Environmental Microbiology
September 2015, Volume 17, Issue 9, Pages 3278-3288
http://dx.doi.org/10.1111/1462-2920.12783
http://archimer.ifremer.fr/doc/00254/36483/

"Ménage à trois": a selfish genetic element uses a virus to propagate within *Thermotogales*

Lossouarn Julien [1, 2, 3], Nesbo Camilla L. [4, 5], Mercier Coraline [1, 2, 3], Zhaxybayeva Olga [6], Johnson Milo S. [6], Charchuck Rhianna [4], Farasin Julien [1, 2, 3], Bienvenu Nadège [1, 2, 3], Baudoux Anne-Claire [7, 8, 9], Michoud Grégoire [1, 2, 3], Jebbar Mohamed [1, 2, 3], Geslin Claire [1, 2, 3, *]

[1] Université de Bretagne Occidentale (UBO, UEB), Institut Universitaire Européen de la Mer (IUEM) – UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes (LMEE), rue Dumont d'Urville, Plouzané, France
[2] CNRS, IUEM – UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes (LMEE), rue Dumont d'Urville, Plouzané, France
[3] Ifremer, UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes (LMEE), Technopôle Pointe du diablea, Plouzané, France
[4] CEES, Department of Biology, University of Oslo, Oslo, Norway
[5] Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada
[6] Department of Biological Sciences, Dartmouth College, Hanover, NH, USA
[7] Sorbonne Universités, UPMC Univ Paris 06, Paris, France
[8] UMR 7144, Equipe DIPO, Station Biologique de Roscoff, Roscoff, France
[9] CNRS, UMR 7144, Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff, Roscoff, France

* Corresponding author : Claire Geslin, Tel. +33 (0)2 98 49 88 58; Fax +33 (0)2 98 49 87 05 ; email address : claire.geslin@univ-brest.fr

Abstract :

Prokaryotic viruses play a major role in the microbial ecology and evolution. However, the virosphere associated with deep-sea hydrothermal ecosystems remains largely unexplored. Numerous instances of lateral gene transfer have contributed to the complex and incongruent evolutionary history of Thermotogales, an order well represented in deep-sea hydrothermal vents. The presence of clustered regularly interspaced short palindromic repeats (CRISPR) loci has been reported in all Thermotogales genomes, suggesting that these bacteria have been exposed to viral infections that could have mediated gene exchange. In this study, we isolated and characterized the first virus infecting bacteria from the order Thermotogales, Marinitoga piezophila virus 1 (MPV1). The host, Marinitoga piezophila is a thermophilic, anaerobic and piezophilic bacterium isolated from a deep-sea hydrothermal chimney. MPV1 is a temperate Siphoviridae-like virus with a 43.7 kb genome. Surprisingly, we found that MPV1 virions carry not only the viral DNA but preferentially package a plasmid of 13.3 kb (pMP1) also carried by M. piezophila. This 'ménage à trois' highlights potential relevance of selfish genetic elements in facilitating lateral gene transfer in the deep-sea biosphere.

1. Introduction

Deep-sea hydrothermal vents are home to many thermophilic and hyperthermophilic *Bacteria* and *Archaea,* whose genetic diversity, metabolic capacity and physiological adaptations are being actively explored (reviewed in Flores and Reysenbach, 2011). However, the virosphere associated with these deep-sea hydrothermal microorganisms remains largely unstudied. The observation of abundant populations of lysogens and temperate viruses indicates that virus-mediated gene exchange and recombination may be critical to the survival and stability of prokaryotes within extreme environments (Williamson *et al*., 2008). Additionally, metagenomic studies suggest that a wide diversity of deep-sea hydrothermal microorganisms is exposed to viral infections (Williamson *et al*., 2008; Anderson *et al*., 2011a). Yet, published viromes mostly contain reads with no homologs in public databases and do not provide any information about the infected hosts (Anderson *et al*., 2011b). Therefore, to understand the impact of viruses on the microbial diversity in the deep-sea ecosystems, it remains crucial to isolate new viruses and to characterize host-virus model systems.

To date, only seven viruses, isolated from deep sea vents, have been described. Two of them are lemon-shaped archaeoviruses, PAV1 and TPV1, which infect *Thermococcales*, obligate anaerobic sulfur-metabolizing hyperthermophiles (Geslin *et al.,* 2003a; Geslin *et al*., 2007; Gorlas *et al*., 2012). The other five isolates are tailed bacterioviruses. Among these, four lytic bacterioviruses infect aerobic, thermophilic, and heterotrophic strains belonging to *Bacillus* and *Geobacillus* genera with optimal growth temperatures around 65°C. The bacteriovirus BVW1 is unclassified whereas GVE1 and GVE2 are affiliated to the *Siphoviridae* family (Liu *et al*., 2006; Liu and Zhang, 2008), and D6E belongs to the *Myoviridae* family (Wang and Zhang, 2010). The remaining bacteriovirus, NrS-1, is a temperate siphovirus which infects the chemolithoautotrophic, anaerobic and microaerobic, moderately thermophilic *Epsilonproteobacterium Nitratiruptor* sp. cultivated at 55°C (Yoshida-Takashima *et al*., 2013).

