LETTER

doi:10.1038/nature17146

MIMIVIRE is a defence system in mimivirus that confers resistance to virophage

Anthony Levasseur[1,2]*, Meriem Bekliz[1,2]*, Eric Chabrière[1,2], Pierre Pontarotti[3], Bernard La Scola[1,2] & Didier Raoult[1,2]

Since their discovery, giant viruses have revealed several unique features that challenge the conventional definition of a virus, such as their large and complex genomes, their infection by virophages and their presence of transferable short element transpovirons[1–5]. Here we investigate the sensitivity of mimivirus to virophage infection in a collection of 59 viral strains and demonstrate lineage specificity in the resistance of mimivirus to Zamilon[6], a unique virophage that can infect lineages B and C of mimivirus but not lineage A. We hypothesized that mimiviruses harbour a defence mechanism resembling the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas system that is widely present in bacteria and archaea[7–10]. We performed *de novo* sequencing of 45 new mimivirus strains and searched for sequences specific to Zamilon in a total of 60 mimivirus genomes. We found that lineage A strains are resistant to Zamilon and contain the insertion of a repeated Zamilon sequence within an operon, here named the 'mimivirus virophage resistance element' (MIMIVIRE). Further analyses of the surrounding sequences showed that this locus is reminiscent of a defence mechanism related to the CRISPR–Cas system. Silencing the repeated sequence and the MIMIVIRE genes restores mimivirus susceptibility to Zamilon. The MIMIVIRE proteins possess the typical functions (nuclease and helicase) involved in the degradation of foreign nucleic acids. The viral defence system, MIMIVIRE, represents a nucleic-acid-based immunity against virophage infection.

Bacteria and archaea acquire immunity to invading genetic elements such as plasmids and phages through the incorporation of short sections of foreign DNA into their genomes[7]. Prokaryotic immunity covers several mechanisms including (1) prevention of viral adsorption and genome injection, (2) cleavage of the invading genome based on the self/non-self-discrimination principle and (3) blockage of phage replication[8,9]. In terms of prokaryotic immunity, the best characterized models are the restriction–modification (R–M) system and the CRISPR–Cas system[7,9,11]. The CRISPR system incorporates short fragments of DNA (21–72 nucleotides) and then uses the transcribed RNA as a guide for destroying the invading element[7]. The CRISPR system is therefore able to memorize and discriminately attack the invaders: that is, nucleic acids. The components of the CRISPR–Cas system differ broadly in terms of occurrence, sequence, number of loci and size across bacterial and archaeal genomes. CRISPRs are found in about 48% of bacteria and 80% of archaea, on the basis of the investigation of publicly available genomes[12]. The features of the CRISPR–Cas system are determined by Cas proteins, which carry diverse functional domains, such as helicase, nuclease and DNA binding motifs[8,9]. Thus far, the CRISPR–Cas system has been found in bacteria and archaea[7,10] and in only one bacteriophage[13]. In this former example, the CRISPR–Cas acquisition is used to counteract a phage inhibitory chromosomal island of the bacterial host, *Vibrio cholerae*[13]. The discovery of giant viruses living together with microbes in an amoeba-filled battlefield has challenged the traditional definition of a virus[1–3,14]. mimiviruses are visible with photonic microscopy, have a large and complex genome containing sequences transferred from other organisms[15], can be infected with viral parasites known as virophages and contain transferable short elements that resemble transposons from bacteria[4,5]. As mimiviruses behave similarly to intra-amoebal microbes[16,17], we speculated that they could also harbour several defence mechanisms in the microbial arms race, and specifically searched for a system resembling the CRISPR–Cas system.

Recently, we reported the identification of a novel virophage, Zamilon, which was found to be associated with giant viruses from the *Mimiviridae* family[6]. In the founding members of the family *Mimiviridae*, three lineages, A, B and C, have been identified among the amoebae mimiviruses. Zamilon was able to infect strains of the B (2/2) and C (2/2) lineages of mimivirus but not the two lineage A strains (0/2). Here, we infected with two virophages a collection of 59 *Acanthamoeba polyphaga* mimivirus (APMV) strains, including 28, 8 and 23 strains from the A, B and C lineages, respectively (Extended Data Fig. 1). Two virophages, Sputnik 3 (as positive control) and Zamilon, were selected for analysis and, after 24 h, an increase in Sputnik 3 DNA was observed in all the APMVs (59/59). In contrast, Zamilon was able to replicate in APMV lineages B (8/8) and C (23/23) but not in the strains from lineage A (0/28). These results confirmed and extended our initial observation that all group A strains of mimivirus are resistant to the Zamilon virophage.

As a hallmark of the CRISPR–Cas system, the acquisition of foreign DNA into the CRISPR array is a prerequisite of resistance to foreign genetic elements. Therefore, to identify potential CRISPR–Cas sequences, we performed *de novo* sequencing on 45 mimivirus strains, including lineages A (21 strains), B (5 strains) and C (19 strains). Combining these with 15 APMV genomes that were already available, we then screened all 60 APMV genomes for foreign virophage DNA sequences. A 28-nucleotide-long stretch that was identical to Zamilon DNA was found in all genomes belonging to lineage A (APMV-A) and in one single strain, the *Megavirus chilensis* strain, of the 24 different lineage C genomes (Extended Data Table 1). This sequence is located in open reading frame 4 (ORF4 encoding a protein distantly related to transposase A) of the Zamilon genome (gi|563399744) but absent in Sputnik and is integrated into mimivirus gene *R349* and the corresponding orthologous genes in all APMV-A *Mimiviridae*. The RNA predicted from the 28-nucleotide-long stretch of virophage perfectly matched the sequence of the sense strand in all APMV-A excluding the potential formation of RNA duplex. Strikingly, a 15-nucleotide-long sequence derived from this homologous sequence was repeated four times in all APMV-A genomes (28/28) but was not found in group B and C genomes (Extended Data Table 1). There was a significant correlation between Zamilon resistance and presence of the repeated Zamilon sequence in mimiviruses ($P < 0.001$). We therefore suggest that the four 15-nucleotide-long repeated sequences that were

[1]Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63, CNRS 7278, IRD 198, INSERM U1095, Marseille, France. [2]IHU Méditerranée Infection, Pôle des Maladies Infectieuses, Assistance Publique-Hôpitaux de Marseille, Faculté de Médecine, 27 Boulevard Jean Moulin, 13005 Marseille, France. [3]Aix-Marseille Université, CNRS, Centrale Marseille, I2M, UMR7373, FR 4213 - FR Eccorev 3098, équipe EBM, 13331 Marseille, France.
*These authors contributed equally to this work.

exclusively found in all APMV-A genomes are linked to resistance and immunity against Zamilon virophages. We then investigated the chromosomal environment around the repeated insertion, to identify CRISPR-like elements.

