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

Niyaz Yoosuf[1,†], Natalya Yutin[2,†], Philippe Colson[1,3,†], Svetlana A. Shabalina[2], Isabelle Pagnier[1], Catherine Robert[1], Said Azza[1], Thomas Klose[4], Jimson Wong[4], Michael G. Rossmann[4], Bernard La Scola[1,3], Didier Raoult[1,3], and Eugene V. Koonin[2,*]

[1]Aix-Marseille University, URMITE, Faculté de Médecine et de Pharmacie, Marseille, France

[2]National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, Bethesda, Maryland

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

[4]Department of Biological Sciences, Purdue University

[†]The authors contributed equally to this work.

*Corresponding author: E-mail: koonin@ncbi.nlm.nih.gov.

Accepted: November 23, 2012

Data deposition: Moumouvirus genome sequence has been deposited in GenBank under the accession number JX962719.

Abstract

The 1,021,348 base pair genome sequence of the *Acanthamoeba polyphaga moumouvirus*, a new member of the *Mimiviridae* family infecting *Acanthamoeba polyphaga*, is reported. The moumouvirus represents a third lineage beside mimivirus and megavirus. Thereby, it is a new member of the recently proposed *Megavirales* order. This giant virus was isolated from a cooling tower water in southeastern France but is most closely related to *Megavirus chiliensis,* which was isolated from ocean water off the coast of Chile. The moumouvirus is predicted to encode 930 proteins, of which 879 have detectable homologs. Among these predicted proteins, for 702 the closest homolog was detected in *Megavirus chiliensis*, with the median amino acid sequence identity of 62%. The evolutionary affinity of moumouvirus and megavirus was further supported by phylogenetic tree analysis of conserved genes. The moumouvirus and megavirus genomes share near perfect orthologous gene collinearity in the central part of the genome, with the variations concentrated in the terminal regions. In addition, genomic comparisons of the *Mimiviridae* reveal substantial gene loss in the moumouvirus lineage. The majority of the remaining moumouvirus proteins are most similar to homologs from other *Mimiviridae* members, and for 27 genes the closest homolog was found in bacteria. Phylogenetic analysis of these genes supported gene acquisition from diverse bacteria after the separation of the moumouvirus and megavirus lineages. Comparative genome analysis of the three lineages of the *Mimiviridae* revealed significant mobility of Group I self-splicing introns, with the highest intron content observed in the moumouvirus genome.

Key words: moumouvirus, mimivirus, giant virus, megavirus, *Mimiviridae*, *Megavirales*, horizontal gene transfer, viral genome, nucleo-cytoplasmic large DNA viruses.

The family *Mimiviridae* consists of giant viruses that together with five previously recognized viral families and the candidate Marseilleviridae family comprise a monophyletic group of viruses known as nucleo-cytoplasmic large DNA viruses (NCLDV) (Iyer et al. 2001, 2006; Yutin and Koonin 2012). Recently, it has been proposed to combine all the NCLDV families into a new virus order tentatively named the *Megavirales* (Colson et al. 2012). The family *Mimiviridae* includes by far the largest viral genomes sequenced to date (La Scola et al. 2003; Claverie et al. 2009; Claverie and Abergel 2010). This is the only group of viruses with genomes larger than 1 megabase, which exceeds the genome size of numerous parasitic and symbiotic bacteria. The genomes of three *Mimiviridae* members have been completely sequenced

Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution 2012.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

and characterized in detail: *Acanthamoeba polyphaga mimivirus* (Raoult et al. 2004), the prototype of the family; *A. castellanii mamavirus,* which is a close relative, effectively a strain of the mimivirus (Colson et al. 2011); and *Megavirus chiliensis* that has been recently isolated from a marine environment (Arslan et al. 2011). In addition, 16 virus isolates of the family *Mimiviridae* have been identified and characterized by proteomic methods and/or partial sequencing (La Scola et al. 2010). Furthermore, marine metagenome analysis has revealed numerous homologs of mimivirus genes indicating that *Mimiviridae* is an abundant and diverse family of giant viruses whose host range remains unknown but includes organisms from habitats as different as marine water, fresh water, and soil (Monier et al. 2008; Kristensen et al. 2010; Yamada 2011).