Here we report and characterize the first bacteriovirus that infects cells from the bacterial order *Thermotogales*, a lineage broadly represented in deep biosphere ecosystems and deep-sea hydrothermal vents in particular. This order consists of mesophilic (Nesbø *et al*., 2012; Hania *et al*., 2013) and hyper/thermophilic, anaerobic, chemo-organotrophic microorganisms (Bonch-Osmolovskaya, 2008). Phylogenetic analyses of 16S rRNA gene and 29 concatenated ribosomal proteins place *Thermotogales* as a sister group to *Aquificales*, and consequently as the second deepest branching order within domain *Bacteria* (Zhaxybayeva *et al*., 2009). However, numerous instances of lateral gene transfer have contributed to the evolutionary history of these bacteria: many genes were exchanged with *Firmicutes* and, to a lesser extent, with *Archaea* and especially *Thermococcales* (Nelson *et al*., 1999; Zhaxybayeva *et al*., 2009). Similarly to many thermophilic prokaryotes, clustered regularly interspaced short palindromic repeats (CRISPRs) are present in all sequenced *Thermotogales* genomes (Zhaxybayeva *et al*., 2009), suggesting that *Thermotogales* have been exposed to viral infections. Yet, prior to this study, no viruses infecting *Thermotogales* had been isolated. The only known extrachromosomal genetic elements in these lineages are two cryptic miniplasmids in the genus *Thermotoga* (Harriott *et al.*, 1994; Akimkina *et al*., 1999) and a 1724 bp miniplasmid in *Mesotoga prima* (Zhaxybayeva *et al*., 2012).

The newly discovered virus infects *Marinitoga piezophila* KA3, a member of the *Thermotogales* isolated from a deep-sea hydrothermal vent chimney located at a depth of 2,630 m on the East-Pacific Rise (Alain *et al*., 2002). *M. piezophila* virus 1 (MPV1) is a temperate *Siphoviridae*-like virus and is mitomycin C inducible. Its genome of 43.7 kb is identical to the provirus integrated in *M. piezophila* KA3 chromosome (Lucas *et al*., 2012). Additionally, we discovered that MPV1 virions mainly contain a plasmid of 13.3 kb (referred

as pMP1), which corresponds not to the viral genome but to a second mobile genetic element also carried by *M. piezophila*. We demonstrate that pMP1 uses the MPV1 viral capsid to propagate, highlighting a complex evolutionary relationship between a bacterial host, an extrachromosomal element and a virus.

2. Results

2.1. Virion morphology.

Particles of MPV1 have a head with a hexagonal outline of ~55 nm in diameter and a flexible non-contractile tail of ~200 nm in length and ~10 nm in width (Fig. 1). Based on these morphological characteristics, we classify MPV1 as a siphovirus from the order *Caudovirales*. Despite the piezophilic character of *M. piezophila*, which grows optimally at high hydrostatic pressure (40 MPa), this isolate can adapt to the growth at atmospheric pressure (0.1 MPa) after several subcultures. The virion morphology was identical under these two growth conditions. Since assays indicated that the viral production is functional and similar at both atmospheric and high hydrostatic pressures of 40 MPa (Table S1), all subsequent experiments were carried out at atmospheric pressure.

2.2. Virus-host relationship.

MPV1 is a mitomycin C inducible temperate bacteriovirus causing host cells lysis upon induction (Fig. 2). A concentration of 5 µg/mL of mitomycin C was sufficient for optimal virus induction, as estimated by viral counts (Fig. S1). The viral production increased, reaching ~3 $x10^9$ viral particles mL^{-1}, 2 h after the induction by 5 µg/mL of mitomycin C (Fig. 2). A concomitant drastic drop to 2.19 x 10^6 cells mL^{-1} in host cell counts was observed, suggesting the lysis of the cells (Fig. 2). Regrowth of bacteria was observed 5h after induction, suggesting that some cells withstood the mitomycin C effects. Untreated cells produced 30-fold less viral particles than mitomycin treated culture, with a maximal virus abundance of 1.11 x 10^8 viral particles mL^{-1} after 9 h of incubation (Fig. 2). This basal viral production could be explained by a spontaneous induction of MPV1 within a fraction of the growing host cells. Similar spontaneous induction was observed during the growth of *Nitratiruptor* sp., the deep-sea hydrothermal isolate lysogenized by the mitomycin C inducible temperate siphovirus NrS-1 (Yoshida-Takashima *et al*., 2013).

2.3. DNA content of viral particles.

The genome sequence of *M. piezophila* KA3 contains a provirus sequence of ~40 kb integrated into the strain's chromosome and a cccDNA of 13 386 bp denoted as pMARPI01 (Lucas *et al*., 2012). pMARPI01 has an *in silico* RFLP pattern identical to that of pMP1, the plasmid that we isolated from *M. piezophila* KA3 by an alkaline lysis extraction (Fig. S2A). For brevity, we will refer to both pMARPI01 and the discovered plasmid as pMP1 throughout the rest of the manuscript. Analysis of DNA extracted from the purified MPV1 viral capsids confirmed that the provirus is the integrated form of the MPV1 genome. Surprisingly, RFLP analyses indicated that the viral DNA is much less frequently present in MPV1 virions (Fig. S2B). Instead, the viral capsids mainly contain pMP1 and hence preferentially package the plasmid. Sequencing of the DNA from the viral particles further confirmed this hypothesis, since not only both a 43 715 bp sequence corresponding to the provirus inserted in *M. piezophila* genome (Fig. 3) and a 13 386 bp plasmid DNA corresponding to pMP1 (Fig. 4) were fully assembled, but also a 20-fold higher coverage of the plasmid reads compared to the “true viral” reads was observed. We rule out possible extracapsid DNA contamination,

since the purified viral suspensions from induced cultures were treated with DNase before DNA extraction. Moreover, specific PCR amplifications targeting the provirus, the plasmid and the 16S rRNA gene in the DNA extracted from the viral capsids, resulted in no PCR product corresponding to the 16S rRNA gene (Fig. S3). The same results were obtained with DNA extracted from a purified viral suspension of an uninduced culture (not shown). We concluded that even without induction both viral and plasmid DNA were incorporated into the viral capsids.

