bacteria, and eukaryotes (Table 5).

well as re-annotate 108 proteins from metagenomic (dark matter) databases and approximately 1,000 hypothetical proteins archived in public sequence databases and encoded by archaea, bacteria, and eukaryotes (Table 5).

Discussion

We demonstrated, using siRNA, that four proteins are involved in Mimivirus fiber formation. A disturbance in the expression of one of these proteins significantly altered the size or shape of these fibers, which indicates that these proteins are either elements of the fiber or involved in fiber formation. To our knowledge, this is the first study that described a modification of Mimivirus virions, and that used siRNA to determine the function of a Mimivirus gene.

In this article, we identified, using RNAi, the function of four proteins, including the L725 and L829 proteins with previously unknown function, a putative oxidoreductase (R135) and a TPR-containing protein (R856; Supplementary Figure S10). Comparative genomic analyses indicated that the L725 encoding gene is an ORFan, while R135, L829, and R856 are unique amongst viruses to mimiviruses, but have homologs in amoeba,

TABLE 4 | The distribution of fibers and homologs to fiber associated proteins (FAPs) among viruses that infect *Acanthamoeba* sp. and were isolated in our laboratory or by other teams.

Virus	Fibers	R135 (FAP1)	L725 (FAP2)	L829 (FAP3)	R856 (FAP4)
APMV§	Yes	+ *	+ *	+ *	+ *
APMV-M4§	No	#	+ *	#	#
ACMaV§	Yes	+ *	+ *	+ *	+ *
APLenV§	Short	+	+	+	+
APMoV§	Yes	+	+	+	
Monv§	No	+	+	+	
Goul§	No	+	+	+	+
Crdo11§	Yes	+	+	+	
Crdo7§	Yes	+	+	+	
MegCV	Yes	+	+	+	
MarsV§	No				
LauV	No				
PsV	Yes				
PdV	Yes				
Pvs					
Sputnik§	Yes				

(+) means detected as a hit in the NCBI BLASTp search, with an e-value <0.01 and identity >40%. (|) identity is 20–30%; (||) identity is 30–40%; *Reliable qPCR threshold cycle (Ct) values; #deleted gene; §the virus was isolated in our laboratory (see also Supplementary Figure S7).

APMV, *Acanthamoeba polyphaga mimivirus*; APMV-M4, *Acanthamoeba polyphaga mimivirus* isolate M4; ACPMaV, *Acanthamoeba castellanii mamavirus*; APMoV, *Acanthamoeba polyphaga moumouvirus*; CroV, *Cafeteria roenbergensis virus BV-PW1*; APLenV, *Acanthamoeba polyphaga lentillevirus*; LauV, *Lausannevirus*; MarsV, *Marseillevirus*; MegCV, *Megavirus chilensis*; Crdo11, *Courdo11 virus*; Crdo7, *Ccourdo7 virus*; Goul, *Moumouvirus goulette*; and Monv, *Moumouvirus monve*; PsV, *Pandoravirus salinus*; PdV, *Pandoravirus dulcis*; Pvs, *Pithovirus sibericum* (La Scola et al., 2008; Boyer et al., 2009, 2011; Fischer et al., 2010; Arslan et al., 2011; Colson et al., 2011; Yoosuf et al., 2012; Philippe et al., 2013).

TABLE 5 | Number of BLASTp hits corresponding to each Mimivirus fiber-associated protein.

	Mimiviridae ¹	Viruses	Archaea	Bacteria	Eukaryota	Metagenome ²
R135 (FAP1)	12	0	0	415	71	88
L725 (FAP2)	5	0	0	0	0	0
L829 (FAP3)	14	0	0	18	4	0
R856 (FAP4)	2	0	7	174	304	20

The numbers represent the number of hits per group of organisms using BLASTp against the NCBI or UniProt Archive (UniParc)² sequence databases. For details, see Supplementary Table S1. L725 gene is an ORFan and conserved only in mimiviruses. FAP means fiber associated protein.