We studied the genomic environment for the presence of putative *cas* genes in the vicinity of the four 15-nucleotide repeated sequences found in all the lineage A strains, as identified by bacterial CRISPR. We found a putative phage-type endonuclease (R354) downstream of the four 15-nucleotide repeated sequence locus (Extended Data Table 2). On the basis of structural similarity searches, this protein has been modelled as a lambda exonuclease protein (36% identity), which is a relative of the Cas4 nuclease family[18]. Adjacent to the *R349* gene containing the inserted Zamilon sequence, we also identified a putative helicase domain associated with a SNF2 domain (ORF R350). This protein contains motifs that are characteristic of the Cas3 protein, which is involved in the type I bacterial CRISPR–Cas system. The R350 SNF2 domain could be involved in a variety of processes including DNA recombination, chromatin unwinding and DNA repair. We also identified a probable RNase III-encoding gene (ORF R343) localized upstream of the repeated sequences (Extended Data Table 2). In bacterial CRISPR, RNase III is responsible for CRISPR-like transcript processing. Additionally, a putative ATP-dependent DNA helicase (L364) was found downstream of the locus (Extended Data Table 2). The putative ATP-dependent DNA helicase has a multi-domain carboxy (C) terminus that includes a conserved domain from superfamily 2 (SF2), a helicase C domain and a DExD domain, as previously described for the Cas3 family.

In summary, the genomic environment in the vicinity of the four 15-nucleotide repeated sequences found in the entire A lineage contains several distant proteins reminiscent of those associated to the CRISPR–Cas system, and these proteins could play a major role in nucleic-acid-based immunity. We propose that this region of the mimivirus genome should be named MIMIVIRE, representing 'mimivirus virophage resistance element'.

A comparative model between the CRISPR–Cas system and MIMIVIRE is depicted in Extended Data Fig. 2. Important discrepancies exist between the two systems, notably in relation to the sequence-specific recognition of the invading nucleic acids, provided by the derived spacers in prokaryote and by the repeated sequences in MIMIVIRE. Contrary to the prokaryotic system in which the repeats are involved in the structural organization of the CRISPR array, MIMIVIRE is assumed to use the four-time repeated sequence inserted in an open reading frame to provide immunity against Zamilon virophage. These four repeated units appear to be essential for immunity because the presence of only one 15-nucleotide-long unit found in some B and C lineages (inserted in non-orthologous genes) did not confer resistance to Zamilon. In addition, the CRISPR system contains multiple integrated virus-derived spacers and, until now, MIMIVIRE was a priori able to target one virophage from the two known virophage strains. Investigation of forthcoming virophages could help us to unravel the MIMIVIRE system, the generality of the system and, possibly, its adaptive immune mechanism. The occurrence of MIMIVIRE was investigated in each of the APMV strains on the basis of the presence and syntenic organization of potential *cas*-related genes. These genes were conserved in all lineages of APMV-sensitive or -resistant Zamilon virophages, whereas no conservation was found with other *Megavirales* families.

To validate our hypothesis, we systematically investigated the silencing of all potential MIMIVIRE genes in mimivirus by short interfering RNA (siRNA)[19]. Consequently, we silenced all genes in the vicinity of the inserted sequence to delimitate and decipher the proteins involved in the MIMIVIRE system. A total of 27 genes were silenced and susceptibility to Zamilon infection was subsequently reported (Fig. 1c). By using quantitative PCR (qPCR), we observed an increased virophage DNA concentration after silencing the gene *R354* (encoding the endonuclease), the *R350* gene (encoding helicase and SNF2 domains), and the *R349* gene (containing the repeated insert). After 48 h, multiplication of the virophage DNA was 14-fold higher for the *R354* gene, 18-fold higher for the *R350* gene and 65-fold higher for the *R349* gene compared with the control mimivirus (Fig. 1a and Supplementary Table 1). In addition, we also combined silencing of the three MIMIVIRE genes and multiplication of the virophage DNA was 32-fold higher compared with the control. The propagation of the virophages is no higher than the unique silencing of the *R349* gene, meaning this gene containing the inserted Zamilon sequences is the central component of the MIMIVIRE system. Additionally, we also demonstrated the propagation of Zamilon virophage particles using transmission electron microscopy (Fig. 1b). No multiplication of the Zamilon virophage was observed following silencing of the other surrounding genes, as confirmed both by qPCR and by transmission electron microscopy. According to these experimental results, we delimitated the MIMIVIRE operon and demonstrated that silencing of three different MIMIVIRE genes could restore mimivirus susceptibility to Zamilon.

Nuclease and helicase activities are known to be central enzymatic functions of the prokaryotic CRISPR–Cas system, in which the Cas3 (type I CRISPR–Cas system) catalyses the unwinding and cleavage of foreign double-stranded DNA (dsDNA) and makes it possible to complete the interference process by destroying the invader nucleic acid. According to our *in silico* inference, the R354 and R350 proteins possess typical nuclease and helicase activities, respectively. To validate the function of the R350 and R354 proteins and to compare the MIMIVIRE system with the CRISPR–Cas model, the two corresponding genes were successfully overexpressed in *Escherichia coli* and the putative nuclease and helicase activities were assayed. Nuclease R354 is assumed to cleave the invading nucleic acid and, as expected, the nuclease activity of the R354 protein was evidenced by unspecific cleavage and partial degradation of dsDNA templates (Extended Data Fig. 3). Moreover, nuclease R354 was more active in the degradation of low GC per cent dsDNA templates (that is, 28–38%) than high GC per cent templates (that is, 50–55%). We found that mimiviruses and virophage (∼29%) genes were degraded but not *A. polyphaga* genes (59%) (Extended Data Fig. 3). Consequently, GC per cent cleavage specificity was in total agreement with the MIMIVIRE system immune function against virus propagation, while protecting the host organism. The R350 protein has motifs that are characteristic of helicases (SF2 superfamily) that play a central role in many aspects of the CRISPR-mediated adaptive immune systems. Helicases are known to unwind dsDNA but some helicases can rewind, or anneal, complementary strands of polynucleic acids. The annealing helicases could generate non-specific DNA hybridization and produce chimaeric aggregations of high molecular size. To determine the function of the R354 protein, we used dsDNA templates to study the unwinding/rewinding activities. We systematically observed high molecular aggregates, confirming the biochemical activity of unzipping and zipping the dsDNA, followed by aspecific hybridization of complementary sequences (Extended Data Fig. 3). These high molecular aggregates disappeared after heating and we observed a single band of the expected DNA fragment size that corresponded to the dehybridized molecules.

As demonstrated for the prokaryotic CRISPR system using Cas3 and CASCADE proteins, the helicase–nuclease R350 and nuclease R354 of the MIMIVIRE system confer central enzymatic activities that may be involved in the cleavage of foreign nucleic acid.