2.4. Attempts to infect a putative sensitive host.

Despite several attempts, we never succeeded in curing *M. piezophila KA3* strain of MPV1 or pMP1, thus we could not test MPV1's ability to re-infect its host. A purified viral suspension was added to growing cultures of two other potential host cells (*Thermosipho sp.* strain AT1244-VC14 and *Marinitoga camini* strain DV1142). For both cultures, no lysis was observed under a phase-contrast microscope. *Marinitoga camini* strain DV1142 did not get infected by MPV1 (data not shown). After 3 subcultures of the *Thermosipho sp.* strain AT1244-VC14, no viral DNA was detected, but a 198 bp fragment from the pMP1 plasmid was amplified by PCR from the total DNA (Fig. 5). Moreover, the complete plasmid pMP1 was extracted by alkaline lysis from the *Thermosipho sp.* strain (Fig. S4). These findings highlight an inter-genus plasmid (pMP1) propagation by means of a viral capsid.

2.5. Effect of the mitomycin C induction on the evolution of the copy numbers of chromosomal, proviral and plasmid target genes by qPCR analyses.

The relative quantity of the chromosomal DNA in *M. piezophila* KA3 induced with mitomycin C was 2 times lower than in the culture not exposed to the antibiotic (Figs. 6 and S5). This might be a result of a chromosomal DNA damaging effect of mitomycin C, which induces the lytic phase of MPV1. On the other hand, the relative concentration of viral DNA was almost 3-fold higher in mitomycin C treated cells than in untreated ones (Figs. 6 and S5). This post-induction increase of the viral DNA replication was consistent with the lysogenic relationship between MPV1 and its host. Finally, the relative concentration of plasmid DNA was 10 fold higher in mitomycin C treated cells than in the uninduced cells (Figs. 6 and S5). The induction of the plasmid was unexpected and could be a response, as a survival strategy, to escape to the host lysis.

2.6. Sequence analysis of the bacterioviral DNA and the plasmid DNA.

The viral DNA sequences obtained from the purified MPV1 capsids was identical to the provirus sequence on *M. piezophila* KA3's chromosome (Fig. 3). *De novo* assembly of the viral reads revealed that while some of the packaged DNA molecules are circular, others are linearized, disrupting the Open Reading Frame (ORF) corresponding to Marpi_0300 in the provirus (marked by a star in Fig. 3). The viral genome has 54 predicted ORFs, which include all genes necessary for viral particle production (Fig. 3). The gene order in the viral genome is similar to that seen in *Enterococcus* viruses (Yasmin *et al*., 2010). Based on BLAST searches, 23 of the 54 predicted viral ORFs have no detectable homologs in GenBank (Table S2). Among the 31 ORFs with homologs in GenBank, 8 ORFs have homologs in proviruses identified by us in 2 *Thermosipho sp.* genomes (C. Nesbø, unpublished), and 3 ORFs have homologs in the genomes of other *Thermotogales* (Table S2). Therefore, it is likely that MPV1 and closely related viruses have exchanged genes with other *Thermotogales* in the past. However, in the available *Thermotogales* genomes,

CRISPR arrays are >85% identical to MPV1 DNA across the entire spacer in only one *Marinitoga* genome (*Marinitoga* sp. 1137, C. Nesbø, unpublished) (data not shown).

When BLAST "self matches" to the *M. piezophila* genome and matches to the above-mentioned *Thermosipho* proviruses were excluded, 22 out of the 31 ORFs had *Firmicutes* genes as their top-scoring BLAST hit. Moreover, among these 31 ORFs, 12 were similar to genes in other bacterioviruses. The majority of these bacterioviral homologs belong to siphoviruses, but some are from the *Myoviridae* family of bacterioviruses with contractile tails (Krupovic *et al*., 2011), exemplified by the thermophilic *Geobacillus bacteriovirus* GBSV1 (Liu *et al*., 2009) (Table S2). On phylogenetic trees of 11 out of 19 MPV1 ORFs with more than three GenBank homologs, the sister taxon of an MPV1 gene was either a *Firmicute* or a *Firmicute* bacteriovirus (Table S2). These analyses suggest a close evolutionary relationship of MPV1 to *Firmicutes* bacterioviruses.