¹Includes hits in mimiviruses and closely related viruses, including pandoraviruses and Pithovirus sibericum; ²Retrieved from BLASTp search against UniProt Archive (UniParc) database, which includes UniProt Metagenomic and Environmental Sequences (UniMES) database.

bacteria, fungi, and metazoa, and might have been exchanged by horizontal gene transfer. It is noteworthy that Mimivirus protein R856 belongs to the TPR superfamily of proteins that were reported to be involved in protein–protein interactions (Das et al., 1998; Blatch and Lassel, 1999; D'Andrea and Regan, 2003; Cortajarena and Regan, 2006), virus–host interactions (Callahan et al., 1998; Jeshtadi et al., 2010), and regulation of virus replication (Lin et al., 2012; Tani et al., 2013). We provide evidence for a new function of this protein in the formation of Mimivirus fibers.

ORFan and un-annotated genes occupy more than 50% of the gene repertoire of Mimivirus. Here, in addition to providing evidence that four proteins are FAPs by siRNA, our new strategy allowed us to re-annotate 30 proteins in mimiviruses and closely related giant viruses that share sequences homology with FAPs. In addition, this strategy opens a way to re-annotate proteins from sequence databases and genomic dark matter, as was the case here for > 500 bacterial proteins and approximately 100 proteins from metagenomes, (Table 5).

In this work, we targeted genes of Mimivirus fibers that are easy to observe and study by electron microscopy, immunogold, and proteomics. However, our strategy of silencing ORFan genes in giant viruses opens the way to identify the function of their complete gene repertoires. In particular, the proteins of giant viruses of amoeba, like those from other intracellular species, are poorly expressed and difficult to crystallize, making their functional analysis difficult. This proposed approach will lead to the annotation of hundreds of proteins without known function found in public databases and differentiate between junk DNA and truly used genes.

Author Contributions

HS, PC, BS, and DR conceived and designed the experiments; HS, BS, IP, DR, and PC analyzed the data; HS performed the experiments; and HS, PC and DR wrote the manuscript.

Acknowledgments

We are thankful to Said Azza, Jean-Pierre Baudoin, and Claude Nappez for technical help. HS received a grant from the Fondation Infectiopoie Sud.