Phylogenetic analysis of genes that are conserved in the majority of the NCLDV (Iyer, Aravind, et al. 2001; Iyer, Balaji, et al. 2006; Koonin and Yutin 2010) has shown that another giant virus that has been isolated from the marine microflagellate *Cafeteria roenbergensis* (CroV) is a distant member of the *Mimiviridae* (Fischer et al. 2010; Colson et al. 2012). In addition to genes that are shared with the other NCLDV, the giant viruses of the *Mimiviridae* family possess many genes that have not been previously detected in any viruses, in particular genes encoding components of the translation system such as aminoacyl-tRNA synthetases as well as a variety of metabolic enzymes (Raoult et al. 2004; Colson and Raoult 2010). The comparison of the mimivirus/mamavirus and the megavirus genomes has shown that only 77% of the megavirus proteins have readily detectable homologs in the mimivirus, suggestive of a large pangenome of the *Mimiviridae* (Arslan et al. 2011). Clearly, additional complete genomes of diverse members of the *Mimiviridae* are required for the characterization of this pangenome. Here, we describe the genome of another member of the *Mimiviridae* that we denoted *A. polyphaga moumouvirus*. The moumouvirus was isolated from water collected in a cooling tower but perhaps unexpectedly is most closely related to the megavirus that was identified in a marine environment.

The moumouvirus was isolated in February 2008 by inoculating *A. polyphaga*, as previously described, with water from an industrial cooling tower located in the south-east of France (La Scola et al. 2008). Some features of this virus have been briefly described previously (La Scola et al. 2010). Morphologically, the moumouvirus particles resemble the particles of other *Mimiviridae* (Klose et al. 2010; Arslan et al. 2011). The icosahedral capsid is approximately 420 nm in size and is covered by a dense layer of fibers (fig. 1). In comparison, *A. polyphaga mimivirus* and *Megavirus chiliensis* exhibit larger capsids with a diameter of approximately 500 and 520 nm, respectively (Klose et al. 2010; Arslan et al. 2011). In addition, these two mimiviruses harbor fibers that are approximately 125- and 75-nm long, respectively, whereas the size of the moumouvirus fibers is approximately 100 nm. Some of the

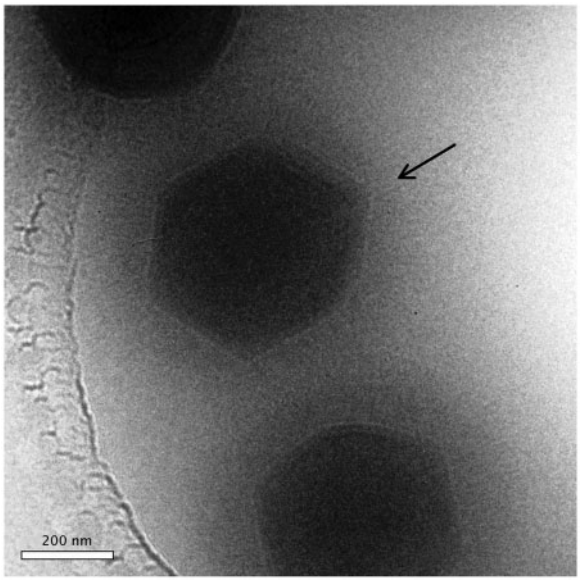

Fig. 1.—Cryo-electron micrograph of moumouvirus particles. The viral particles have a dense layer of fibers and their morphology resembles the shape of other *Mimiviridae* members, including a distinctive, starfish like vertex (arrow). Scale bar: 200 nm.

moumouvirus particles also exhibit, similar to other *Mimiviridae* members, a distinctive, starfish-like vertex (Klose et al. 2010; Arslan et al. 2011). Finally, viral factories were observed within the *A. polyphaga* cytoplasm during the replication cycle of the moumouvirus; the morphology of the moumouvirus factories is similar to that observed previously for *A. polyphaga mimivirus* and *Megavirus chiliensis* (Suzan-Monti et al. 2007; Arslan et al. 2011).

The moumouvirus genome DNA was sequenced using the 454-Roche GS20 device (Roche Diagnostics Corp., Branford, CT) (Raoult et al. 2004; Margulies et al. 2005) and then the AB SOLiD instrument (Life Technologies Corp., Carlsbad, CA). The genome assembly was performed using a combination of Roche 454 paired-end and AB SOLiD sequencing reads (supplementary methods, Supplementary Material online). The moumouvirus genome is 1,021,348 base pairs (bp) in length which is more than 200 kilobase (kb) shorter than the megavirus genome (the current record holder in viral genome size) and more than 100 kb shorter than the mimivirus and mamavirus genomes (the moumouvirus genome sequence was deposited in GenBank under the Accession Number JX962719). Using pulse-field gel electrophoresis, the moumouvirus genome was characterized as a linear DNA molecule of approximately 1 megabase (not shown). Using a combination of prediction tools (supplementary methods and file 1, Supplementary Material online), 930 open reading frames (ORFs) were identified as putative protein-coding genes, with the mean predicted protein size of 290 amino acids (aa). These ORFs are evenly distributed on both DNA strands,