Lysis in Gram-negative bacteria involves three types of proteins: (i) endolysins (to degrade murein); (ii) holins (to increase inner membrane permeability) and (iii) spannins (to disrupt the outer membrane) (Wang *et al*., 2000). Based on sequence similarity and composition, we have identified candidate lysis genes in the MPV1 genome. We assigned Marpi_340 a putative endolysin function, since similarity searches place the gene with phosphoesterases and metallophosphatase (MPP) superfamily. We designated Marpi_343 and Marpi_344 as holins: Marpi_343 has significant similarity to holin homologs in GenBank, while physicochemical features of Marpi_344 (102 amino acids in length, predicted transmembrane helix, dual start motif and a charged C-terminal sequence) are compatible with those of known holins (Young, 2013). Similarity searches placed Marpi_345 within the NlpC/P60 family (secreted cysteine proteinases that degrade cell walls), making this gene a candidate for encoding a protein that disrupts the outer membrane. Marpi_341 possibly encodes another membrane-degrading enzyme, since it exhibits a weak similarity to type I phosphodiesterase/nucleotide pyrophosphatase superfamily in *Heliobacterium modesticaldum* Ice1 (a *Firmicute*).

pMP1 plasmid contains 13 ORFs: 6 have Genbank homologs with a predicted function (Fig. 4 and Table S3), while remaining 7 could be assigned to the gene families in the Superfamily database (Gough *et al*., 2001). No known structural genes involved in virion formation were identified. Nine of the putative 13 proteins have a predicted pI above 9.0 (Table S3). Moreover, all 7 protein-coding genes with homologs in the Superfamily database were assigned to gene families involved in DNA interactions (i.e., transcription, regulation, and DNA metabolic processes, Table S3). Although pMP1 was not found integrated into the sequenced *M. piezophila* chromosome, one of its genes (Marpi_2122) encodes a site-specific recombinase XerD (an integrase family protein). Genes for DNA polymerases and primases are not commonly found on plasmids, however, Marpi_2127 encodes a protein with both a DnaB (helicase) and a DnaG (primase) domain. Recently, a DNA primase/polymerase protein (PolpTN2) encoded by the pTN2 plasmid from *Thermococcus nautilus* has been characterized (Gill *et al.*, 2014). To summarize, a large fraction of pMP1 encoded proteins might be involved in DNA binding and possibly replication and propagation of the plasmid.

Pairwise BLASTN comparison of the pMP1 and MPV1 genomes revealed a 17 bp region with 100% identity [E-value = 0.14; located in Marpi_300 (MPV1) and Marpi_2128 (pMP1)] and three short regions with > 80% identity (E-value < 0.05): a 28 bp region with 89% identity [located in Marpi_300 (MPV1) and Marpi_2127 (pMP1)], a 33 bp region with 85% identity [located in Marpi_315 (MPV1) and Marpi_2120 (pMP1)], and a 23 bp region with 91% identity [located in Marpi_306 (MPV1) and Marpi_2127 (pMP1)] (Figs 3 and 4, Tables S2 and S3). These matching sequence motifs could serve as recognition sequences during packing of pMP1 genome into the viral particles. Interestingly, one of the pMP1 ORFs (Marpi_2120) has significant similarity to the proviral gene (Marpi_0315; green in Fig. 3) located two ORFs upstream of a putative small terminase subunit, which is involved in packing of viral DNA

(Feiss and Rao, 2012). Both Marpi_2120 and Marpi_0315 have a DNA binding helix-turn-helix domain and show significant similarity to Sigma-70 family RNA polymerase genes. Taken together, these observations strongly suggest that both Marpi_0315 and Marpi_2120 could be involved in packing DNA into the viral particles.

3. Discussion

Deep-sea hydrothermal vents are harsh environments, characterized by steep physicochemical gradients, high hydrostatic pressures, high temperatures, lack of solar energy, and prevalence of microbial chemosynthesis. The dynamic, gradient-dominated nature of the vent environment provides niches for diverse communities of microorganisms and makes it a particularly attractive site for studies of viral diversity and evolution (Anderson *et al*., 2011a). In this ecosystem, *Thermotogales* share the ecological niche with diverse prokaryotes, such as frequently isolated bacteria from the *Firmicutes* phylum and archaea from the *Thermococcales* order (Prieur *et al.*, 2004). The bacterial order *Thermotogales* is known for its complex evolutionary history shaped by large amounts of gene transfer from the above prokaryotic lineages (Nelson *et al*., 1999; Zhaxybayeva *et al*., 2009). Viruses represent one possible vector for this gene exchange, but prior to this study no viruses infecting *Thermotogales* cells have been identified. In support of this hypothesis, the reported genome of the first *Thermotogales* bacteriovirus, MPV1, reveals a connection to genomes of *Firmicutes* and bacterioviruses known to infect them (Table S2). Since *Thermotogales* and *Firmicutes* appear to share large number of genes (Zhaxybayeva *et al*., 2009), our findings imply that some of the genes exchanges may have been mediated by viruses. Further studies of *Thermotogales* and *Firmicutes* viruses will be needed to explore this hypothesis.

