Virology 452-453 (2014) 125–132

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Draft genome sequences of Terra1 and Terra2 viruses, new members of the family *Mimiviridae* isolated from soil

Niyaz Yoosuf [a], Isabelle Pagnier [a], Ghislain Fournous [a], Catherine Robert [a], Didier Raoult [a,b], Bernard La Scola [a,b], Philippe Colson [a,b,*]

[a] Aix-Marseille Univ., Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63 CNRS 7278 IRD 198 INSERM U1095, Facultés de Médecine et de Pharmacie, Marseille, France
[b] Fondation 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

ARTICLE INFO

Article history:
Received 11 September 2013
Returned to author for revisions
30 September 2013
Accepted 21 December 2013
Available online 31 January 2014

Keywords:
Terra1 virus
Terra2 virus
Mimivirus
Giant virus
"Megavirales"
Amoeba
Acanthamoeba
Soil
Environment

ABSTRACT

Since the discovery of Mimivirus, the founding member of the family *Mimiviridae*, three lineages, A–C, have been delineated among the mimiviruses of amoebae. To date, all giant viruses with annotated genomes have been isolated from water samples. Here, we describe the genome of two mimiviruses, Terra1 virus and Terra2 virus, which were recovered by co-culturing on *Acanthamoeba* spp. from soil samples. These genomes are predicted to harbor 1055 and 890 genes, respectively. Comparative genomics and phylogenomics show that Terra1 virus and Terra2 virus are classified within lineages C and A of the amoebae-associated mimiviruses, respectively. The genomic architecture of both viruses show conserved collinear central regions flanked by less conserved areas towards the extremities, when compared with other mimivirus genomes. A cluster of genes that are orthologous to bacterial genes and have no counterpart in other viral genomes except in lineage C mimiviruses was identified in Terra1 virus.

Introduction

Mimiviruses are giant viruses with particle and genome sizes that are on the same order of magnitude of small bacteria (Yutin and Koonin, 2012; Yutin et al., 2013; Colson et al., 2012). The founding member of the family *Mimiviridae* is *Acanthamoeba polyphaga* mimivirus that was discovered in 2003 by coculturing on *A. polyphaga* from a water sample collected in a cooling tower in England in 1992 (La Scola et al., 2003). The Mimivirus genome (GenBank Accession no. HQ336222) is 1.18 megabase pairs (Mbp) and is predicted to encode for approximately 1000 proteins, including proteins with functions that were believed to be the trademarks of cellular organisms, such as aminoacyl-tRNA synthetases (Raoult et al., 2004). Since the discovery of Mimivirus, several dozens other mimiviruses have been isolated from freshwater, saltwater and soil using the amoebal coculture method (La Scola et al., 2010; Boughalmi et al., 2013; Colson et al., 2013; Arslan et al., 2011; Pagnier et al., 2013). To date, five mimivirus genomes have been described in detail. Three giant viruses, namely Mimivirus, Mamavirus (another Mimivirus strain; JF801956) and Moumouvirus (JX962719), have been recovered from water collected in cooling towers in the center of England, in Paris, France; and in southern France, respectively (La Scola et al., 2003, 2008; Raoult et al., 2004; Colson et al., 2011; Yoosuf et al., 2012). Two other viruses were isolated from marine coastal water. *Megavirus chilensis* (JN258408) was recovered from water collected in Chile by culturing with *Acanthamoeba* spp. (Arslan et al., 2011), and the *Cafeteria roenbergensis* virus (Crov; GU244497) was isolated from water collected in Texas, USA, from *C. roenbergensis*, a phagocytic protist belonging to the phylum *Chromalveolata* (Fischer et al., 2010). Three lineages, A–C, have been delineated among the mimiviruses of amoebae, and the leading members of these lineages are Mimivirus, Moumouvirus and *M. chilensis*, respectively (Colson et al., 2012; Desnues et al., 2012). More distantly related mimiviruses, including Crov, infect green algae, heterokonts and haptophyta (Yutin et al., 2013). In addition, smaller, though still giant, viruses that infect *Acanthamoeba* spp. have been isolated from environmental water samples since 2008, and these viruses compose a new proposed viral family called the family

* Corresponding author at: URMITE, Faculté de Médecine et de Pharmacie, Aix-Marseille Université, 27 Boulevard Jean Moulin, 13385 Marseille CEDEX 05, France. Tel.: +33 491 324 375; fax: +33 491 387 772.
E-mail address: philippe.colson@univ-amu.fr (P. Colson).

0042-6822/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2013.12.032

"Marseilleviridae", as the first of these viruses has been named Marseillevirus (Pagnier et al., 2013; Colson et al., 2013). Mimiviruses and marseilleviruses are common in environmental water, as demonstrated by the isolation of these viruses from up to 18% of such samples and the frequent detection of sequences matching the DNA of these viruses in metagenomic studies (La Scola et al., 2010; Pagnier et al., 2013; Ghedin and Claverie, 2005; Monier et al., 2008; Kristensen et al., 2010; Colson et al., 2013). In addition, metagenomic reads matching the genomes of these giant viruses have also been detected from human samples (Colson et al., 2013). In addition, two mimiviruses of amoebae classified within lineage C, LBA111 virus and Shan virus, have recently been recovered from the bronchoalveolar fluid and the stool, respectively, of Tunisian patients presenting pneumonia (Pagnier et al., 2013; Saadi et al., 2013a; Saadi et al., 2013b).

Mimiviruses and marseilleviruses have been assigned to the nucleo-cytoplasmic large DNA viruses (NCLDVs), a monophyletic group of viruses that encompasses members of the families *Poxviridae*, *Phycodnaviridae*, *Iridoviridae*, *Ascoviridae* and *Asfarviridae* primarily based on a limited set of core genes shared by all of these viruses (Iyer et al., 2006; Raoult et al., 2004; Yutin and Koonin, 2012; Yutin et al., 2009; Iyer et al., 2001). It has recently been proposed to the International Committee on Taxonomy of Viruses that the NCLDVs should be reclassified into a new viral order called the "Megavirales" (Colson et al., 2013) (http://talk.ictvonline.org/files/proposals/taxonomy_proposals_fungal1/m/fung02/4649.aspx). Two new giant viruses, *Pandoravirus salinus* and *Pandoravirus dulcis*, were recently isolated by coculturing on *Acanthamoeba* spp. from marine sediment collected in Chile and from the mud of a freshwater pond collected in Australia, respectively (Philippe et al., 2013). These viruses have the largest particle and genome sizes among viruses. Notably, these viruses harbor 2.5 and 1.9 Mbp-long genomes, respectively. In addition, ORFans compose approximately 93% of these viral genomes, and their morphology is unique among viruses.

To date, only mimiviruses recovered from water samples have been described in detail. Here, we describe the genome of two mimiviruses recovered by coculturing with *Acanthamoeba* spp. from soil samples.

Results

Terra1 virus

The Terra1 virus was isolated from a soil sample collected in March 2009 in Marseille, France. Some of the features of this virus have been briefly described (La Scola et al., 2010).