with 470 predicted genes located on the "direct" strand and 460 on the "reverse" strand. The mean size of intergenic regions is 130 ± 166 nucleotides, with the predicted protein-coding density of 0.91 genes/kb (as compared with 0.89 genes/kb for the megavirus). In addition, three tRNA genes were predicted using the tRNAscan-SE method (Schattner et al. 2005). The ORFs were analyzed for evolutionary conservation, protein domain content and predicted functions by using PSI-BLAST search (Altschul et al. 1997) of the Refseq database at the NCBI (one iteration by default and up to 3 iterations when initial functional prediction was ambiguous), domain identification by RPS-BLAST search of the Conserved Domain Database (Marchler-Bauer and Bryant 2004), and assignment of proteins to clusters of orthologous NCLDV genes (NCVOGs) (Yutin et al. 2009).

Of the 930 predicted proteins of the moumouvirus, for 879 homologs were detected by protein sequence database search, and for the great majority, the most similar homolog was a megavirus protein (supplementary file 1, Supplementary Material online). For 656 predicted proteins of the moumouvirus, the megavirus homolog was a bidirectional best hit (BBH) (with the expect value cut-off of 10^{-3}); that is, a probable ortholog. The putative moumouvirus–megavirus orthologous protein pairs ranged in identity from 91% to 23%, with a median of 62%. An analogous comparison between moumouvirus and mimivirus yielded 548 putative orthologs, with a median 52% identity, indicating that the moumouvirus is more similar to the megavirus than it is to the mimivirus in terms of both the gene repertoire and sequence conservation. The evolutionary affinity of the moumouvirus and the megavirus was clearly supported by the results of phylogenetic analysis of concatenated conserved NCLDV proteins (fig. 2 and supplementary file 2, Supplementary Material online). Although the trees for individual conserved genes showed topological differences for other branches within *Phycodnaviridae* and *Mimiviridae*, the moumouvirus–megavirus clade and its monophyly with the mimivirus–mamavirus clade were invariably recovered (supplementary file 2, Supplementary Material online). In addition, a genomic dot-plot of the moumouvirus against the megavirus reveals near perfect collinearity of orthologous genes in the middle part of the genome (~650 kb), with rearrangements found only in the peripheral parts of the genomes (fig. 3*A*). This similarity of genome architectures contrasts the results of the comparison of the moumouvirus and mimivirus genomes that shows shorter, interrupted collinear regions and in addition a large inversion in the central part of the genomes (fig. 3*B*), similar to that described from the comparison of the megavirus and mimivirus genomes (Arslan et al. 2011). Conservation of the gene order in the middle of the genome with divergence at the genome ends seems to be a general feature of NCLDV evolution that was first noted in poxviruses (Senkevich et al. 1997) and has been more recently pointed out for *Chlorella phycodnaviruses* (Filee et al. 2007), marseillevirus and lausannevirus

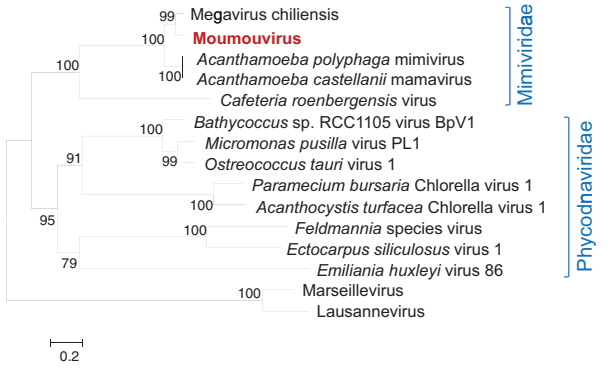

Fig. 2.—Phylogenetic tree of the *Mimiviridae* and selected *Phycodnaviridae* constructed from concatenated alignments of DNA polymerase, A32-like packaging ATPase, and A2-like Transcription Factor. Marseillevirus and lausannevirus were used as an outgroup. The alignment included 1,429 positions that were deemed reliably aligned. The bootstrap values (percentage points) are indicated for each internal branch (*Mimiviridae*).