Its distant analogy to the bacterial CRISPR–Cas model raises the question of the origin of the MIMIVIRE system. We therefore investigated its evolutionary history by conducting a phylogenetic analysis of the experimentally validated proteins R350 and R354. In APMV, these genes were grouped together and outside their bacterial homologues and other nucleocytoplasmic large DNA viruses (Extended Data Fig. 4). This result suggests that these MIMIVIRE genes were present in the ancestors of these viruses. These two genes could also be found

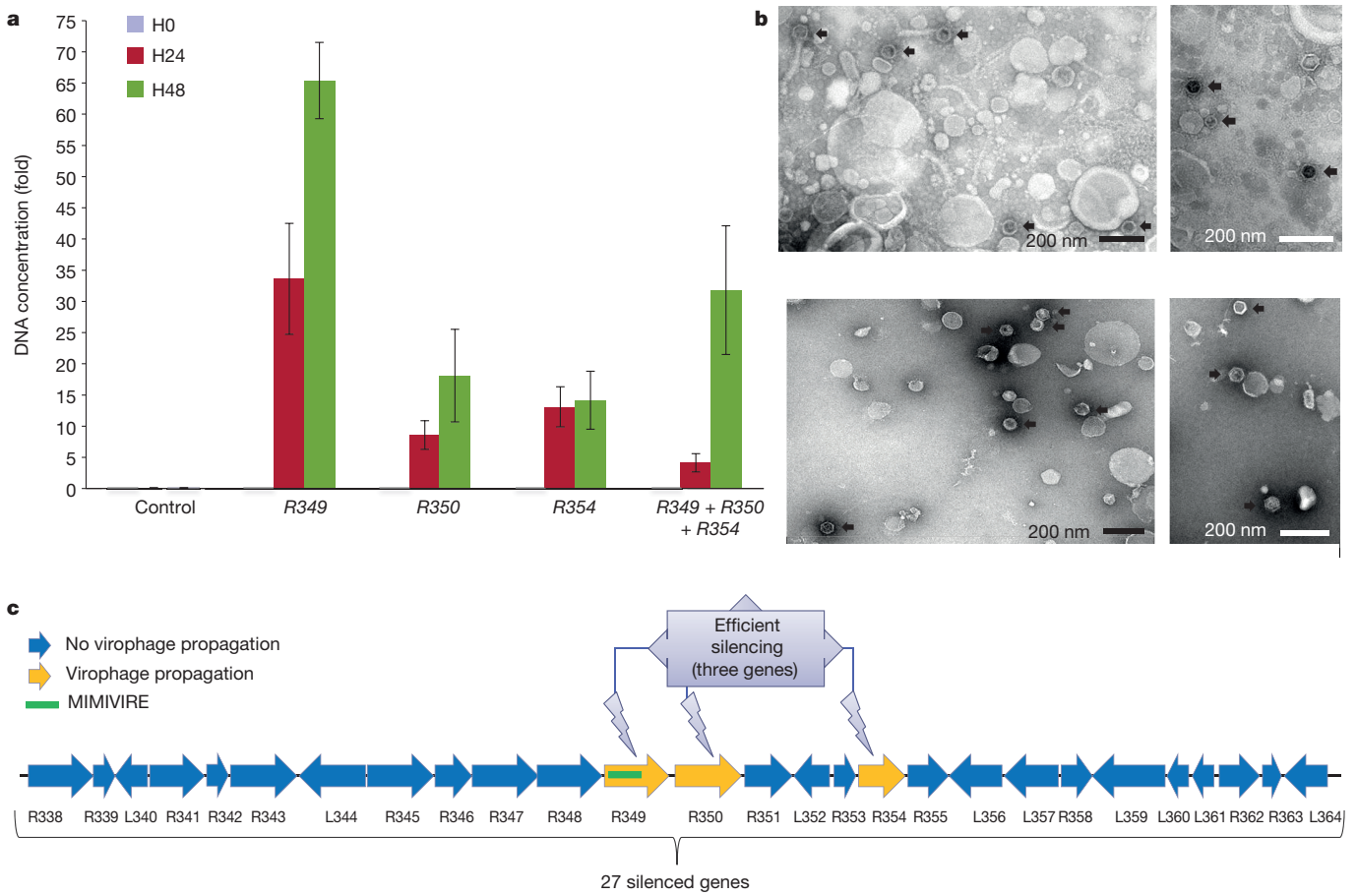


Figure 1 | Chromosomal environment of the MIMIVIRE locus of *Mimiviridae* and virophage infection. a, Quantification of Zamilon propagation after 0 h, 24 h and 48 h (H0, H24 and H48) in the wild-type mimivirus (control) and in the three silenced mimivirus strains (genes *R349*, *R350* and *R354*). The *y* axis represents the increase of the DNA concentration of Zamilon (*x*-fold) compared with the control. Mean values (±s.d.) of three independent experiments. **b**, Negative staining electron microscopy after 48 h of growth; the Zamilon virophage is identified graphically by black arrows. **c**, The 27 silenced genes are indicated with blue (no virophage infection) and yellow (virophage infection) arrows.

in many other viruses, but are scattered along the genome and their role remains to be established. Concerning the *R349* gene, no orthologous gene was retrieved in nucleoplasmic large DNA viruses, with the exception of the three APMV lineages.

Giant viruses have extraordinary features that render them unique in the viral world. We therefore tried to identify whether they may also have defence mechanisms similar to those that have been described in bacteria and archaea. We have identified sequences of foreign repeated DNA in these viruses that suggest they have also developed prokaryotic-type defence mechanisms to inhibit the genetic parasitism that they inevitably encounter in their protist hosts[20]. In this study, we identified a distant CRISPR–Cas-like mechanism called the MIMIVIRE system that explains the resistance of lineage A mimiviruses to the Zamilon virophage. We here unveil this novel immune system in giant viruses, as a result of our computational analysis as previously performed for the initial identification of the CRISPR–Cas system in prokaryotes[21]. We additionally confirmed the biological role of the MIMIVIRE system by silencing and overexpressing two of the genes that are incorporated in it. Both experimental results (silencing of MIMIVIRE genes and functional characterization of MIMIVIRE proteins) confirmed our hypothesis about the fundamental role of MIMIVIRE in the susceptibility of mimivirus to virophage infection and indicated that MIMIVIRE is a defence system against invading elements such as nucleic acids. Besides eliminating competing parasite virophages, MIMIVIRE could also function as a means of maintaining the lytic and infective capacity of the giant virus[4]. In the future, further experimental studies will be required to unravel the molecular bases of the mechanism that drives the MIMIVIRE system. Our findings illustrate that giant viruses have undergone genetic evolution that is similar to other microbes, via the incorporation of viral parasites (virophages), mobile elements (transpovirons, polintons) and lateral gene transfer[22], and that MIMIVIRE confers a nucleic-acid-based immunity in giant viruses.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 April 2015; accepted 18 January 2016.
Published online 29 February 2016.

1. La Scola, B. *et al.* A giant virus in amoebae. *Science* **299,** 2033 (2003).
2. Koonin, E. V. & Dolja, V. V. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Mol. Biol. Rep.* **78,** 278–303 (2014).
3. Forterre, P. Giant viruses: conflicts in revisiting the virus concept. *Intervirology* **53,** 362–378 (2010).
4. La Scola, B. *et al.* The virophage as a unique parasite of the giant mimivirus. *Nature* **455,** 100–104 (2008).
5. Raoult, D. How the virophage compels the need to readdress the classification of microbes. *Virology* **477,** 119–124 (2015).
6. Gaia, M. *et al.* Zamilon, a novel virophage with *Mimiviridae* host specificity. *PLoS ONE* **9,** e94923 (2014).
7. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nature Rev. Microbiol.* **13,** 722–736 (2015).
8. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* **41,** 4360–4377 (2013).
9. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526,** 55–61 (2015).