MPV1 is a temperate virus and lysogeny is indeed a common lifestyle in extreme environments such as hydrothermal vents (Williamson *et al*., 2008, Anderson *et al*., 2011a). From the viral point of view, lysogeny could be favored due to reduced viability of viral particles outside of the cell. For the cell, on another hand, carrying a virus could increase its fitness by enhancing the survivability in fluctuating environmental conditions via introduced novel genes and/or mutations (Anderson *et al*., 2011a). Further analyses will be necessary to better understand the MPV1 viral cycle. In particular, the mechanisms of MPV1 adsorption and extrusion through the “toga”, an unusual cell envelope forming a loose sheet envelope around cells, remain unknown. The toga is mainly composed of two proteins, the anchor protein OmpA1 and the porin OmpB (Petrus *et al*., 2012). Porins are known to serve as viral receptors: In *E. coli*, for example, T4 viral particles use the OmpC porin as a receptor (Hashemolhosseini *et al*., 1994; Marti *et al*., 2013). Therefore, we hypothesize that OmpB may serve as MPV1 receptors in *Thermotogales*.

Peculiarly, MPV1 viral capsids mainly package a plasmid of 13.3 kb (pMP1), facilitating the inter-genus plasmid propagation. Thus, this system is the first virus-mediated plasmid exchange reported in deep-sea hydrothermal vents. Although the interaction between MPV1 and pMP1 is not fully understood, the ratio of packed plasmid/viral DNA is high, which suggests that pMP1/ MPV1 represent a new example of molecular piracy. This term has been coined to describe a biological phenomenon in which one replicon (the pirate) uses the structural proteins encoded by another replicon (the viral helper) to pack its own genome and thus allow the pirate’s propagation (Christie and Dokland, 2012). The type of relationship between a virus and another mobile genetic element within a same host, the so-called "ménage à trois", has been described in organisms from all three domains of life. A distinct plasmid (a transpoviron) discovered in giant eukaryoviruses infecting amoebae depends on giant viruses for its replication and propagation (Desnues *et al*., 2012). Analogies can be made between the eukaryotic transpovirons and virus-associated plasmids described in archaea and bacteria. For example, non-conjugative plasmids pSSVi and pSSVx exploit

fuselloviruses in order to propagate within strains of the hyperthermophilic archaeon *Sulfolobus* (Arnold *et al*., 1999; Wang *et al*., 2007). Similar strategies were described for the well-studied *E. coli* P4/P2 system in which the plasmid P4 uses the capsids of the bacteriovirus helper P2 to propagate in *E. coli* cells (Six, 1975; Lindqvist *et al*., 1993; Briani *et al*., 2001) and the *S. aureus* SaPIs pathogenicity islands/bacterioviruses helper systems (Ram *et al*., 2012). In the two bacterial systems described above, the pirate element is mobilized and packed into viral-like transducing particles assembled from proteins supplied by a helper caudovirus. These particular relationships involve a complex web of interactions between the helper and the pirate that occur at several levels, from transcriptional control to macromolecular assembly (Christie and Dokland, 2012).

The MPV1/pMP1 system shares many features with the *E. coli* and *S. aureus* pirates. Both pMP1 and P4 are plasmids of similar sizes, and are able to persist in their respective host cells in a self-replicating multicopy cccDNA state. Similar to P4 and SapIs, pMP1 harbors an integrase gene, suggesting that it might recombine within host chromosome. Also, like the SaPIs pirates, pMP1 is induced by mitomycin C along with the provirus, and responds by replicating and escaping from the bacterial hosts by using helper viral capsids. The induction of the plasmid was unexpected and, to our knowledge, has not yet been reported in prokaryotes.

It has been demonstrated that pirates have evolved different strategies of interference with their helpers to favor the packing of their own pirate genomes into the helper capsids (Christie and Dokland, 2012). Specific DNA packing strategy of the pMP1 element remains to be discovered. For example, we have not observed small MPV1 capsids that are likely to contain a single copy of the plasmid pMP1. We hypothesize that multiple copies of pMP1 (x3) are packaged in MPV1 capsids. Indeed, the discontinuous headful packaging of multiple plasmid DNAs in viral capsid has been demonstrated (Bravo and Alonso, 1990; Coren *et al*., 1995; Leffers and Rao, 1996). It has been shown that the frequency of these plasmid-containing particles is enhanced when there is DNA similarity between a plasmid and a virus (Bravo and Alonso, 1990). We detected three regions of significant sequence similarity between pMP1 and MPV1 genomes. Among these, the strongest candidate for a common packing signal is a 28 bp region in Marpi_0300, which is disrupted in the linearized version of the MPV1 genome. The pMP1 protein encoded by Marpi_2120 might be involved in the binding to the putative packing signal.

Would there be any benefit for the host from this "ménage à trois"? Carrying a provirus often confers resistance against related bacterioviruses (Canchaya *et al*., 2003), while bearing the pirate that hijacks the viral capsid could increase the survival of the bacterial host population, due to lesser exposure to the bacterioviral DNA (Christie and Dokland, 2012). Resistance of the helper bacterioviruses against the pirate elements, which alter the propagation capacity of the bacteriovirus for their own survival, is expected to evolve but have not yet been observed. Thus, the MPV1/pMP1 system likely represents an ongoing evolutionary “arms race”.

Mobile genetic elements (MGEs), such as viruses, plasmids, membrane vesicles, gene transfer agents (GTAs), transposons and transpovirons (all collectively referred as a mobilome) interact with cellular organisms from all three domains of life, including those thriving in most extreme environments. The MPV1/pMP1 system demonstrates the existence of an intricate network of interactions in the deep-sea vent mobilome, and further studies are likely to unravel unknown mechanisms for the host-MGEs interactions. The MGEs have a potential to be powerful drivers of cellular host adaptations to the extreme marine environments, and comparative studies of these elements from hyper/thermophilic *Archaea* and *Bacteria* will help us understand the dynamics of the microbial communities in the deep biosphere.