(Thomas et al. 2011), mamavirus, mimivirus, and CroV (Boyer et al. 2011; Colson, Gimenez, et al. 2011; Colson et al. 2011).

Together, these observations indicate that megavirus and moumouvirus comprise a distinct branch of the *Mimiviridae*. Given the moderate sequence conservation between the orthologs and differences in the gene repertoire (discussed later), moumouvirus and megavirus clearly are distinct virus species unlike mimivirus and mamavirus that, at >98% mean identity between orthologous proteins and near perfect genomic collinearity, are most appropriately considered strains of the same species (Colson et al. 2011). These findings are in line with the recent demonstration of three evolutionary lineages within the *Mimiviridae* (Colson et al. 2012).

Although moumouvirus is the sister group of megavirus in the phylogenetic tree of the *Mimiviridae* (fig. 2), its genome is more than 200 kb smaller than the megavirus genome. Of the 1,120 predicted protein-coding genes of the megavirus (Arslan et al. 2011), for 464 no one-on-one ortholog has been detected in the moumouvirus. Analysis of these megavirus proteins showed that 219 are members of paralogous families common for *Mimiviridae* members; 139 are ORFans without detectable homologs; 21 apparently were acquired from sources outside the *Mimiviridae* (mainly from bacteria); and 85 are shared by megavirus and mimivirus/mamavirus but absent in the moumouvirus (supplementary file 3, Supplementary Material online). Thus, these 85 genes that are located in the terminal regions of the genome apparently have been lost in the moumouvirus lineage; an alternative, less parsimonious evolutionary scenario would involve independent acquisition of these genes in the mimivirus and megavirus lineages. Most of the genes that are inferred to have been lost by the moumouvirus are functionally uncharacterized but for some functions could be predicted, in particular in DNA repair (supplementary file 3, Supplementary Material online).

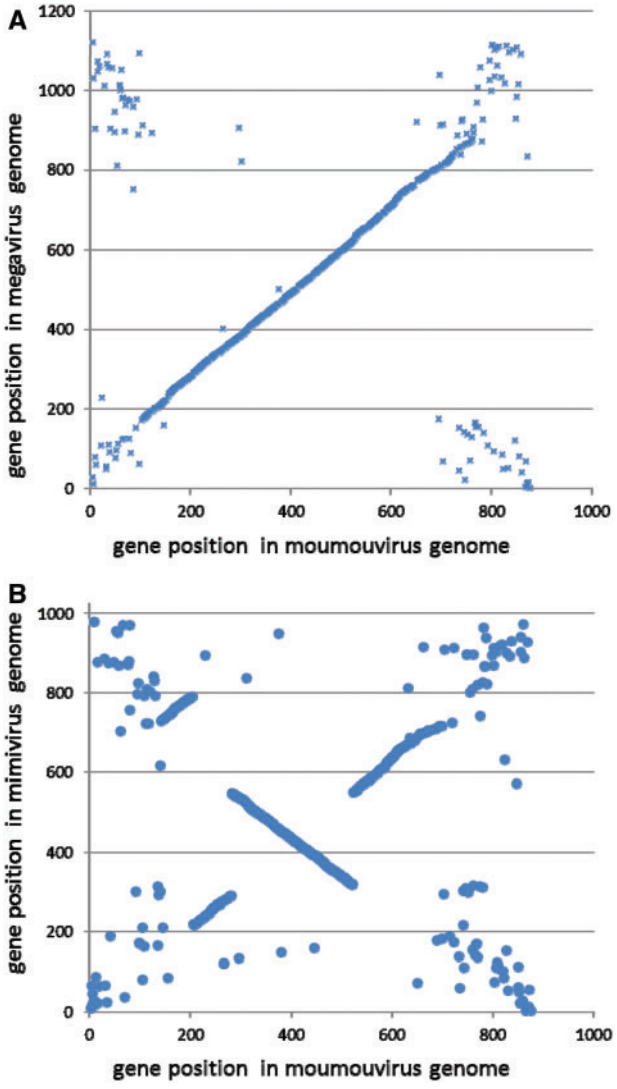

Fig. 3.—Genomic dot plots for the moumouvirus and other members of the *Mimiviridae*. (*A*) Moumouvirus versus *Megavirus chiliensis*. (*B*) Moumouvirus versus *Acanthamoeba polyphaga mimivirus*. Each point represents a pair of orthologous genes (BBHs in BLASTP searches).