The final assembly of the Terra1 virus genome yielded 12 contigs, including 8 large (>1500 bp) contigs and 4 small (<1500 bp) contigs with mean coverages of 17 and 10X, respectively. The Terra1 virus genome (KF527229) is a double-stranded DNA molecule composed of approximately 1233,835 bp. This genome is AT-rich (74.8%), which is similar to other mimiviruses. A total of 1055 predicted proteins were tentatively identified in this genome. In addition, the Terra1 virus genome encodes two tRNAs (1 Leu-tRNA and 1 Trp-tRNA). The predicted genes were unevenly distributed on both DNA strands, with 640 located on the negative strand and 415 located on the positive strand. The length of the Terra1 virus predicted proteins ranges from 99 to 1903 amino acids, with a mean length of 337 amino acids. A total of 1044 (99.0%) of the 1055 predicted proteins are homologous to a *M. chilensis* protein with a mean amino acid identity of 95.4%. In addition, 904 (85.7%) and 877 (82.2%) proteins from the Terra1 virus are homologous to Moumouvirus and Mimivirus proteins, respectively, with mean identities of 56.4% and 47.8%, respectively. *M. chilensis* proteins were the best hit for 409 Terra1 virus proteins. In addition, 438 best hits were proteins of the Courdo7 virus that belongs to lineage C mimiviruses (La Scola et al., 2010; Desnues et al., 2012). Also, 4 best hits were proteins of Moumouvirus monve (La Scola et al., 2010; Desnues et al., 2012) that belongs to the lineage B, and 2 best hits were proteins of Mimivirus that belongs to the lineage A, although homologs for these 6 proteins were present in mimiviruses from lineage C. Moreover, a BLASTp search against the predicted proteins from all previously published mimiviruses and against the clusters of orthologous groups of proteins (COGs) identified hits for 1047 and 292 Terra1 virus proteins, respectively. The comparative analysis of the Terra1 virus using the Proteinortho tool with other annotated mimivirus genomes showed that the Terra1 virus shares a maximum number of orthologous genes (512, 48.5% of the gene repertoire) with *M. chilensis* (lineage C). The Terra1 virus shares 433 orthologous genes (41.0%) with Moumouvirus (lineage B), 376 (35.6%) with Mimivirus and 367 (34.78%) with Mamavirus (lineage A). The *C. roenbergensis* virus, a distant member of the family *Mimiviridae*, shares 40 orthologs (3.8% of its gene content) with Terra1 virus proteins. Altogether, Mimivirus, Mamavirus, Moumouvirus and *M. chilensis* share 287 orthologous genes with the Terra1 virus. Genomic dot plots of the Terra1 virus against the amoebae-associated mimiviruses of lineages A–C showed substantial levels of synteny, which decrease close to the genome extremities. The highest level of collinearity is with *M. chilensis* (Fig. 1). The Terra1 virus genome shares a perfect collinearity with *M. chilensis*, with two inverted regions at the 5′ extremity of the genome. The genomic dot-plot of the Terra1 virus against Mimivirus shows shorter, interrupted collinear regions with a larger inverted region in the central part of the genome. The phylogeny reconstruction based on family B DNA polymerase and a concatenation of core genes of the proposed order "Megavirales" indicates that the Terra1 virus belongs to lineage C of mimiviruses of amoebae, which is congruent with results from the comparative genomic analyses (Figs. 2 and 3).

The BLASTp search against the GenBank nr database found hits for 1050 of 1055 Terra1 proteins and identified five ORFans (0.47% of the predicted gene repertoire) (Fig. 4). The Terra1 virus genome harbors 643 lineage ORFans, which represents 61% of the gene repertoire of this virus (Fig. 4). The detailed analysis of the Terra1 virus gene repertoire shows a set of genes that were either shared only by members of lineage C or that were shared by the members of lineages B and C but not lineage A. For example, Terra1_282 encodes a vacuolar sorting-associated protein that is essential for vacuolar biogenesis and maturation and is widely present in cellular organisms. This protein has been reported in *M. chilensis*. Interestingly, this gene is only present in lineage C, which suggests that the lineage C ancestor might have acquired this gene. A less parsimonious scenario is that this gene was present in the ancestor of all of the lineages and was lost in lineages A and B. In addition, the Terra1_1006 gene that encodes a Cu/Zn superoxide dismutase, the oxidoreductase enzyme that converts toxic superoxide radicals into molecular oxygen, has orthologs in all of the members of lineages B and C and in eukaryotic genomes. The distribution of this gene in mimiviruses suggests an evolutionary scenario in which this gene was present in the ancestor of all of the lineages and was lost in lineage A or that this gene was transferred in the ancestor of lineages B–C.

We have identified a set of proteins of putative bacterial origin in Terra1 virus that are contiguous and annotated as an UDP-*N*-acetylglucosamine2-epimerase, a dTDP-4-dehydrorhamnose reductase, a dTDP-*d*-glucose 4–6 dehydratase, a hypothetical protein, and an ExoV-like protein (Fig. 5). The detailed comparative analysis of this set of genes with other mimivirus genomes identified orthologs for these genes only among the members of lineage C, with no counterparts in the genomes of members