4. Experimental procedures

4.1. Screening of Thermotogales strains for mobile genetic elements.

A total of 60 strains of *Thermotogales* were retrieved from the UBO culture collection "UBOCC" (www.univ-brest.fr/UBOCC) and from our laboratory culture collections. As described in *Supporting information*, these strains were isolated from different deep-sea vents in the Lau Basin, the East Pacific Rise and the Middle Atlantic Ridge and were screened for mobile genetic elements as previously described (Geslin *et al*, 2003a,b; Gorlas *et al*, 2012). One of them, the reference strain *M. piezophila* KA3 (Alain *et al*., 2002), was discovered to be the host of the virus MPV1 and the plasmid pMP1.

4.2. Culture conditions.

Thermotogales strains were cultured in a modified Ravot medium as previously described (Ravot *et al*., 1995; Geslin *et al*., 2003a,b) with minor modifications as described in *Supporting information*. Most experiments were performed with the *M. piezophila KA3* cultured at 65°C and 0.1 MPa. For the high hydrostatic pressure growth experiments, *M. piezophila KA3* cells were grown at 65°C and 40 MPa pressure as described in *Supporting information.*

4.3. Viral induction assays by addition of mitomycin C.

Attempts were made to increase the viral production by induction using mitomycin C (Yoshida-Takashima *et al*., 2013). Several final concentrations of mitomycin C (0.1 µg/mL, 1 µg/mL, 5 µg/mL and 10 µg/mL) were added to cultures of *M. piezophila* at mid-log growth phase. After 3 hours of incubation with mitomycin C, aliquots were collected in duplicate to determine viral production by epifluorescence microscopy. A 'mitomycin C-free' culture was used as a control. Samples were fixed with 25% glutaraldehyde [0.5% final concentration, EM grade] for 15 min at 4°C, flash frozen in liquid nitrogen, and stored at -80°C until analysis. The viral counts were performed according to Noble and Fuhrman (Noble and Fuhrman, 1998). Viral particles collected on anodisc 0.02 µm filters were counted in each of 20 random fields using an Olympus microscope BX60 (magnification x 1000, WIB filter).

4.4. Flow cytometry counts of bacteria and viruses.

Bacteria and virus abundances were monitored by flow cytometry using a FACSCanto II flow cytometer (BD Bioscience) equipped with a laser with an excitation wavelength of 488 nm (15 mW). Samples were fixed as previously described. The thawed samples were diluted 100 to 5,000-fold in autoclaved 0.2 µm filtered TE buffer (10:1 Tris–EDTA, pH 8.0) and stained with the nucleic acid-specific dye SYBR Green I (Invitrogen-Molecular Probes) for 15 min at room temperature and 10 min at 80°C for bacteria and viruses, respectively. The trigger was set on green fluorescence. The sample was delivered at a rate of 50 µl.min^{-1} and analyzed during at least 1 min. TE-buffer with autoclaved 0.2 µm filtered seawater was used as a control.

4.5. Concentration and purification of MPV1.

300 mL of *M. piezophila* culture were induced by addition of mitomycin C at early log-phase (5 µg/mL). Two hours later, debris and cells were pelleted by centrifugation at 7,500 g, 4°C for 15 min and discarded. The supernatant was ultracentrifuged at 80,000 g, 10°C for 2 h (Beckman Optima LE-80 K 70.1 Ti rotor). The viral pellet was resuspended in 1 mL of buffer (10 mM Tris-HCL, 100 mM NaCl, 5 mM $CaCl_2$, 20 mM $MgCl_2$). The viral suspension was purified by ultracentrifugation in a linear Iodixanol gradient [OptiPrep, 30–45% diluted in a buffer (see above)] at 40, 000 *g* and 10°C for 5 h (Beckman Optima LE-80 K SW 55 rotor). Following ultracentrifugation, the opaque virus band was recovered and stored at 4°C until use. To exclude the possibility of cellular nucleic acids contamination, the concentrated viral particles suspensions were treated with DNase and RNase (50 µg/ml and 100 µg/ml respectively). The viral particles were examined using a JEOL JEM 100 CX II transmission electron microscope as previously described (Geslin *et al*, 2003a, b; Gorlas *et al*, 2012).

4.6. Extraction of cellular and viral DNA.

Total DNA from *M. piezophila* cells and viral DNA from purified virions were prepared as previously described (Geslin *et al*, 2003a, b; Gorlas *et al*, 2012). Plasmid DNA was extracted from cells in log growth phase by alkaline lysis method (Bimboim and Doly, 1979). To confirm the nature of the DNA packaged in the viral capsids and to check potential chromosomal contamination, PCR amplifications were performed using primer sets for 16S rRNA gene (*M. piezophila* chromosomal DNA marker), major capsid protein encoding gene (Marpi_326, provirus and virus marker), and hypothetical protein encoding gene (Marpi_2124, pMP1 plasmid marker). The primers and PCR protocol are described in *Supporting information*.

4.7. RFLP analyses.

Marinitoga piezophila KA3 total DNA and plasmid DNA were restricted by the enzymes HhaI and HindIII (Promega ®); packaged viral DNA was restricted by HhaI, in accordance with the manufacturer's instructions. Restriction fragments obtained were separated on a 0.8 % agarose gel electrophoresis.