10. Barrangou, R. & Marraffini, L. A. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell* **54,** 234–244 (2014).
11. Bickle, T. A. & Kruger, D. H. Biology of DNA restriction. *Microbiol. Rev.* **57,** 434–450 (1993).
12. Pourcel, C. & Drevet, C. in *CRISPR-Cas Systems* 33–59 (Springer, 2012).
13. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494,** 489–491 (2013).
14. Raoult, D. TRUC or the need for a new microbial classification. *Intervirology* **56,** 349–353 (2013).
15. Filée, J. & Chandler, M. Gene exchange and the origin of giant viruses. *Intervirology* **53,** 354–361 (2010).
16. Boyer, M. *et al.* Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc. Natl Acad. Sci. USA* **108,** 10296–10301 (2011).
17. Raoult, D. & Boyer, M. Amoebae as genitors and reservoirs of giant viruses. *Intervirology* **53,** 321–329 (2010).
18. Zhang, J., Kasciukovic, T. & White, M. F. The CRISPR associated protein Cas4 is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster. *PLoS ONE* **7,** e47232 (2012).
19. Sobhy, H., La Scola, B., Pagnier, I., Raoult, D. & Colson, P. Identification of giant Mimivirus protein functions using RNA interference. *Front. Microbiol.* **6,** 345 (2015).
20. Slimani, M., Pagnier, I., Raoult, D. & La Scola, B. Amoebae as battlefields for bacteria, giant viruses, and virophages. *J. Virol.* **87,** 4783–4785 (2013).
21. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1,** 7 (2006).
22. Yutin, N., Raoult, D. & Koonin, E. V. Virophages, polintons, and transpovirons: a complex evolutionary network of diverse selfish genetic elements with different reproduction strategies. *Virol. J.* **10,** 158 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Benamar, A. Caputo and C. Robert for technical support, and L. Pinault for her assistance on protein production.

Author Contributions A.L., B.L. and D.R. conceived the project and designed the study and experiments; A.L., M.B., P.P., B.L. and D.R. analysed the results; A.L. and D.R. wrote the manuscript. All authors read and approved the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R. (didier.raoult@gmail.com) or B.L. (bernard.la-scola@univ-amu.fr).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Genome sequencing and bioinformatic analyses. Forty-five new *Mimiviridae* strains were isolated and subsequently sequenced using either a 454-Roche GS FLX Titanium system (Roche Diagnostics), AB SOLiD instrument (Life Technologies) or MiSeq sequencer (Illumina) (Supplementary Table 2). In detail, 24 *Mimiviridae* strains were sequenced on the MiSeq Technology (Illumina) with a paired-end and barcode strategy on different flowcells using a Nextera XT library kit (Illumina). The DNA was quantified by a Qubit assay with a high sensitivity kit (Life Technologies) and dilution was performed to require 1 ng of each sample as input. The construction of the library was performed by a 'tagmentation' step to fragment the genomic DNA, followed by limited cycle PCR amplification to complete the tag adapters and introduce dual-index barcodes. Automated cluster generation and paired-end sequencing was performed on a MiSeq instrument in a single 39-h run to 2×250 bp. The sequencing strategies of 23 *Mimiviridae* strains were performed through the SOLiD 4_Life technologies in NGS technologies. The paired-end library was constructed from 1 μg of purified genomic DNA of each strain. The sequencing was performed to 50×35 bp using SOLiD V4 chemistry on one full slide on an Applied Biosystems SOLiD 4 machine. All of these 96 genomic DNAs were barcoded with the module 1-96 barcodes provided by Life Technologies. Thirteen strains of the *Mimiviridae* paired-end library were pyrosequenced on the 454 Roche Titanium. Each project was loaded on a 1/4 region on PTP Picotiterplate. The library was constructed with 5 μg of DNA according to the 454 Titanium paired-end protocol and the manufacturer's instructions. It was mechanically fragmented on a Covaris device (KBioScience-LGC Genomics) through a miniTUBE-Red 5 kb. The library was clonally amplified in emPCR reactions with a GS Titanium SV emPCR Kit (Lib-L) version 2, then loaded on a GS Titanium PicoTiterPlates PTP Kit 70×75 sequenced with a GS Titanium Sequencing Kit XLR70 and reads generated with an average of 280 bp.

Genome assembly and structural annotation. The Newbler assembler version 2.7 and Abyss genomics version 2.3 assembler were used to assemble *Mimiviridae* genomes (Supplementary Table 2). SOLiD reads were mapped on assembled *Mimiviridae* genome using the CLC Genomics Workbench version 7.5. Gene predictions were performed using GeneMarkS software with default parameters[23].

Virophage DNA screening in APMV. The genomes of Zamilon (NC_022990), Sputnik 1 (EU606015), Sputnik 2 (NC_023846) and Sputnik 3 (NC_023847) were downloaded from the National Center for Biotechnology Information (NCBI). The genomes were fragmented into short fragments of 40 nucleotides using a sliding window of size 10 nucleotides. All fragments were blasted against the respective APMV genomes using an *e*-value threshold $= e^{-3}$.We then looked for all fragments (with 100% identity) present in the entire lineage A of APMV and mostly absent in lineage B and C. One hit, 28 nucleotides in length, fulfilled these criteria and was selected.

Phylogenetic tree construction. From each query sequence, a data set of putative homologous sequences was built by a BLAST[24] run on the NCBI non-redundant (NR) database. The raw data set was manually filtered to eliminate potentially non-homologous sequences, disturbing alignments and duplicates. Alignments were conducted using MUSCLE[25]. For phylogenetic reconstruction, we used the maximum likelihood method.

APMV and virophage production. The *A. polyphaga* Link-AP1 trophozoite strain[26] was cultured in peptone-yeast extract glucose (PYG) medium at 32 °C for 3 days, as described previously[27]. The giant viruses in our collection were co-cultured with fresh *A. polyphaga* in PYG medium. To purify the giant viruses, the co-culture was centrifuged at low speed (1,700*g* per 10 min), and the supernatant was filtered across a 0.8 μm membrane to remove residual amoebas and cysts. Each supernatant was then washed three times with Page's modified Neff's amoeba saline (PAS) by centrifugation at high speed (10,300*g* per 10 min) to pellet the virus. Sputnik 3 and Zamilon virophages were produced in co-culture with Mamavirus and Mont1, respectively, in PYG medium containing the amoeba *A. polyphaga*. After complete lysis, the supernatant that was obtained following centrifugation at high speed (10,300*g* per 10 min) was successively filtered with 0.8 μm, 0.45 μm and 0.22 μm membranes to obtain a pure virophage suspension. A final ultracentrifugation was performed at 13,900*g* for 1.5 h to concentrate each virophage filtrate.