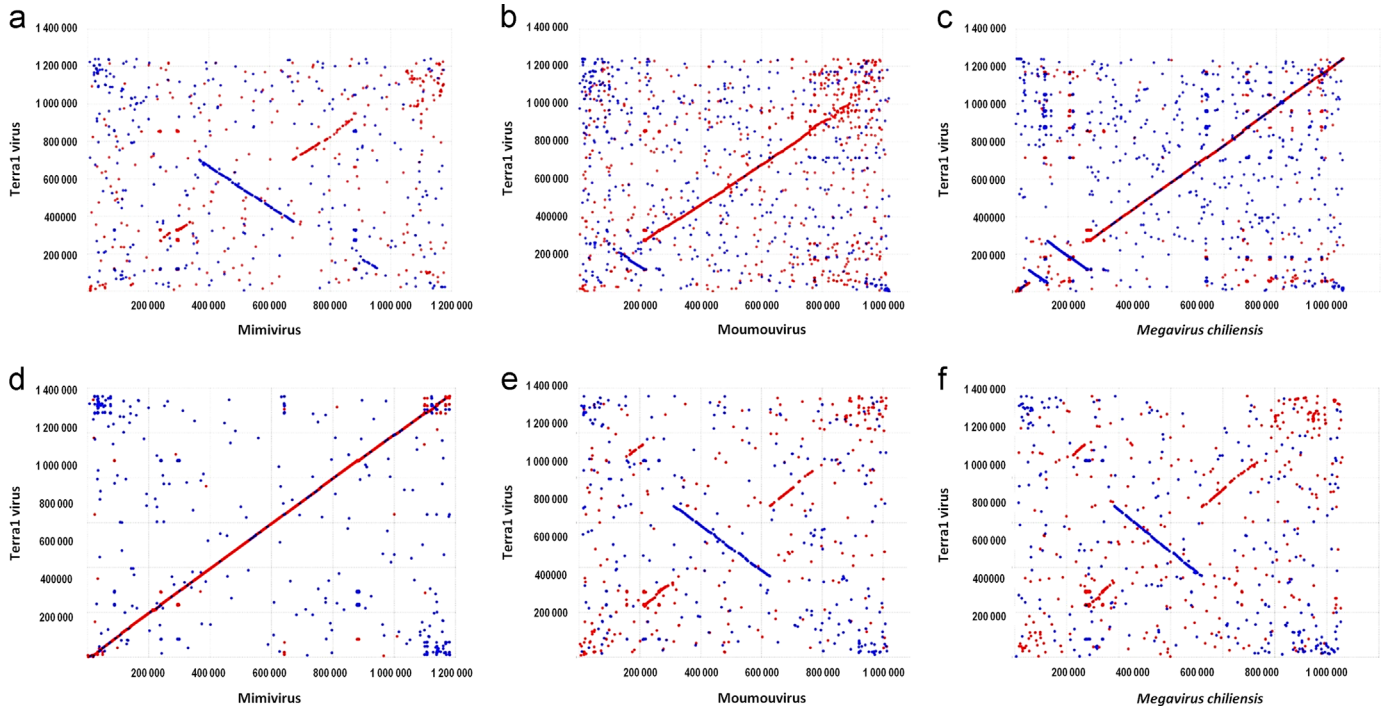

Fig. 1. Genomic dot plots of the Terra1 virus against amoebae-associated mimiviruses of lineages A–C. The Terra1 virus and Terra2 virus genomes were compared to the Mimivirus (a) and (d), Moumouvirus (b) and (e) and *M. chilensis* (c) and (f) genomes. Dot plots were constructed using the MUMmer3.22 software (Delcher et al., 2003): nucleotide-based alignments were performed with MUMmer. Dot plots were generated by the MUMmerplot script and the program gnuplot (www.gnuplot.info/docs_4.0/gnuplot.html). Aligned segments are represented by dots or lines. Colors indicate forward matches in red and reverse complement matches in blue.

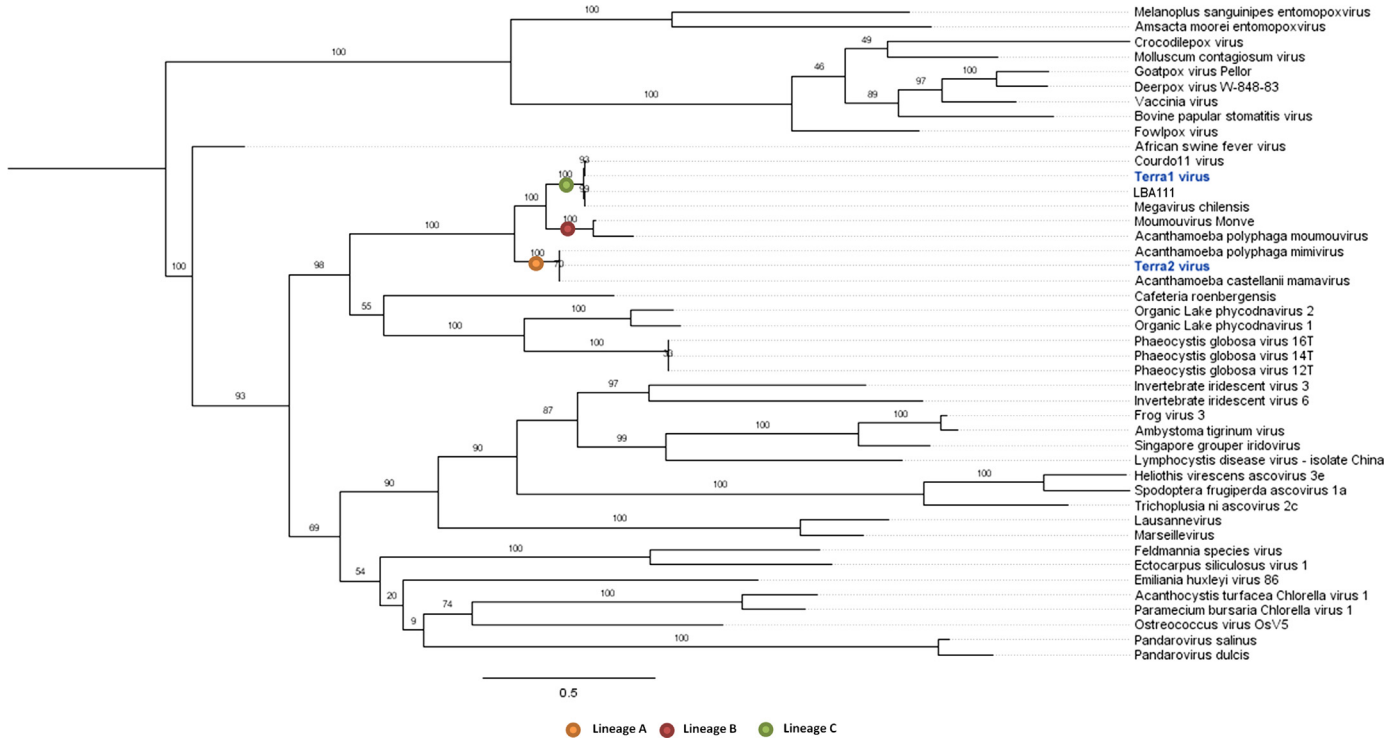

Fig. 2. Phylogenetic tree constructed based on the family B DNA polymerase from selected members of the family "Megavirales" and *P. dulcis* and *P. salinus* using the maximum likelihood method. The numbers at tree nodes indicate bootstrap replicates of 100. The sequence alignment was generated using the muscle program (Edgar, 2004), and the trimAl tool (Capella-Gutierrez et al., 2009) was used for automated alignment trimming. The phylogenic tree was constructed for 44 sequences (1441 conserved positions) using PhyML (Guindon et al., 2005) and visualized with FigTree software (http://www.umiacs.umd.edu/~morariu/figtree/).

of lineages A and B (except for the ExoV-like protein that is orthologous to mimivirus protein L143 and Mamavirus protein L199). A dTDP-4-dehydrorhamnose reductase and a dTDP-*d*-glucose 4–6 dehydratase are present in lineage A viral genomes, but these genes have different genomic locations and correspond to Mimivirus genes R141 and L780, respectively, and they are not

