LETTER

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

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 delimited 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
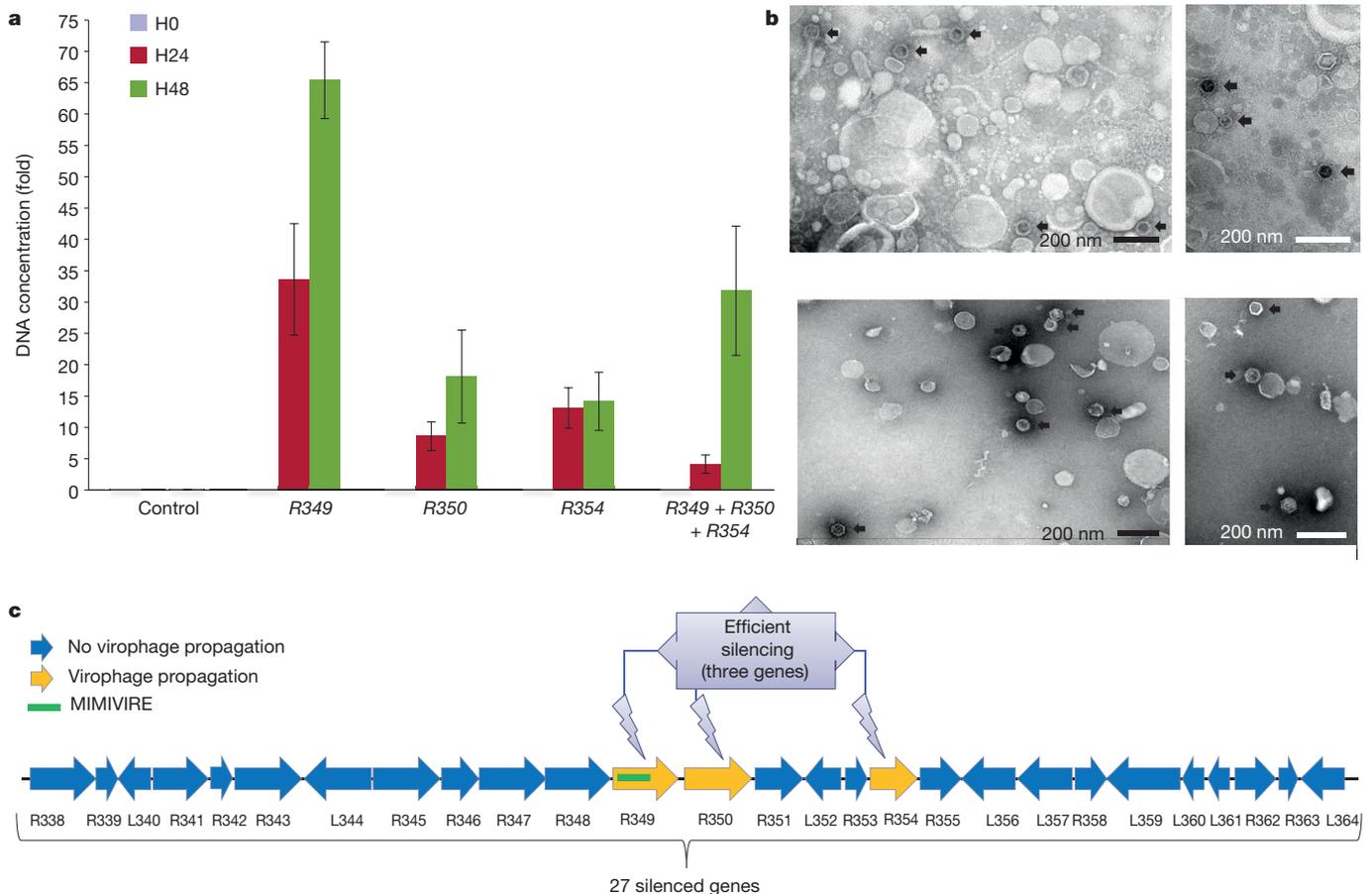
Figure 1 | Chromosomal environment of the MIMIVIRE locus of *Mimiviridae* and virophage infection. a, Quantification of Zamilon propagation after 0 h, 24 h and 48 h in the wild-type mimivirus (control) and in the three silenced mimivirus strains (genes *R349*, *R350* and *R354*). The *y* axis represents 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.

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).