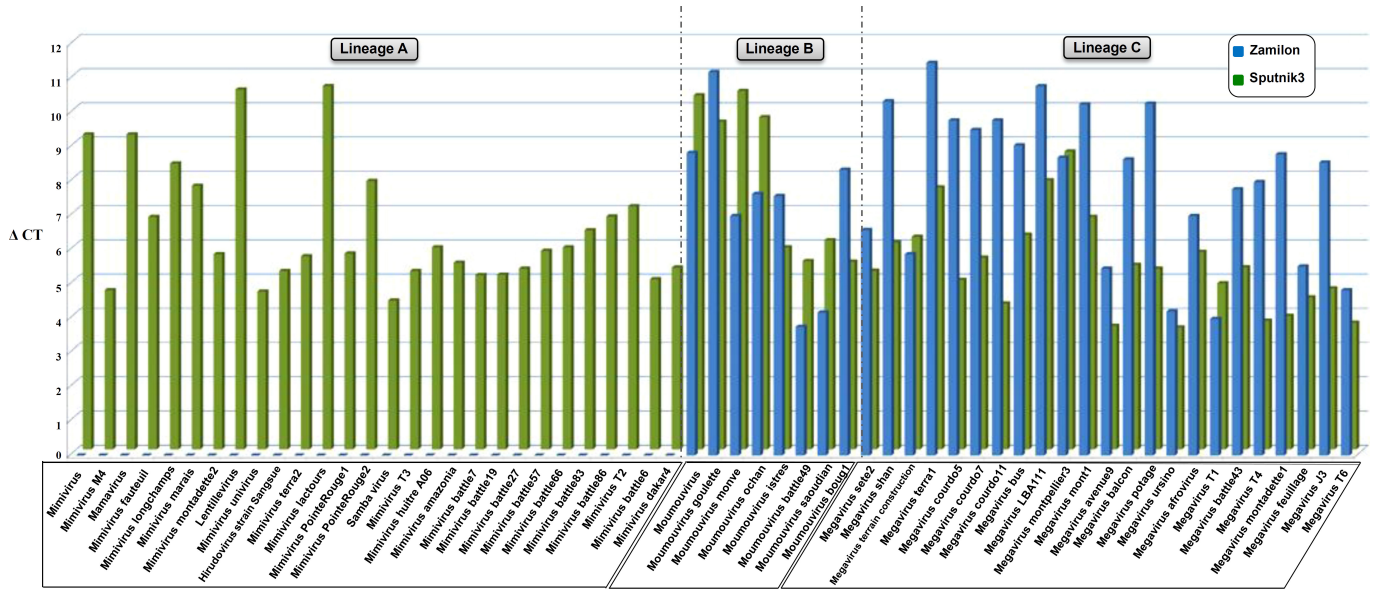
Virophage co-culture with *Mimiviridae*. *A. polyphaga* were suspended three times in PAS. One million APMV virions were inoculated individually into 10 ml of 5×10^5 cells per millilitre of rinsed *A. polyphaga* that contained 100 μl of either Sputnik 3 or Zamilon suspension. The co-culture was incubated for 1 h at 32 °C, and the supernatant was delicately removed to purge the virophage and APMV particles that did not enter the amoebas. Following this procedure, 10 ml of fresh PAS was added. This time point was defined as H0. Each virus was separately incubated without virophage to serve as a negative control. Mamavirus and Mont1 virus that were naturally infected with Sputnik 3 and Zamilon, respectively, were used as positive controls. At 0 and 24 h after infection, a 200 μl aliquot of co-culture was removed for DNA extraction and qPCR to enable the evaluation of virophage multiplication. The DNA extraction was performed using an EZ1 DNA Tissue Kit (Qiagen) according to the manufacturer's instructions. The qPCRs were performed in a CFX96 thermal cycler (Bio-Rad Laboratories) using a SYBR Green PCR Master Mix (Qiagen). Virophages were detected and quantified using primers targeting ORF20 for Sputnik 3 (forward primer 5′-GAGATGCTGATGGAGCCAAT-3′, reverse primer 5′-CATCCCACAAGAAAGGAGGA-3′) and ORF06 for Zamilon (forward primer 5′-GGGATGAACATCAAGCTGGT-3′, reverse primer 5′-GGGTTGTTGGAAGCTGACAT-3′).

Co-culture and mimivirus silencing. We targeted the mimivirus operon genes using siRNA, an oligonucleotide primer system, which was purchased from Invitrogen (http://rnaidesigner.invitrogen.com/rnaiexpress/) (Supplementary Table 3). We diluted each 20 μM solution of duplex siRNA and 50 μl of Lipofectamine RNAiMAX (Invitrogen) in 200 μl PAS according to the manufacturer's instructions and recommendations. To improve siRNA specificity, we used duplex siRNA and checked for specific and non-specific pairing. One hour before transfection, 1.5×10^6 *A. polyphaga* were put onto a plate with 5 ml of PAS to allow them time to adhere. After this, the siRNA-Lipofectamin suspension, 10^6 mimivirus particles and 10^{10} Zamilon virophage particles were all added to the plate containing the amoeba. The co-culture was incubated for 1 h at 32 °C, then the supernatant was delicately removed after centrifugation (1,700 *g* per minute), to eliminate the mimivirus and Zamilon particles that did not enter into the amoebas. The supernatant was replaced by 5 ml of fresh PAS containing the original concentration of siRNA-Lipofectamine, and the culture was submitted for a second incubation for 24–48 h at 32 °C. This time point was defined as H0. The same procedure was used with the omission of Zamilon and/or of mimivirus to serve as negative controls. To control siRNA transfection inside amoeba, a DMI6000 (Leica DMI 6000B) fluorescence microscope was used to visualize the green fluorescence of the oligonucleotides that were transfected into the amoeba. At 0, 24 and 48 h after infection, a 200 μl aliquot of co-culture was removed for DNA extraction and for real-time qPCR to evaluate Zamilon virophage multiplication. Twofold serial dilutions of Zamilon DNA from virophages that were cultivated either with mimivirus (wild type or silenced strains) or with *Mimiviridae* lineage B (Moumouvirus) and lineage C (Courdo7) strains were subjected to qPCR. The Zamilon DNA concentration was subsequently estimated at 0, 24 and 48 h for each condition. For the co-silencing of several genes of mimivirus, we used the same procedure previously mentioned according to the manufacturer's instructions and recommendations.