Interestingly, one of the lost genes encodes the small polyA polymerase subunit/cap *O*-methyltransferase, a gene that is shared by megavirus, mamavirus, and poxviruses but is missing in the rest of the NCLDV, suggestive of multiple losses (Colson et al. 2011). The demonstration of extensive gene loss in the moumouvirus echoes the dramatic reduction in the mimivirus genome size after cultivation in germ-free amoeba; notably, the size of the terminal regions that have been eliminated from the mimivirus genome after multiple passages is approximately the same (~200 kb) as the difference in genome size between megavirus and moumouvirus (Boyer et al. 2011).

In addition to being the largest viral genome sequenced to date, the megavirus is notable for encoding the largest number of translation system components among all viruses including 7 aminoacyl-tRNA synthetases (aaRS) (Arslan et al. 2011). The moumouvirus encodes apparent orthologs of many but not all of these proteins, in particular 5 aaRS (supplementary files 1 and 3, Supplementary Material online).

The majority of the moumouvirus protein-coding genes have apparent conserved orthologs in the megavirus but the remaining genes showed some interesting evolutionary patterns. Two genes encoding metabolic enzymes, cysteine dioxygenase and NAD-dependent epimerase/dehydratase, are shared by moumouvirus and CroV to the exclusion of the other NCLDV. Phylogenetic analysis of both genes demonstrated monophyly of the two giant viruses, along with some uncharacterized environmental sequences (fig. 4*A* and *B*). This phylogeny implies the presence of these genes in the common ancestor of the *Mimiviridae* with at least two subsequent losses (in the mimivirus and megavirus branches) or the less likely evolutionary scenarios involving gene exchange between moumouvirus and CroV or independent acquisition of genes from related sources by the two viruses. Phylogenetic analysis of the moumouvirus genes with closest bacterial homologs supported the origin of these genes from diverse bacteria (two examples are shown in fig. 4*C* and *D*), in agreement with the previously noticed extensive gene exchange among symbionts and parasites of amoeba (Ogata et al. 2006; Moreira and Brochier-Armanet 2008; Boyer et al. 2009; Raoult and Boyer 2010).

Similarly to other *Mimiviridae* members, moumouvirus genes were found to contain 8 Group I self-splicing introns and three inteins (supplementary file 4, Supplementary Material online). All sequenced members of the *Mimiviridae* share an apparent ancestral intron in the gene for the largest subunit of the RNA polymerase (RNAP) and an ancestral intein in the DNA polymerase gene. The moumouvirus contains only a single intron in the major capsid protein gene, similar to the mamavirus, whereas the megavirus and the mimivirus contain two introns in this gene. In the HSP70 chaperone gene, megavirus and moumouvirus share an intron to the exclusion of the other *Mimiviridae* members but moumouvirus lacks the intron-encoded endonuclease ORF. In addition, the moumouvirus contains short inteins that consist of the cis-acting HINT protease domain alone in the genes encoding the repair ATPase MutS and an uncharacterized protein (supplementary file 4, Supplementary Material online). The positions of other introns in the genomes of the *Mimiviridae* members vary and several new introns were identified. In particular, unlike other *Mimiviridae* members, the moumouvirus contains the largest number of introns (5), along with an intein, in the second largest RNAP subunit gene (supplementary file 4, Supplementary Material online). These findings emphasize the dynamic evolution of the introns and inteins in the *Mimiviridae*.

Analysis of the mimivirus and megavirus genomes has revealed two distinct features of the transcripts, namely the