4.8. Attempts to infect a putative sensitive host.

Strains belonging to *Marinitoga* and *Thermosipho* genera, obtained from our laboratory cultures collections, were screened for the presence of genetic elements and strains scored negative were selected as potential hosts. As a result, *Thermosipho sp.* AT1244-VC14 and *Marinitoga camini* DV1142 strains were selected. A purified MPV1 suspension was added in exponentially growing host (virus/host=30) cultured in 5 ml modified Ravot medium at 65°C. Non infected cultures were used as negative control. After 3 subcultures, total and plasmid DNAs were extracted from each culture. PCR amplifications were performed on total DNA as previously described in *Supporting information* by using the specific primers targeting the bacteriovirus MPV1 and the plasmid DNA pMP1.

4.9. QPCR assays for quantifications of the copy numbers of chromosomal, proviral and plasmid target genes.

A culture of *M. piezophila* KA3 was split into two aliquots. One aliquot was induced with mitomycin C (5 µg/mL) at early-log growth phase whereas the second was not induced. The cultures were incubated for 80 min at 65°C and total DNA was extracted. The DNA concentration was estimated by NanoDrop ND-1000 and diluted to a concentration of 0.2 ng/µL. Real-time PCR amplifications were performed in triplicate using Syber Green Supermix (Taq Core, Qiagen) according to the manufacturer's instructions with modifications. Primers sets are the same than those previously used; real-time PCR protocol is described in *Supporting information*. The copy numbers of target genes were calculated from mitomycin C treated and non-treated cultures, respectively.

4.10. Sequencing and assembly of bacteriovirus and plasmid DNA.

Purified viral DNA and plasmid DNA were sequenced on Ion Torrent Personal Genome Machine (PGM) using the Ion PGM sequencing 200 kit v.2 (LifeTechnologies) and the Ion Xpress Barcode adapters 1-16 kit (LifeTechnologies). The sequences were assembled using the CLC Genomics Workbench 6.5.1 (http://www.clcbio.com) as well as MIRA 4 (Chevreux *et al*., 1999). Sequence analyses of protein-coding genes of bacteriovirus and pMP1 are described in *Supporting information*.

Acknowledgments

This work was financially supported by the French Ministry of Higher Education to JL and the Agence Nationale de la Recherche, project Thermovésicles (ANR-12-BSV3-OO23-01) to CG, a Norwegian Research Council award (project no. 180444/V40) to CLN, and Dartmouth's Walter and Constance Burke Research Initiation Award and the Dean of Faculty start-up funds to OZ. We thank S. Dupont for his technical help and G. Sinquin for assistance in electron microscopy.

Conflict of interest

The authors declare no conflict of interest.

References

Akimkina, T., Ivanov, P., Kostrov, S., Sokolova, T., Bonch-Osmolovskaya, E.A., Firman, K., *et al*. (1999) A Highly Conserved Plasmid from the Extreme Thermophile *Thermotoga maritima* MC24 Is a Member of a Family of Plasmids Distributed Worldwide. *Plasmid* **42:** 236–240.

Alain, K., Marteinsson, V.T., Miroshnichenko, M.L., Bonch-Osmolovskaya, E.A., Prieur, D., and Birrien, J.-L. (2002) *Marinitoga piezophila* sp. nov., a rod-shaped, thermo-piezophilic bacterium isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* **52:** 1331–1339.

Anderson R.E., Brazelton W.J., and Baross, J.A. (2011a). Is the genetic landscape of the deep subsurface biosphere affected by viruses? *Front Microbiol* **2:** 219. doi: 10.3389/ fmicb.2011.00219.

Anderson R.E., Brazelton W.J., and Baross, J.A. (2011b). Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol Ecol* **77:**120–133.

Arnold, H.P., She, Q., Phan, H., Stedman, K., Prangishvili, D., Holz, I., *et al*. (1999) The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Mol Microbiol* **34:** 217–226.

Bimboim, H., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7:** 1513–1523.

Bonch-Osmolovskaya, E.A. (2008) *Thermotogales*. In, *Encyclopedia of Life Sciences (ELS)*. Chichester: John Wiley & Sons, Ltd.

Bravo, A., and Alonso, J.C. (1990) The generation of concatemeric plasmid DNA in *Bacillus* subtilis as a consequence of bacteriophage SPP1 infection. *Nucleic Acids Res* **18:** 4651–4657.

Briani, F. Deho, G., Forti, F., and Ghisotti, D. (2001) The plasmid status of satellite bacteriophage P4. *Plasmid* **45:** 1–17.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., and Brüssow, H. (2003) Prophage genomics. *Microbiol Mol Biol Rev* **67:** 238–276.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A.J., Müller, W.E., Wetter, T., and Suhai, S. (1999) Genome sequence assembly using trace signals and additional sequence information. In *Computer Science and Biology: German Conference on Bioinformatics (GCB)* 45, pp. 45–56.

Christie, G.E., and Dokland, T. (2012) Pirates of the *Caudovirales*. *Virology* **434:** 210–221.

Coren, J.S., Pierce, J.C., and Sternberg, N. (1995) Headful packaging revisited: the packaging of more than one DNA molecule into a bacteriophage P1 head. *J Mol Biol* **249:** 176–184.