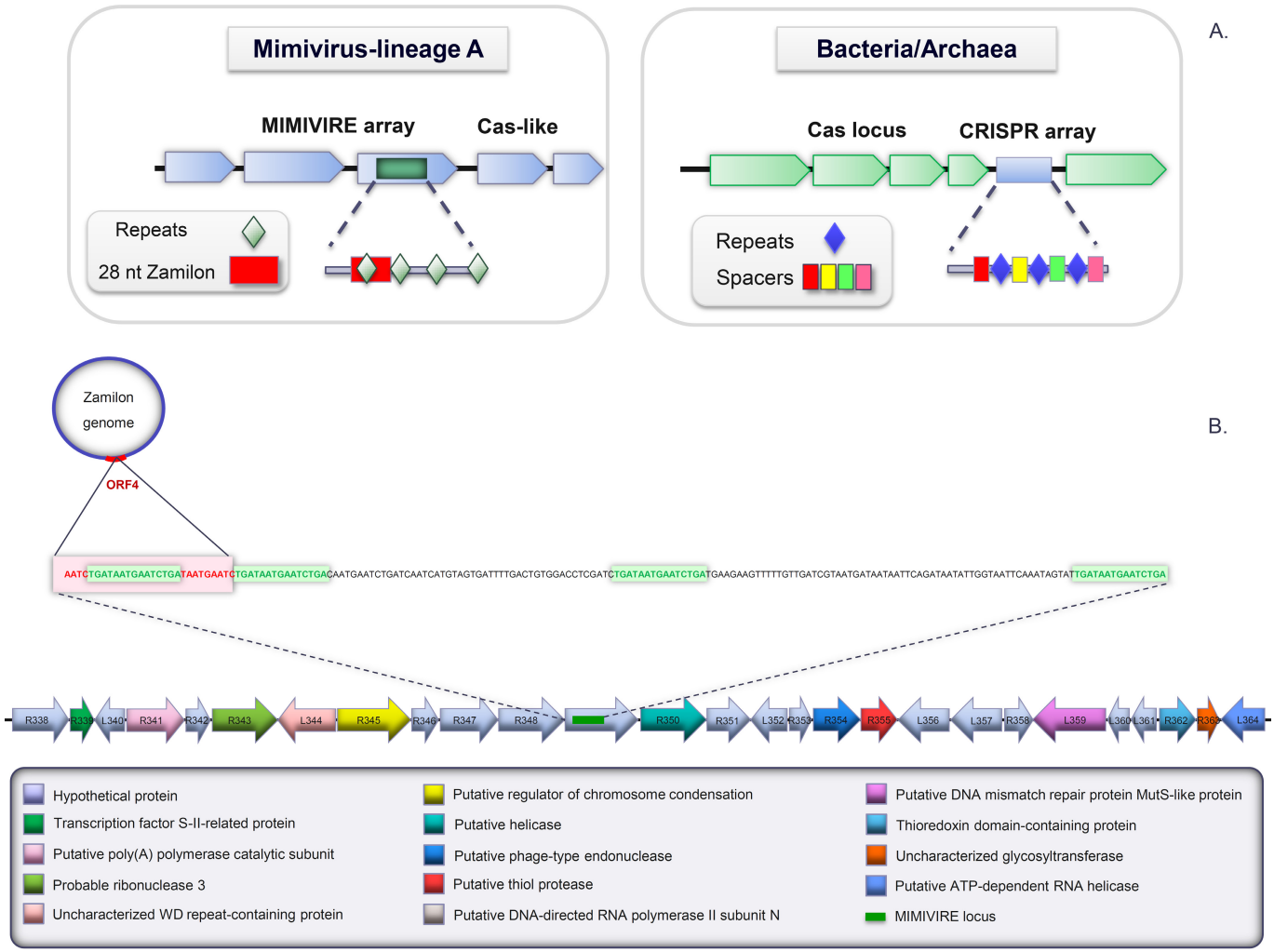
Cloning, expression and purification. Genes encoding the proteins R350 and R354 from APMV were codon-optimized for *E. coli* expression and synthesized by GenScript. Those optimized genes were designed to include a polyhistidine tag at the amino (N) terminus of each protein. Each gene was inserted between the NdeI and NotI cutting sites of a pET22b(+) plasmid. Recombinant proteins were expressed in *E. coli* BL21(DE3)-pGro7/GroEL (TaKaRa) using ZYP-5052 media. Each culture was grown at 37 °C until reaching an absorbance at 600 nm of 0.8 followed by addition of L-arabinose (0.2% m/v) and induction with a temperature transition to 18 °C over 20 h. Cells were harvested by centrifugation (4,250*g*, 30 min, 4 °C) and the resulting pellets were resuspended in wash buffer (50 mM Tris pH 8, 300 mM NaCl, 10 mM imidazole) and stored at −80 °C overnight. Frozen cells were thawed and incubated on ice for 1 h after adding lysozyme, DNase I and phenylmethylsulfonyl fluoride (PMSF) to final concentrations of, respectively, 0.25 mg ml^{-1}, 10 μg ml^{-1} and 0.1 mM. Partly lysed cells were then disrupted by three consecutive cycles of sonication (30 s, amplitude 45) performed on a Q700 sonicator system (QSonica). Cellular debris was discarded after a centrifugation step (21,640*g*, 20 min, 4 °C). The recombinant proteins were purified using immobilized metal affinity chromatography (wash buffer: 50 mM Tris pH 8, 300 mM NaCl, 10 mM imidazole; elution buffer: 50 mM Tris pH 8, 300 mM NaCl, 500 mM imidazole) on a 5 ml HisTrap FF crude column (GE Healthcare). Fractions containing each protein of interest were pooled and further purified using size-exclusion chromatography (buffer: 50 mM Tris pH 8, 300 mM NaCl) on a Superdex 75 16/60 column (GE Healthcare). Protein purity was assessed using 10% SDS–PAGE analysis (Coomassie stain). Bands matching the masses of the two proteins of interest were submitted to mass spectrometry analysis, which confirmed the expression of both desired proteins. Protein concentrations were measured using a Nanodrop 2000c spectrophotometer (Thermo Scientific).

Enzymatic treatments. Enzymatic reactions were performed by incubating each PCR product in combination with one of the enzymes (nuclease R354 or helicase–nuclease R350) or both enzymes together. The enzymatic reactions were conducted in PAS buffer solution at 32 °C for 2 h, using a protein concentration of 0.5 mg ml^{-1} for each enzyme. After incubation, the material was loaded onto agarose gel electrophoresis (1.5%). The DNA products used are listed in Extended Data Fig. 3. Controls were performed with different treatment parameters, such as the denaturation of the enzyme by heating at 94 °C for 10 min, denaturation of the enzyme coupling with DNA by heating at 94 °C for 10 min, 2 h after incubation, denaturation of DNA product by heating at 94 °C for 10 min before coupling with an enzyme and treatment time.