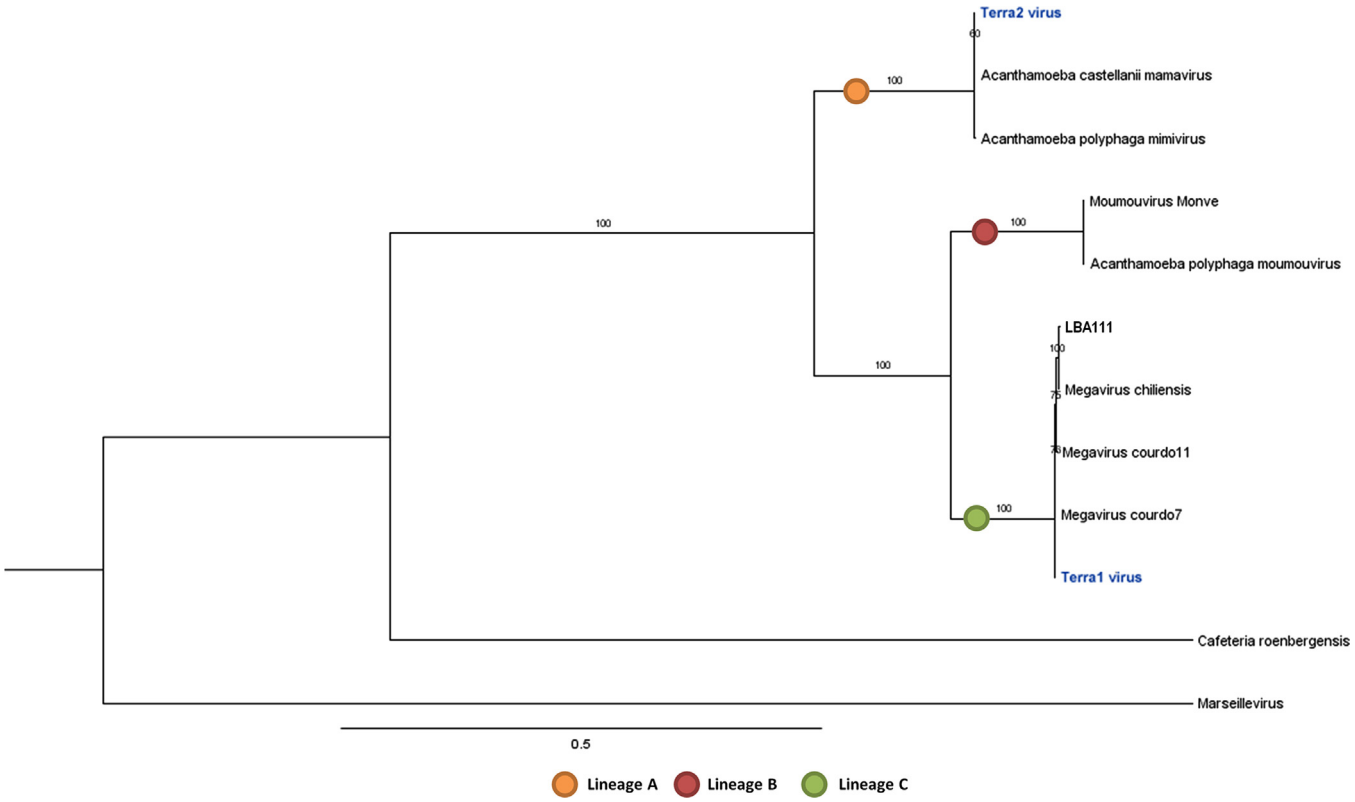

Fig. 3. Phylogenetic tree of the mimiviruses constructed from concatenated alignments of core genes of the proposed order "Megavirales". Core genes used for the phylogeny reconstruction included primase-helicase, family-B DNA polymerase, packaging ATPase and A2L-like transcription factor. Marseillevirus was used as an outgroup. The alignment included 3230 positions that were deemed reliably aligned. The bootstrap values are indicated for each internal branch.

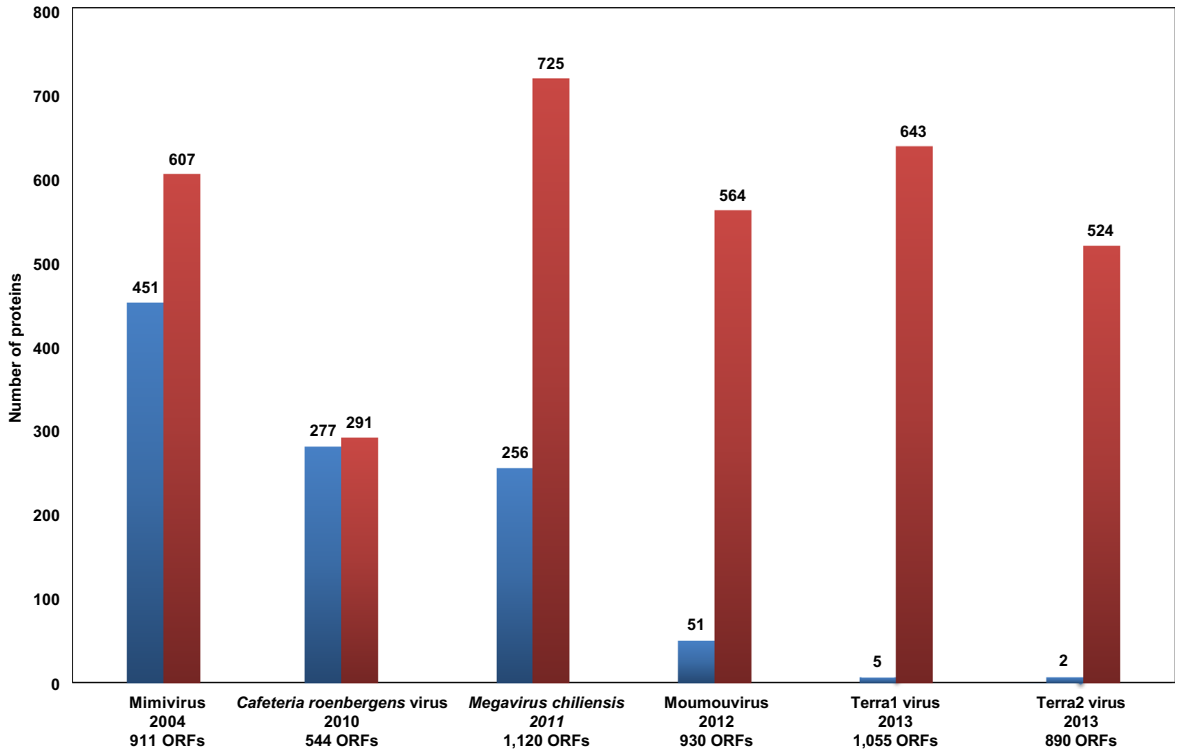

Fig. 4. The distribution of ORFans and lineage ORFans in the family *Mimiviridae* according to the timescale of the virus description. Legend for *X*-axis includes the name of the virus, the year of description, and the number of ORFs.