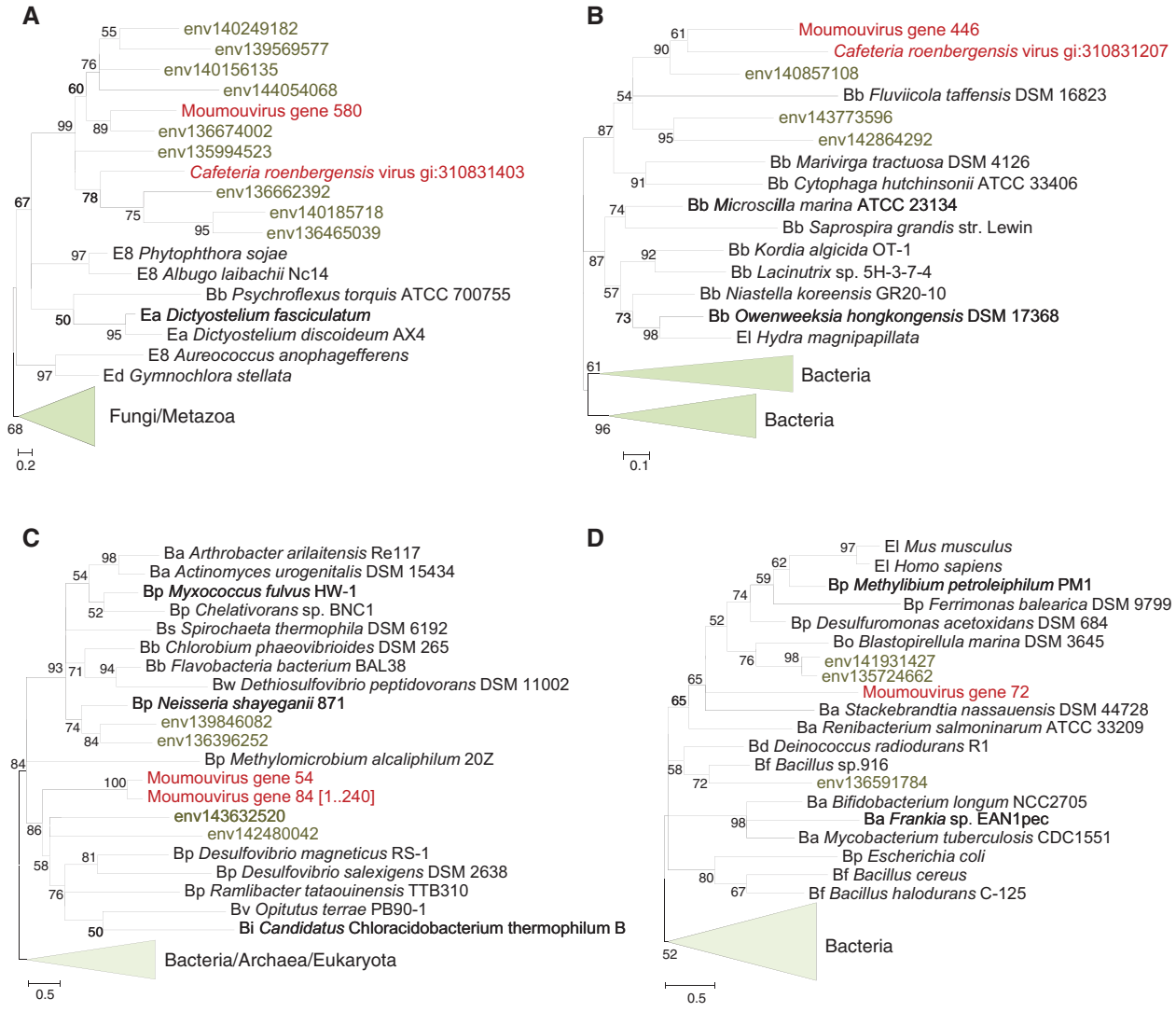

Fig. 4.—Phylogenetic trees of moumouvirus genes missing in other members of the *Mimiviridae*. (*A*) Cysteine dioxygenase, (*B*) NAD-dependent epimerase/dehydratase, (*C*) Methyltransferase, and (*D*) Nudix Hydrolase. The bootstrap values (percentage points) are indicated for each internal branch. Each sequence is denoted by taxa abbreviation and species name. GenBank Identification (GI) numbers for *Cafeteria roenbergensis* virus are shown on the trees. For other sequences, GI numbers are as follows: *Phytophthora sojae*, 348690046; *Albugo laibachii* Nc14, 325183169; *Psychroflexus torquis* ATCC 700755, 91214785; *Dictyostelium fasciculatum*, 328865331; *Dictyostelium discoideum* AX4, 66806929; *Aureococcus anophagefferens*, 323447704; *Gymnochlora stellate*, 193875832; *Fluviicola taffensis* DSM 16823, 327404160; *Marivirga tractuosa* DSM 4126, 313675758; *Cytophaga hutchinsonii* ATCC 33406, 110637252; *Microscilla marina* ATCC 23134, 124004204; *Saprospira grandis* str. Lewin, 379730028; *Kordia algicida* OT-1, 163754599; *Lacinutrix sp.* 5H-3-7-4, 336171271; *Niastella koreensis* GR20-10, 375148557; *wenweeksia hongkongensis* DSM 17368, 375013912; *Hydra magnipapillata*, 221105922; *Arthrobacter arilaitensis* Re117, 308177223; *Actinomyces urogenitalis* DSM 15434, 227495863; *Myxococcus fulvus* HW-1, 338534578; *Chelativorans sp.* BNC1, 110632790; *Spirochaeta thermophila* DSM 6192, 307718496; *Chlorobium phaeovibrioides* DSM 265, 145219574; *Flavobacteria bacterium* BAL38, 126664257; *Neisseria shayeganii* 871, 349575093; *Dethiosulfovibrio peptidovorans* DSM 11002, 288575048; *Methylomicrobium alcaliphilum* 20Z, 357407184; *Desulfovibrio magneticus* RS-1, 239906771; *Desulfovibrio salexigens* DSM 2638, 242280667; *Ramlibacter tataouinensis* TTB310, 337278313; *Opitutus terrae* PB90-1, 182415517; *Candidatus Chloracidobacterium thermophilum B*, 347756407; *Mus musculus*, 148701036; *Homo sapiens*, 344217763; *Methylibium petroleiphilum* PM1, 124266300; *Ferrimonas balearica* DSM 9799, 308049455; *Desulfuromonas acetoxidans* DSM 684, 95930587; *Blastopirellula marina* DSM 3645, 87312225; *Stackebrandtia nassauensis* DSM 44728, 291297650; *Renibacterium salmoninarum* ATCC 33209, 163840224; *Deinococcus radiodurans* R1, 6458004; *Bacillus* sp.916, 394994508; *Bifidobacterium longum* NCC2705, 23465838; *Frankia sp.* EAN1pec, 68197430; *Mycobacterium tuberculosis* CDC1551, 13879930; *Escherichia coli*, 50513417; *Bacillus cereus*, 51975946; *Bacillus halodurans* C-125, 10176350. Taxa abbreviations: Ba, *Actinobacteria*; Bb, *Bacteroidetes/Chlorobi* group; Bd, *Deinococcus-Thermus*; Bf, *Firmicutes*; Bi, *Acidobacteria*; Bo, *Planctomycetes*; Bp, *Proteobacteria*; Bs, *Spirochaetes*; Bv, *Chlamydiae/Verrucomicrobia* group; Bw, *Synergistetes*; E8, *stramenopiles*; Ea, *Amoebozoa*; Ed, *Rhizaria*; El, *Opisthokonta*.