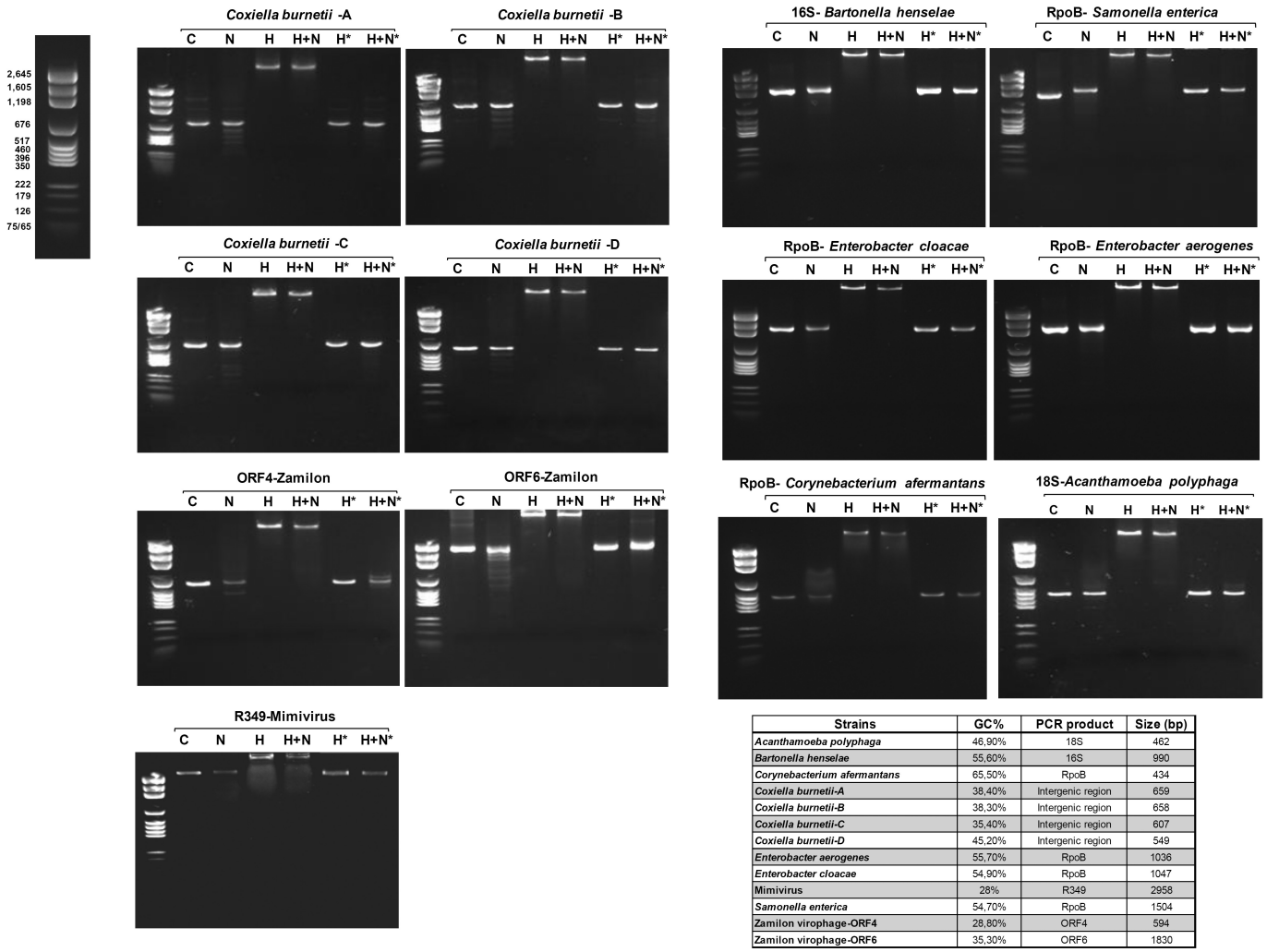
23. Besemer, J. Lomsadz, A. & Borodovsky, M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29,** 2607–2618 (2001).
24. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
25. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
26. La Scola, B., Mezi, L., Weiller, P. J. & Raoult, D. Isolation of *Legionella anisa* using an amoebic coculture procedure. *J. Clin. Microbiol.* **39,** 365–366 (2001).
27. Mba Medie, F., Ben Salah, I., Henrissat, B., Raoult, R. & Drancourt, M. *Mycobacterium tuberculosis* complex mycobacteria as amoeba-resistant organisms. *PLoS ONE* **6,** e20499 (2011).



Extended Data Figure 1 | Histogram depicting the replication of Zamilon and Sputnik 3 DNA in *Mimiviridae* after its phylogenetic classification into lineages A, B and C. The replication of each virophage was measured after 24 h using qPCR. The term Δ CT corresponds to the difference between the CT value specific to virophage at H0 and H24.