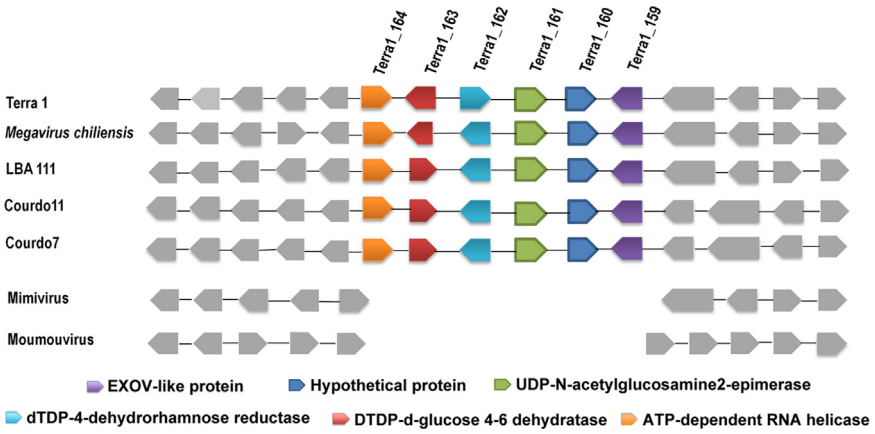

Fig. 5. Comparative gene organization and gene synteny for a region of interest in the genomes of mimiviruses. The genes and their orientation are depicted by polygons. Genes of interest across the genomes are depicted with colors. The orientation of Terra1 virus, Mimivirus and Moumouvirus genomes are reversed for the feasibility of syntenic regions.

orthologous to mimiviruses C proteins despite having the same functional annotation. In addition, we checked for the synteny of these genes in lineage C genomes and for corresponding regions in Mimivirus and Moumouvirus. The gene arrangement was conserved in the upstream and downstream regions of all the mimiviral genomes except in Moumouvirus (upstream). A similar arrangement of contiguous genes of bacterial origin with predicted functions linked to carbohydrate metabolism was previously identified in Crov (Fischer et al., 2010). Fischer et al. hypothesized that the presence of these genes in the Crov genome could be the result of frequent encounters of this mimivirus with bacteria ingested by *C. roenbergensis*, the phagocytic protistan host, inside the host cytoplasm, and the presence of virally encoded transposases that might have enabled the integration of foreign DNA into the viral genome. The Terra1 virus genome also encodes a transposase (Terra1_894) and an integrase/resolvase (Terra1_895), which might have promoted the gain of the five aforementioned proteins that are widely distributed among bacterial species by mimiviruses of lineage C. A less parsimonious evolutionary scenario could be that the genes were gained by a common mimivirus A-C ancestor, and then the genes were lost in mimivirus lineages A and B.

Finally, we compared the gene repertoire of Terra1 virus with those of *P. dulcis* and *P. salinus* using BLASTp and 1e-3 as e-value cut-off. Orthologs were identified in the gene content of *P. dulcis* and *P. salinus* for 130 and 148 Terra1 virus proteins, respectively. Pandoraviruses share several core genes of the "Megavirales" with the Terra1 virus, including the class I core genes with the exception of the capsid protein.

Terra2 virus

The Terra2 virus was also isolated in March 2009 by inoculating *A. polyphaga* from a soil sample collected in Marseille, France. The icosahedral capsid is approximately 370 nm in size and is covered by a dense layer of fibers. The Terra2 virus capsid is smaller than that of the Terra1 virus, which is 420 nm in size.

The final assembly of the Terra2 virus genome yielded 18 contigs, including five large contigs (>1500 bp) and 13 small contigs (<1500 bp) with mean coverages of 20 and 10X, respectively. The Terra2 virus genome (KF527228) is a double-stranded, AT-rich (72%) DNA molecule composed of approximately 1167,289 bp. A total of 890 predicted proteins were tentatively identified in this genome. In addition, six tRNAs (1 His-tRNA, 1 Cys-tRNA, 1 Trp-tRNA and 3 Leu-tRNA) were detected. The protein-encoding genes were mostly distributed on the negative strand with 490 genes compared to 400 genes on the positive strand. The length of the Terra2 virus predicted proteins ranges from 99 to 2945 amino acids, with a mean length of 381 amino acids. A total of 82 Terra2 virus proteins are larger than 667 amino acids in size (i.e., are encoded by genes larger than 2 kilobase pairs (kbp)), among which three proteins are larger than 2000 amino acids, namely a kinesin-like protein, a capsid protein, and a putative early transcription factor large subunit. A total of 883 (99%) of the 890 predicted Terra2 virus proteins have homologs in Mimivirus with a mean amino acid identity of 95%, 884 (99%) have homologs in Mamavirus with a mean identity of 95%, 734 (82%) have homologs in Moumouvirus with a mean identity of 48% and 769 (86%) have homologs in *M. chilensis* with a mean identity of 48%. A BLASTp search against all previously published mimivirus genomes and against COGs identified hits for 887 and 251 Terra2 virus genes, respectively. The comparative analysis of the Terra2 virus with other annotated mimivirus genomes using the Proteinortho tool yielded a maximum number of orthologs (728 (82% of the gene repertoire)) with Mimivirus, followed by 721 (81%) orthologs with Mamavirus, 423 (48%) with *M. chilensis* and 350 (39%) with Moumouvirus. In addition, 306 orthologs were shared by all these genomes, which represent 34% of the gene content of the Terra2 virus. The *C. roenbergensis* virus has 44 (5%) proteins orthologous to Terra2 virus proteins. These bidirectional best hit analyses showed that the Terra2 virus is most closely related to Mimivirus, the leading member of lineage A. Moreover, Mimivirus and Mamavirus proteins were the best hits for 662 and 182 Terra2 virus proteins, respectively. In addition, 23 of the best hits were from lentillevirus, another mimivirus of amoebae of lineage A (La Scola et al., 2010; Cohen et al., 2011). Besides, for 14 Terra2 virus predicted proteins with homologs in lineage A mimiviruses, the best hits were from mimiviruses of lineage C in three cases, from *M. chilensis*, Courdo11 virus and Courdo7 virus in six, four and two cases, respectively, and from Moumouvirus monve, which belongs to lineage B, for two proteins. Genomic dot plots for the Terra2 virus against the amoebae-associated mimiviruses of lineages A, B and C showed a high level of collinearity with Mimivirus, and a far lower collinearity was observed with Moumouvirus and *M. chilensis* (Fig. 1). The dot-plots of the Terra2 virus against Moumouvirus and *M. chilensis* revealed shorter and interrupted collinear regions and a large inverted region located in the central part of the genome. In addition, the phylogeny reconstruction based on family B DNA polymerase and a concatenation of core genes of the proposed order "Megavirales" indicates that the Terra2 virus belongs to lineage A, which is in agreement with the comparative genomic results (Figs. 2 and 3).

BLASTp searches against the GenBank nr database found hits for 888 Terra2 proteins and identified two ORFans (0.22%) (Fig. 4). In addition, the Terra2 virus genome harbors 524 lineage ORFans, which account for 59% of the viral gene repertoire (Fig. 4). Similar to the Terra1 virus, the Terra2 virus possesses a set of genes shared only by members of a particular mimivirus lineage of amoebae. Terra2_291, a ricin-type lectin protein, only shares homologs with other members of lineage A, with an identity and a coverage higher than 90%. In addition, Terra2_822 gene encodes a probable 7-dehydrocholesterol reductase, an enzyme that catalyzes the production of cholesterol from 7-dehydrocholesterol using NADPH. This gene is present in all the members of lineages A and B but has no counterpart in any member of lineage C. These findings exemplify that gene gains and losses are common amongst the family *Mimiviridae*. Finally, we identified significant BLASTp hits in the gene content of *P. dulcis* and *P. salinus* for 125 and 129 Terra2 virus proteins, respectively, and pandoraviruses were found to share several "Megavirales" core genes with the Terra2 virus, including the class I genes shared by all of the viruses of this proposed order, with the exception of the capsid protein.