References

- Abergel, C., Chenivresse, S., Byrne, D., Suhre, K., Arondel, V., and Claverie, J. M. (2005). *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.* 61, 212–215. doi: 10.1107/S174430910500062X
- Abergel, C., Rudinger-Thirion, J., Giege, R., and Claverie, J. M. (2007). Virus-encoded aminoacyl-tRNA synthetases: structural and functional characterization of mimivirus TyrRS and MetRS. *J. Virol.* 81, 12406–12417. doi: 10.1128/JVI.01107-07
- Aherfi, S., Boughalmi, M., Pagnier, I., Fournous, G., La Scola, B., Raoult, D., et al. (2014). Complete genome sequence of tunisivirus, a new member of the proposed family Marseilleviridae. *Arch. Virol.* 159, 2349–2358. doi: 10.1007/s00705-014-2023-5
- Arslan, D., Legendre, M., Seltzer, V., Abergel, C., and Claverie, J. M. (2011). Distant *Mimivirus* relative with a larger genome highlights the fundamental features of Megaviridae. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17486–17491. doi: 10.1073/pnas.1110889108
- Azza, S., Cambillau, C., Raoult, D., and Suzan-Monti, M. (2009). Revised *Mimivirus* major capsid protein sequence reveals intron-containing gene structure and extra domain. *BMC Mol. Biol.* 10:39. doi: 10.1186/1471-2199-10-39
- Blatch, G. L., and Lassel, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* 21, 932–939. doi: 10.1002/(SICI)1521-1878(199911)21:11<932::AID-BIES5>3.0.CO;2-N
- Boughalmi, M., Pagnier, I., Aherfi, S., Colson, P., Raoult, D., and La Scola, B. (2013a). First isolation of a *Marseillevirus* in the Diptera Syrphidae *Eristalis tenax*. *Intervirology* 56, 386–394. doi: 10.1159/000354560
- Boughalmi, M., Saadi, H., Pagnier, I., Colson, P., Fournous, G., Raoult, D., et al. (2013b). High-throughput isolation of giant viruses of the Mimiviridae and Marseilleviridae families in the Tunisian environment. *Environ. Microbiol.* 15, 2000–2007. doi: 10.1111/1462-2920.12068
- Boyer, M., Azza, S., Barrassi, L., Klose, T., Campocasso, A., Pagnier, I., et al. (2011). *Mimivirus* shows dramatic genome reduction after intraamoebal culture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10296–10301. doi: 10.1073/pnas.110118108
- Boyer, M., Gimenez, G., Suzan-Monti, M., and Raoult, D. (2010). Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses. *Intervirology* 53, 310–320. doi: 10.1159/000312916
- Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., et al. (2009). Giant *Marseillevirus* highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21848–21853. doi: 10.1073/pnas.0911354106
- Callahan, M. A., Handley, M. A., Lee, Y. H., Talbot, K. J., Harper, J. W., and Panganiban, A. T. (1998). Functional interaction of human immunodeficiency virus type 1 Vpu and Gag with a novel member of the tetratricopeptide repeat protein family. *J. Virol.* 72, 5189–5197.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334
- Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D. K., et al. (2013). “Megavirales,” a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch. Virol.* 158, 2517–2521. doi: 10.1007/s00705-013-1768-6
- Colson, P., Yutin, N., Shabalina, S. A., Robert, C., Fournous, G., La Scola, B., et al. (2011). Viruses with more than 1,000 genes: *Mamavirus*, a new *Acanthamoeba* polyphaga mimivirus strain, and reannotation of *Mimivirus* genes. *Genome Biol. Evol.* 3, 737–742. doi: 10.1093/gbe/evr048
- Cortajarena, A. L., and Regan, L. (2006). Ligand binding by TPR domains. *Protein Sci.* 15, 1193–1198. doi: 10.1110/ps.062092506

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.00345/abstract>

- D’Andrea, L. D., and Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem. Sci.* 28, 655–662. doi: 10.1016/j.tibs.2003.10.007
- Das, A. K., Cohen, P. W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* 17, 1192–1199. doi: 10.1093/emboj/17.5.1192
- Fischer, M. G., Allen, M. J., Wilson, W. H., and Suttle, C. A. (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19508–19513. doi: 10.1073/pnas.1007615107
- Forterre, P. (2010). Giant viruses: conflicts in revisiting the virus concept. *Intervirology* 53, 362–378. doi: 10.1159/000312921
- Jensen, O. N., Houthaeve, T., Shevchenko, A., Cudmore, S., Ashford, T., Mann, M., et al. (1996). Identification of the major membrane and core proteins of vaccinia virus by two-dimensional electrophoresis. *J. Virol.* 70, 7485–7497.
- Jeshtadi, A., Burgos, P., Stubbs, C. D., Parker, A. W., King, L. A., Skinner, M. A., et al. (2010). Interaction of poxvirus intracellular mature virion proteins with the TPR domain of kinesin light chain in live infected cells revealed by two-photon-induced fluorescence resonance energy transfer fluorescence lifetime imaging microscopy. *J. Virol.* 84, 12886–12894. doi: 10.1128/JVI.01395-10
- Klose, T., Kuznetsov, Y. G., Xiao, C., Sun, S., McPherson, A., and Rossmann, M. G. (2010). The three-dimensional structure of *Mimivirus*. *Intervirology* 53, 268–273. doi: 10.1159/000312911
- Kuznetsov, Y. G., Xiao, C., Sun, S., Raoult, D., Rossmann, M., and McPherson, A. (2010). Atomic force microscopy investigation of the giant mimivirus. *Virology* 404, 127–137. doi: 10.1016/j.virol.2010.05.007
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- La Scola, B., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299, 2033. doi: 10.1126/science.1081867
- La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., et al. (2008). The viroplasm as a unique parasite of the giant *Mimivirus*. *Nature* 455, 100–104. doi: 10.1038/nature07218
- Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., et al. (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4274–4279. doi: 10.1073/pnas.1320670111
- Lin, J. Y., Mendu, V., Pogany, J., Qin, J., and Nagy, P. D. (2012). The TPR domain in the host Cyp40-like cyclophilin binds to the viral replication protein and inhibits the assembly of the tombusviral replicase. *PLoS Pathog* 8:e1002491. doi: 10.1371/journal.ppat.1002491
- Moreira, D., and Brochier-Armanet, C. (2008). Giant viruses, giant chimeras: the multiple evolutionary histories of *Mimivirus* genes. *BMC Evol. Biol.* 8:12. doi: 10.1186/1471-2148-8-12
- Philippe, N., Legendre, M., Doutre, G., Coute, Y., Poirot, O., Lescot, M., et al. (2013). *Pandoraviruses*: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341, 281–286. doi: 10.1126/science.1239181
- Piacente, F., Marin, M., Molinaro, A., De Castro, C., Seltzer, V., Salis, A., et al. (2012). Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine). *J. Biol. Chem.* 287, 3009–3018. doi: 10.1074/jbc.M111.314559
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., et al. (2004). The 1.2-megabase genome sequence of *Mimivirus*. *Science* 306, 1344–1350. doi: 10.1126/science.1101485