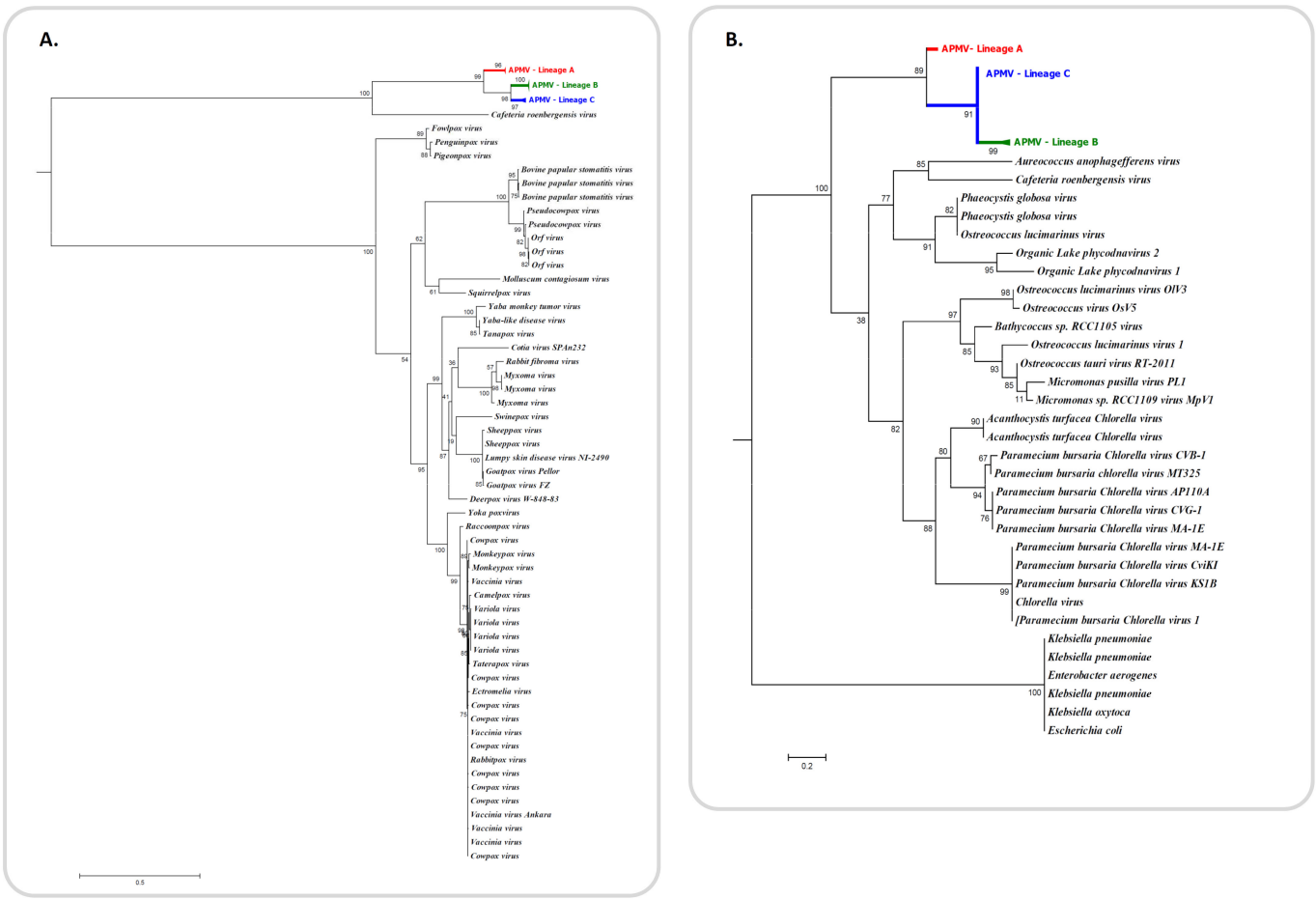


Extended Data Figure 2 | The MIMIVIRE defence system. **a**, A comparative model between prokaryotic CRISPR–Cas system and the viral MIMIVIRE system in APMV-A. **b**, The chromosomal environment of *Mimiviridae* lineage A is illustrated using mimivirus as an example. This organization is conserved across all APMV-A genomes. The 28-nucleotide-long Zamilon insert sequence is AATCTGATAATGAATCTGATAATGAATC, and the derived 15-nucleotide repeated unit is TGATAATGAATCTGA. The four repeats units are separated by 9, 48 and 63 nucleotides, respectively.



Strains	GC%	PCR product	Size (bp)
Acanthamoeba polyphaga	46,90%	18S	462
Bartonella henselae	55,60%	16S	990
Corynebacterium afermantans	65,50%	RpoB	434
Coxiella burnetii-A	38,40%	Intergenic region	659
Coxiella burnetii-B	38,30%	Intergenic region	658
Coxiella burnetii-C	35,40%	Intergenic region	607
Coxiella burnetii-D	45,20%	Intergenic region	549
Enterobacter aerogenes	55,70%	RpoB	1036
Enterobacter cloacae	54,90%	RpoB	1047
Mimivirus	28%	R349	2958
Samonella enterica	54,70%	RpoB	1504
Zamilon virophage-ORF4	28,80%	ORF4	594
Zamilon virophage-ORF6	35,30%	ORF6	1830

Extended Data Figure 3 | Agarose gel electrophoresis of different DNA products treated with and without nuclease and/or helicase enzymes. C, control; N, nuclease treatment for 2 h; H, helicase treatment for 2 h; H+N, helicase and nuclease treatment for 2 h; H*, helicase treatment for 2 h followed by heating at 94 °C for 10 min; H+N*, helicase and nuclease treatment for 2 h followed by heating at 94 °C for 10 min.



Extended Data Figure 4 | Phylogenetic trees based on the sequences of the two Cas proteins. a, R350. **b**, R354.

Extended Data Table 1 | Identification of the Zamilon sequences that were found inserted into the 60 genomes of *Mimiviridae*

Mimiviridae	Lineage	ZAMILON inserted sequence (28 bp length)	ZAMILON repeat CRISPR-like hits (15 bp length)
Mimivirus	A	1	4
Mimivirus M4	A	1	4
Mamavirus	A	1	4
Mimivirus fauteuil	A	1	4
Mimivirus longchamps	A	1	4
Mimivirus marais	A	1	4
Mimivirus montadette2	A	1	4
Lentillevirus	A	1	4
Mimivirus univirus	A	1	4
Hirudovirus strain Sangsue	A	1	4
Mimivirus terra2	A	1	4
Mimivirus lactours	A	1	4
Mimivirus PointeRouge1	A	1	4
Mimivirus PointeRouge2	A	1	4
Samba virus	A	1	4
Mimivirus T3	A	1	4
Mimivirus huitre A06	A	1	4
Mimivirus amazonia	A	1	4
Mimivirus battle7	A	1	4
Mimivirus battle19	A	1	4
Mimivirus battle27	A	1	4
Mimivirus battle57	A	1	4
Mimivirus battle66	A	1	4
Mimivirus battle83	A	1	4
Mimivirus battle86	A	1	4
Mimivirus T2	A	1	4
Mimivirus battle6	A	1	4
Mimivirus dakar4	A	1	4
Moumouvirus	B	0	1
Moumouvirus goulette	B	0	0
Moumouvirus monve	B	0	1
Moumouvirus ochan	B	0	0
Moumouvirus istres	B	0	0
Moumouvirus battle49	B	0	0
Moumouvirus saoudian	B	0	1
Moumouvirus boug1	B	0	0
Megavirus sete2	C	0	0
Megavirus shan	C	0	0
Megavirus terrain en constructi	C	0	0
Megavirus terra1	C	0	0
Megavirus courdo5	C	0	1
Megavirus courdo7	C	0	0
Megavirus courdo11	C	0	0
Megavirus bus	C	0	1
Megavirus LBA111	C	0	0
Megavirus montpellier3	C	0	0
Megavirus mont1	C	0	0
Megavirus chilensis	C	1	1
Megavirus avenue9	C	0	1
Megavirus balcon	C	0	0
Megavirus potage	C	0	0
Megavirus ursino	C	0	1
Megavirus afrovirus	C	0	0
Megavirus T1	C	0	0
Megavirus battle43	C	0	0
Megavirus T4	C	0	0
Megavirus montadette1	C	0	0
Megavirus feuillage	C	0	0
Megavirus J3	C	0	0
Megavirus T6	C	0	0

Lineage	Number of Mimiviridae genomes	Zamilon inserted sequence (28 bp length)	Zamilon repeat CRISPR-like hits (4 x 15 bp length)
A	28	28/28	28/28
B	8	0/8	0/8
C	24	1/24	0/24

All genomes were screened for the presence of the 28-nucleotide-long stretch (AATCTGATAATGAATCTGATAATGAATC) and the repeated 15-nucleotide sequence (TGATAATGAATCTGA).

Extended Data Table 2 | Functional inferences of the open reading frames in the vicinity of the MIMIVIRE locus in mimivirus

ORFs	Proposed gene-loke name	Accession	E-value	Function of ORFs
R343	RNAIII-Like	cd00593	6.85e-31	Probable ribonuclease 3
R350	Helicase domain / SNF2 domain	smart00490 pfam00176	1.66e-04 4.45e-14	Helicase Transcription regulation DNA repair
R354	Cas4-Like	pfam09588	4.69e-34	Putative phage-type endonuclease
L364	Helicase C DEXDc Cas3_I	cd00079 cd00046 cd09639	6.34e-13 2.45e-06 3.97e-03	Putative ATP-dependent RNA helicase

Sequence and structural similarity searches were performed by using BLAST and PHYRE2.