Overall, the numbers of ORFans in the Terra1 virus and Terra2 virus genomes were dramatically lower than those found for previously described mimiviruses, although it must be taken into account that these findings are influenced by the set of sequences available at time of genome description (Fig. 4). Concurrently, we determined that the pan-genome of the family *Mimiviridae*, i.e. the total of the genes specific to one mimivirus genome or shared by two or more genomes, has substantially expanded since the discovery of Mimivirus a decade ago. Indeed, the size of this pan-genome increased from 979 to 2454 genes (Fig. 6).

Discussion

The comparative analyses of the Terra1 virus and the Terra2 virus genomes indicate that these giant viruses are *bona fide* members of the family *Mimiviridae*, are related to other mimiviruses that infect *Acanthamoeba* spp. and belong to lineages C and A, respectively. In addition, these analyses confirm previous findings regarding the architecture of the mimivirus genomes, with conserved collinear central regions flanked by less conserved areas towards the extremities. These features have been previously highlighted in poxviruses and phycodnaviruses, other members of the proposed order "Megavirales" (Senkevich et al., 1997; Filee et al., 2007; McLysaght et al., 2003). The large inverted regions in the central part of the genomes and shorter collinear regions towards the tips identified by genome comparison were similar to those described during the comparisons of the Mimivirus genome with the *M. chilensis* genome or the Moumouvirus genome (Arslan et al., 2011; Yoosuf et al., 2012). In addition, this pattern is similar to X-shaped patterns observed on dot plots comparing genomes or gene contents from bacterial species, which were described to have symmetry around the replication origin and terminus and suggest large chromosomal inversions around the origin of replication (Eisen et al., 2000).

The evolutionary relationship between the Terra1 virus and the Terra2 virus, which were isolated from soil, and the other mimiviruses of amoebae, which have been isolated from water and include Mimivirus, Moumouvirus and *M. chilensis*, is an indication that mimiviruses have a wide habitat and may be found in many environmental samples inhabited by *Acanthamoeba* spp. (Yoosuf et al., 2012; Colson et al., 2013; Pagnier et al., 2013).

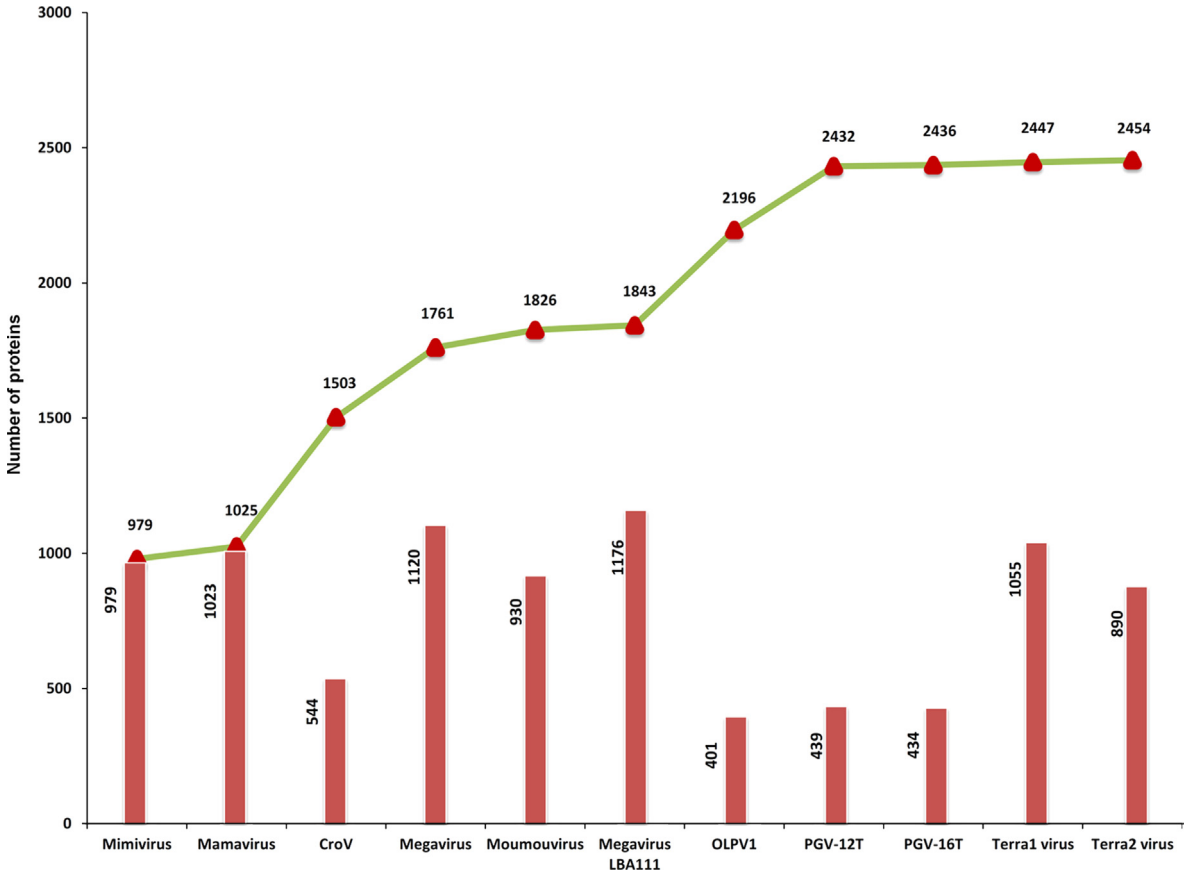

Fig. 6. The number of genes composing the pan-genome of the mimiviruses, according to the time of description of the genomes of these giant viruses. The pan-genome is the total of the genes specific to one mimivirus genome or shared by two or more genomes. Crov, *Cafeteria roenbergensis* virus; OLPV1, Organic lake phycodnavirus 1; PGV, *Phaeocystis globosa* virus.

Acanthamoeba spp., the known hosts of mimiviruses, are phagocytic protists classified in the phylum *Amoebozoa* and are predominant among the organisms in soil and water (Barker and Brown, 1994; Moliner et al., 2010; Thomas and Greub, 2010). These free living amoebae can ingest any particle with a size greater than 0.5 μm at the trophozoite stage and are known to graze on multiple organisms and microorganisms including bacteria, yeasts, fungi, viruses and algae. Therefore, these amoeba engulf large amounts of foreign DNA (Rodriguez-Zaragoza, 1994; Barker and Brown, 1994; Horn and Wagner, 2004). Moreover, *Acanthamoeba* spp. can resist various unfavorable conditions by differentiating into cysts (Raoult and Boyer, 2010; Bertelli and Greub, 2012). Interestingly, Mimivirus-like particles were observed by light microscopy within *Acanthamoeba* spp. in treated sewage sludge from a wastewater treatment plant in the UK (Gaze et al., 2011). This finding suggested that amoebae could promote the dissemination of mimiviruses to land.