- Raoult, D., and Forterre, P. (2008). Redefining viruses: lessons from *Mimivirus*. *Nat. Rev. Microbiol.* 6, 315–319. doi: 10.1038/nrmicro1858
- Renesto, P., Abergel, C., Decloquement, P., Moinier, D., Azza, S., Ogata, H., et al. (2006). *Mimivirus* giant particles incorporate a large fraction of anonymous and unique gene products. *J. Virol.* 80, 11678–11685. doi: 10.1128/JVI.00940-06
- Saadi, H., Pagnier, I., Colson, P., Cherif, J. K., Beji, M., Boughalmi, M., et al. (2013a). First isolation of *Mimivirus* in a patient with pneumonia. *Clin. Infect. Dis* 57, e127–e134. doi: 10.1093/cid/cit354
- Saadi, H., Reteno, D. G., Colson, P., Aherfi, S., Minodier, P., Pagnier, I., et al. (2013b). Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia. *Intervirology* 56, 424–429. doi: 10.1159/000354564
- Sobhy, H., and Colson, P. (2012). Gemi: PCR primers prediction from multiple alignments. *Comp. Funct. Genomics* 2012, 783138. doi: 10.1155/2012/783138
- Tani, J., Shimamoto, S., Mori, K., Kato, N., Moriishi, K., Matsuura, Y., et al. (2013). Ca(2+) /S100 proteins regulate HCV virus NS5A-FKBP8/FKBP38 interaction and HCV virus RNA replication. *Liver Int.* 33, 1008–1018. doi: 10.1111/liv.12151
- Xiao, C., Kuznetsov, Y. G., Sun, S., Hafenstein, S. L., Kostyuchenko, V. A., Chipman, P. R., et al. (2009). Structural studies of the giant mimivirus. *PLoS Biol.* 7:e92. doi: 10.1371/journal.pbio.1000092
- Yoosuf, N., Yutin, N., Colson, P., Shabalina, S. A., Pagnier, I., Robert, C., et al. (2012). Related giant viruses in distant locations and different habitats: *Acanthamoeba polyphaga* mousmouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. *Genome Biol. Evol.* 4, 1324–1330. doi: 10.1093/gbe/evs109

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

Copyright © 2015 Sobhy, La Scola, Pagnier, Raoult and Colson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.