conserved octameric motif AAAATTGA upstream of protein-coding sequences (Suhre et al. 2005) and stable hairpin structures or palindromic sequences at the 3′-ends of transcripts (Byrne et al. 2009). We searched for these two characteristic features in the moumouvirus genome. Overall, AAAATTGA sites were found within 150-nt regions upstream of the predicted start codons in 351 of the 930 predicted moumouvirus genes (37.7%). Among the orthologous genes between moumouvirus and megavirus, the fraction containing this motif was nearly the same. The AAAATTGA motif has been shown to function as an early promoter element in mimivirus and megavirus (Legendre et al. 2010). Together with the previously published comparison of megavirus and mimivirus genes containing this motif (Arslan et al. 2011), our observations on the presence of AAATTGA in upstream regions of moumouvirus genes imply that the expression pattern of orthologous genes is largely conserved among these three members of the *Mimiviridae*.

We also detected palindromic sites and predicted thermodynamically stable hairpins in the 3′ intergenic regions (150-nts downstream of the stop codon) of 725 of the 930 (78%) predicted protein-coding genes of the moumouvirus. Most of these structures are well-conserved between moumouvirus and megavirus. For example, we aligned and predicted the consensus structure of the hairpin element at the 3′-end of the major capsid gene for mimivirus, megavirus, and moumouvirus (supplementary file 5, Supplementary Material online) for which the presence of this element in the mature transcript has been experimentally validated by RNA sequencing of megavirus mg464 (Arslan et al. 2011). The polyadenylation of the megavirus gene occurs in the predicted conserved hairpin. The position of the experimentally verified polyadenylation sites is conserved between mimivirus and megavirus allowing us to predict the polyadenylation site of the moumouvirus gene (supplementary file 5, Supplementary Material online). These findings are in good agreement with the previous results demonstrating the conservation of these structural elements between mimivirus and megavirus (Arslan et al. 2011), suggesting that these hairpins function as transcription termination signals in moumouvirus.