As in previous studies that analyzed the genomes of giant viruses, some evidence of lateral gene transfers between these genomes and genomes of other organisms from other branches of life have been found (Raoult et al., 2004; Boyer et al., 2009; Filee, 2009). These transfers can be related to the particular lifestyle of mimiviruses within the amoebae where the viruses live sympatrically with other viruses and bacteria and can exchange genes with these viruses, bacteria and the eukaryotic host (Raoult and Boyer, 2010; Moliner et al., 2010; Bertelli and Greub, 2012). The identification in the genome of the Terra1 virus of a cluster of genes that are orthologous to bacterial genes and have no counterpart in other viral genomes than those of mimiviruses of lineage C exemplifies this capability to acquire genes. Such clusters of bacterial genes have been observed in the Mimivirus genome (Filee et al., 2007) and the Crov genome (Fischer et al., 2010). In the case of Crov, which infects phagocytic protists other than *Acanthamoeba* spp., a 38-kbp genomic fragment was identified that encompasses 34 predicted genes, among which 14 were most similar to bacterial genes, and 7 were predicted to be involved in carbohydrate metabolism. Strikingly, the cluster of genes identified in the Terra1 virus genome also contains genes involved in carbohydrate metabolism.

Since the discovery of Mimivirus, the founding member of the family *Mimiviridae*, the pan-genome of this viral family has shown a 2.5-fold expansion. Only five and two ORFans were identified among the predicted genes from the Terra1 and Terra2 viruses, respectively. Concurrently, the size of the pan-genome for mimiviruses of amoebae reached a plateau and could be considered closed based on currently available genomes. This finding may rely on the fact that these giant viruses were isolated through the same strategy of co-culturing the viruses with *A. polyphaga* or *A. castellanii*, and the majority of the amoebae were from water or soil samples, although these samples were collected from various geographical areas on three continents (La Scola et al., 2010; Pagnier et al., 2013; Colson et al., 2013). Notwithstanding, it has been recently shown that the family *Mimiviridae* is expanding through the reclassification of viruses formerly classified as phycodnaviruses, and these viruses infect different hosts than *Acanthamoeba* spp. (Yutin et al., 2013). Finally, the present study underlines that the lineage ORFans compose a considerable part of the gene content of mimiviruses that infect amoebae, and this observation, together with the large proportion of hypothetical proteins, stresses that much remains to be known about these viruses.

Materials and methods

The Terra1 and Terra2 viruses were isolated by inoculating *A. polyphaga*, as previously described (La Scola et al., 2010). The genomes of these viruses were sequenced by shotgun sequencing on the 454-Roche GS20 instrument (Boyer et al., 2009). Sequence reads were assembled *de novo* using the Newbler tool (Margulies et al., 2005) with 90% identity and 40 bp as overlap, then using other mimivirus genomes as references with the CLC Bio software (http://www.clcbio.com/index.php?id=28).

Open reading frames were predicted using the GeneMarkS software with default parameters (Besemer and Borodovsky, 2005). The predicted protein sequences were searched against the GenBank non-redundant protein sequence database (nr) and the database of Clusters of Orthologous Groups of proteins (COGs) using BLASTp (Tatusov et al., 2000; Altschul et al., 1990). The tRNAScanSE tool was used to search for transfer RNA genes (tRNAs) (Schattner et al., 2005). The strategy of reciprocal best BLASTp hits (Jordan et al., 2002) was performed to identify the set of orthologous genes using the Proteinortho tool (Lechner et al., 2011). An e-value below 1e-3, an amino acid identity above 30% and a sequence coverage above 70% were used to consider hits as significant. ORFans, which are ORFs that lack detectable homologs in sequence databases (Fischer and Eisenberg, 1999; Boyer et al., 2010), were identified by BLASTp against the NCBI GenBank nr database as ORFs with an e-value greater than 1e-3 considering an alignment length greater than 80 amino acids (for alignment lengths <80 amino acids, we used an e-value cut-off of 1e-05). This e-value cut-off has been used in previous studies to define ORFans (Yin and Fischer, 2008; Yin and Fischer, 2006). Lineage ORFans, which are the ORFs that have homologs in a given taxonomic rank and no outside homolog (Boyer et al., 2010), were found with the same methodology as the ORFans. The genomic architectures were compared using MUMmer, Mauve and r2cat softwares (Kurtz et al., 2004; Darling et al., 2004; Husemann and Stoye, 2010). The sequence alignments were built using the muscle program (Edgar, 2004), and the alignments were trimmed using the trimAl tool (Capella-Gutierrez et al., 2009). The phylogenetic trees were constructed using PhyML (Guindon et al., 2005) and were visualized using the FigTree software (http://www.umiacs.umd.edu/~morariu/figtree/).

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Arslan, D., Legendre, M., Seltzer, V., Abergel, C., Claverie, J.M., 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc. Nat. Acad. Sci. U.S.A. 108, 17486–17491.

Barker, J., Brown, M.R., 1994. Trojan horses of the microbial world: protozoa and the survival of bacterial pathogens in the environment. Microbiology 140 (Pt 6), 1253–1259.

Bertelli, C., Greub, G., 2012. Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. Front Cell Infect. Microbiol 2, 110.

Besemer, J., Borodovsky, M., 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 33, W451–W454.

Boughalmi, M., Saadi, H., Pagnier, I., Colson, P., Fournous, G., Raoult, D., La Scola, B., 2013. High-throughput isolation of giant viruses of the *Mimiviridae* and Marseilleviridae families in the Tunisian environment. Environ. Microbiol. 15, 2000–2007.

Boyer, M., Gimenez, G., Suzan-Monti, M., Raoult, D., 2010. Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses. Intervirology 53, 310–320.

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., Robert, C., Azza, S., Sun, S., Rossmann, M.G., Suzan-Monti, M., La Scola, B., Koonin, E.V., Raoult, D., 2009. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc. Nat. Acad. Sci. U.S.A 106, 21848–21853.

Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Cohen, G., Hoffart, L., La Scola, B., Raoult, D., Drancourt, M., 2011. Ameba-associated Keratitis, France. Emerg. Infect. Dis. 17, 1306–1308.

Colson, P., de Lamballerie, X., Fournous, G., Raoult, D., 2012. Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. Intervirology 55, 321–332.

Colson, P., de Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D.K., Cheng, X. W., Federici, B.A., Van Etten, J.L., Koonin, E.V., La Scola, B., Raoult, D., 2013. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch. Virol. 158, 2517–2521.

Colson, P., Fancello, L., Gimenez, G., Armougom, F., Desnues, C., Fournous, G., Yoosuf, N., Million, M., La Scola, B., Raoult, D., 2013. Evidence of the megavirome in humans. J. Clin. Virol. 57, 191–200.

Colson, P., Pagnier, I., Yoosuf, N., Fournous, G., La Scola, B., Raoult, D., 2013. "Marseilleviridae", a new family of giant viruses infecting amoebae. Arch. Virol. 158, 915–920.

Colson, P., Raoult, D., Pagnier, I., La Scola, B. Evidence of the Presence of Mimiviruses, Their Virophages, and Marseilleviruses in Environmental and Clinical Samples Worldwide. Google Maps Content, URMITE Laboratory. 2013. URL: ⟨http://maps.google.com/maps/ms?ie=UTF&msa=0&msid=200914559094835369589.0004beba4af112f60dcf2⟩.