Conclusions

Analysis of the moumouvirus genome confirms that it represents a third lineage amongst the *Mimiviridae*, in addition to those represented by mimivirus and megavirus. The moumouvirus genome further expands the pangenome of the *Mimiviridae* and emphasizes the dynamic evolution of the giant viruses, in particular extensive gene loss. The evolutionary relationship between the moumouvirus isolated from freshwater amoeba and *Megavirus chiliensis* that was isolated from a marine environment but shown to reproduce in the amoeba host of the other mimiviruses (Arslan et al. 2011) suggests that these giant viruses have a broad host range leading to ecological plasticity.

Supplementary Material

Supplementary methods and files S1–S5 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Gregory Gimenez and Ghislain Fournous for help with bioinformatic analyses and Lina Barrassi for technical assistance in the isolation of the moumouvirus. This work was supported by intramural funds of the US Department of Health and Human Services (National Library of Medicine) to N.Y., S.A.S., and E.V.K.

Literature Cited

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

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

Boyer M, 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.

Boyer M, et al. 2011. Mimivirus shows dramatic genome reduction after intraamoebal culture. Proc Natl Acad Sci U S A. 108:10296–10301.

Byrne D, et al. 2009. The polyadenylation site of mimivirus transcripts obeys a stringent "hairpin rule". Genome Res. 19:1233–1242.

Claverie JM, Abergel C. 2010. Mimivirus: the emerging paradox of quasi-autonomous viruses. Trends Genet. 26:431–437.

Claverie JM, Abergel C, Ogata H. 2009. Mimivirus. Curr Top Microbiol Immunol. 328:89–121.

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, Gimenez G, Boyer M, Fournous G, Raoult D. 2011. The giant *Cafeteria roenbergensis* virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. PLoS One 6: e18935.

Colson P, Raoult D. 2010. Gene repertoire of amoeba-associated giant viruses. Intervirology 53:330–343.

Colson P, 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.

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 MG, Allen MJ, Wilson WH, Suttle CA. 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci U S A. 107:19508–19513.

Iyer LM, Aravind L, Koonin EV. 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J Virol. 75:11720–11734.

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

Klose T, et al. 2010. The three-dimensional structure of mimivirus. Intervirology 53:268–273.

Koonin EV, Yutin N. 2010. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology 53:284–292.
Kristensen DM, Mushegian AR, Dolja VV, Koonin EV. 2010. New dimensions of the virus world discovered through metagenomics. Trends Microbiol. 18:11–19.
La Scola B, et al. 2003. A giant virus in amoebae. Science 299: 2033.
La Scola B, et al. 2008. The virophage as a unique parasite of the giant mimivirus. Nature 455:100–104.
La Scola B, et al. 2010. Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. Intervirology 53: 344–353.
Legendre M, et al. 2010. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in mimivirus. Genome Res. 20:664–674.
Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32:W327–W331.
Margulies M, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380.
Monier A, et al. 2008. Marine mimivirus relatives are probably large algal viruses. Virol J. 5:12.
Moreira D, Brochier-Armanet C. 2008. Giant viruses, giant chimeras: the multiple evolutionary histories of mimivirus genes. BMC Evol Biol. 8:12.
Ogata H, et al. 2006. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. PLoS Genet. 2:e76.
Raoult D, Boyer M. 2010. Amoebae as genitors and reservoirs of giant viruses. Intervirology 53:321–329.
Raoult D, et al. 2004. The 1.2-megabase genome sequence of mimivirus. Science 306:1344–1350.
Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33:W686–W689.
Senkevich TG, Koonin EV, Bugert JJ, Darai G, Moss B. 1997. The genome of molluscum contagiosum virus: analysis and comparison with other poxviruses. Virology 233:19–42.
Suhre K, Audic S, Claverie JM. 2005. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. Proc Natl Acad Sci U S A. 102:14689–14693.
Suzan-Monti M, La Scola B, Barrassi L, Espinosa L, Raoult D. 2007. Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga mimivirus. PLoS One 2:e328.
Thomas V, et al. 2011. Lausannevirus, a giant amoebal virus encoding histone doublets. Environ Microbiol. 13:1454–1466.
Yamada T. 2011. Giant viruses in the environment: their origins and evolution. Curr Opin Virol. 1:58–62.
Yutin N, Koonin EV. 2012. Hidden evolutionary complexity of nucleo-cytoplasmic large DNA viruses of eukaryotes. Virol J. 9:161.
Yutin N, Wolf YI, Raoult D, Koonin EV. 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol J. 6:223.

Associate editor: Emmanuelle Lerat