Colson, P., Yutin, N., Shabalina, S.A., Robert, C., Fournous, G., La Scola, B., Raoult, D., Koonin, E.V., 2011. Viruses with more than 1000 genes: Mamavirus, a new Acanthamoeba castellanii mimivirus strain, and reannotation of mimivirus genes. Genome Biol. Evol 3, 737–742.

Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403.

Delcher, A.L., Salzberg, S.L., Phillippy, A.M., 2003. Using MUMmer to identify similar regions in large sequence sets (Unit). Curr. Protoc. Bioinf. , http://dx.doi.org/10.1002/0471250953.bi1003s00. (Chapter 10:Unit 10.3).

Desnues, C., La Scola, B., Yutin, N., Fournous, G., Robert, C., Azza, S., Jardot, P., Monteil, S., Campocasso, A., Koonin, E.V., Raoult, D., 2012. Provirophages and transpovirons as the diverse mobilome of giant viruses. Proc. Nat. Acad. Sci. U.S. A 109, 18078–18083.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797(%19).

Eisen, J.A., Heidelberg, J.F., White, O., Salzberg, S.L., 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. 1. (RESEARCH0011).

Filee, J., 2009. Lateral gene transfer, lineage-specific gene expansion and the evolution of Nucleo Cytoplasmic Large DNA viruses. J. Invertebr. Pathol 101, 169–171.

Filee, J., Siguier, P., Chandler, M., 2007. I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. Trends Genet 23, 10–15.

Fischer, D., Eisenberg, D., 1999. Finding families for genomic ORFans. Bioinformatics 15, 759–762.

Fischer, M.G., Allen, M.J., Wilson, W.H., Suttle, C.A., 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc. Nat. Acad. Sci. U.S.A 107, 19508–19513.

Gaze, W.H., Morgan, G., Zhang, L., Wellington, E.M., 2011. Mimivirus-like Particles in Acanthamoebae from Sewage Sludge. Emerg. Infect. Dis. 17, 1127–1129.

Ghedin, E., Claverie, J.M., 2005. Mimivirus relatives in the sargasso sea. Virol. J. 2, 62.

Guindon, S., Lethiec, F., Duroux, P., Gascuel, O., 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33, W557–W559.

Horn, M., Wagner, M., 2004. Bacterial endosymbionts of free-living amoebae. J. Eukaryot. Microbiol. 51, 509–514.

Husemann, P., Stoye, J., 2010. r2cat: synteny plots and comparative assembly. Bioinformatics 26, 570–571.

Iyer, L.M., Aravind, L., Koonin, E.V., 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. 75, 11720–11734.

Iyer, L.M., Balaji, S., Koonin, E.V., Aravind, L., 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. 117, 156–184.

Jordan, I.K., Rogozin, I.B., Wolf, Y.I., Koonin, E.V., 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Res. 12, 962–968.

Kristensen, D.M., Mushegian, A.R., Dolja, V.V., Koonin, E.V., 2010. New dimensions of the virus world discovered through metagenomics. Trends Microbiol. 18, 11–19.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5, R12.

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.M., Raoult, D., 2003. A giant virus in amoebae. Science 299, 2033.

La Scola, B., Campocasso, A., N′Dong, R., Fournous, G., Barrassi, L., Flaudrops, C., Raoult, D., 2010. Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. Intervirology 53, 344–353.

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E., Raoult, D., 2008. The virophage as a unique parasite of the giant mimivirus. Nature 455, 100–104.

Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P.F., Prohaska, S.J., 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinf. 12 (124), http://dx.doi.org/10.1186/1471-2105-12-124. (124-12).

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M. L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K. E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in micro-fabricated high-density picolitre reactors. Nature 437, 376–380.

McLysaght, A., Baldi, P.F., Gaut, B.S., 2003. Extensive gene gain associated with adaptive evolution of poxviruses. Proc. Nat. Acad. Sci. U.S.A 100, 15655–15660.

Moliner, C., Fournier, P.E., Raoult, D., 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS Microbiol. Rev. 34, 281–294.

Monier, A., Larsen, J.B., Sandaa, R.A., Bratbak, G., Claverie, J.M., Ogata, H., 2008. Marine mimivirus relatives are probably large algal viruses. Virol. J. 5, 12.

Pagnier, I., Reteno, D.G., Saadi, H., Boughalmi, M., Gaia, M., Slimani, M., Ngounga, T., Bekliz, M., Colson, P., Raoult, D., La Scola, B., Pagnier, I., Raoult, D., La Scola, B., 2013. A decade of improvements in *Mimiviridae* and Marseilleviridae isolation from amoeba.A collection of giant viruses from amoebae. Intervirology 56, 354–363.

Philippe, N., Legendre, M., Doutre, G., Coute, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., Garin, J., Claverie, J.M., Abergel, C., 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341, 281–286. (%19).

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.M., 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350.

Raoult, D., Boyer, M., 2010. Amoebae as genitors and reservoirs of giant viruses. Intervirology 53, 321–329.

Rodriguez-Zaragoza, S., 1994. Ecology of free-living amoebae. Crit. Rev. Microbiol. 20, 225–241.

Saadi, H., Ikanga Reteno, D.G., Colson, P., Aherfi, S., Minodier, P., Pagnier, I., Fenollar, F., Robert, C., Raoult, D., La Scola, B., 2013a. Shan virus, isolation of a new Mimivirus from the stool of a Tunisian patient with pneumonia. Intervirology 56, 424–429.

Saadi, H., Pagnier, I., Colson, P., KanounCherif, J., Beji, M., Boughalmi, M., Azza, S., Armstrong, N., Robert, C., Fournous, G., LaScola, B., Raoult, D., Saadi, et al., 2013b. First isolation of Mimivirus in a patient with pneumonia. Clin. Infect Dis. 57, e127–134.

Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, W686–W689.

Senkevich, T.G., Koonin, E.V., Bugert, J.J., Darai, G., Moss, B., 1997. The genome of molluscum contagiosum virus: analysis and comparison with other poxviruses. Virology 233, 19–42.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28, 33–36.

Thomas, V., Greub, G., 2010. Amoeba/amoebal symbiont genetic transfers: lessons from giant virus neighbours. Intervirology 53, 254–267.

Yin, Y., Fischer, D., 2006. On the origin of microbial ORFans: quantifying the strength of the evidence for viral lateral transfer. BMC Evol. Biol. 6, 63.

Yin, Y., Fischer, D., 2008. Identification and investigation of ORFans in the viral world. BMC Genomics 9, 24.

Yoosuf, N., Yutin, N., Colson, P., Shabalina, S.A., Pagnier, I., Robert, C., Azza, S., Klose, T., Wong, J., Rossmann, M.G., La Scola, B., Raoult, D., Koonin, E.V., 2012. 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 4, 1324–1330.

Yutin, N., Colson, P., Raoult, D., Koonin, E.V., 2013. *Mimiviridae*: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. Virol. J. 10, 106.

Yutin, N., Koonin, E.V., 2012. Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes. Virol. J. 9, 161.

Yutin, N., Wolf, Y.I., Raoult, D., Koonin, E.V., 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol. J. 6, 223.