Desnues, C., La Scola, B., Yutin, N., Fournous, G., Robert, C., Azza, S., *et al*. (2012) Provirophages and transpovirons as the diverse mobilome of giant viruses. *Proc Natl Acad of Sci* **109:** 18078–18083.

Feiss, M., and Rao, V.B. (2012) The bacteriophage DNA packaging machine. In *Viral Molecular Machines*. Rossmann, M.G., and Rao, V.B. (eds). New York City, USA: Springer, pp. 489–509.

Flores, G.E., and Reysenbach, A.-L. (2011) Hydrothermal Environments, Marine. In *Encyclopedia of Earth Science Series: Encyclopedia of Geobiology*. Reitner, J., and Thiel, V. (eds). Dordrecht, The Netherlands: Springer, pp. 456–467.

Geslin, C., Le Romancer, M., Erauso, G., Gaillard, M., Perrot, G., and Prieur, D. (2003a) PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, *Pyrococcus abyssi*. *J Bacteriol* **185:** 3888–3894.

Geslin, C., Le Romancer, M., Gaillard, M., Erauso, G., and Prieur, D. (2003b) Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. *Res Microbiol* **154:** 303–7.

Geslin, C., Gaillard, M., Flament, D., Rouault, K., Le Romancer, M., Prieur, D., *et al.* (2007) Analysis of the first genome of a hyperthermophilic marine virus-like particle, PAV1, isolated from *Pyrococcus abyssi*. *J Bacteriol* **189:** 4510–4519.

Gill, S., Krupovic, M., Desnoues, N., Béguin, P., Sezonov, G., and Forterre, P. (2014) A highly divergent archaeo-eukaryotic primase from the *Thermococcus nautilus* plasmid, pTN2. *Nucleic Acids Res* **42:** 3707–3719.

Guy, L., Kultima, J.R., and Andersson, S.G. (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26:** 2334–2335.

Gorlas, A., Koonin, E.V., Bienvenu, N., Prieur, D., and Geslin, C. (2012) TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus*. *Environ Microbiol* **14:** 503–516.

Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* **313**: 903–919.

Hania, W.B., Postec, A., Aüllo, T., Ranchou-Peyruse, A., Erauso, G., Brochier-Armanet, C., *et al*. (2013) *Mesotoga infera* sp. nov., a mesophilic member of the order *Thermotogales*, isolated from an underground gas storage aquifer. *Int J Syst Evol Microbiol* **63:** 3003–3008.

Harriott, O.T., Huber, R., Stetter, K.O., Betts, P.W., and Noll, K.M. (1994) A cryptic miniplasmid from the hyperthermophilic bacterium *Thermotoga* sp. strain RQ7. *J bacteriol* **176:** 2759–2762.

Hashemolhosseini, S., Montag, D., Krämer, L., and Henning, U. (1994) Determinants of receptor specificity of coliphages of the T4 family: a chaperone alters the host range. *J Mol Biol* **241:** 524–533.

Krupovic, M., Prangishvili, D., Hendrix, R.W., and Bamford, D.H. (2011) Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol Mol Biol Rev* **75:** 610–635.

Leffers, G., and Basaveswara Rao, V. (1996) A discontinuous headful packaging model for packaging less than headful length DNA molecules by bacteriophage T4. *J Mol Biol* **258:** 839–850.

Lindqvist, B.H., Dehò, G., and Calendar, R. (1993) Mechanisms of genome propagation and helper exploitation by satellite phage P4. *Microbiol Rev* **57:** 683–702.

Liu, B., Wu, Suijie, Song, Q., Zhang, X., and Xie, L. (2006) Two novel bacteriophages of thermophilic bacteria isolated from deep-sea hydrothermal fields. *Curr Microbiol* **53:** 163–166.

Liu, B., and Zhang, X. (2008) Deep-sea thermophilic *Geobacillus* bacteriophage GVE2 transcriptional profile and proteomic characterization of virions. *Appl Microbiol Biotechnol* **80:** 697–707.

Liu, B., Zhou, F., Wu, Suijie, Xu, Y., and Zhang, X. (2009) Genomic and proteomic characterization of a thermophilic *Geobacillus* bacteriophage GBSV1. *Res Microbiol* **160:** 166–171.

Lucas, S., Han, J., Lapidus, A., Cheng, J.-F., Goodwin, L.A., Pitluck, S., *et al*. (2012) Complete genome sequence of the thermophilic, piezophilic, heterotrophic bacterium *Marinitoga piezophila* KA3. *J Bacteriol* **194:** 5974–5975.

Marti, R., Zurfluh, K., Hagens, S., Pianezzi, J., Klumpp, J., and Loessner, M.J. (2013) Long tail fibres of the novel broad-host-range T-even bacteriophage S16 specifically recognize *Salmonella* OmpC. *Mol Microbiol* **87:** 818–834.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., *et al*. (1999) Evidence for lateral gene transfer between *Archaea* and *Bacteria* from genome sequence of *Thermotoga maritima*. *Nature* **399:** 323–329.

Nesbø, C.L., Bradnan, D.M., Adebusuyi, A., Dlutek, M., Petrus, A.K., Foght, J., *et al*. (2012) *Mesotoga prima* gen. nov., sp. nov., the first described mesophilic species of the *Thermotogales*. *Extremophiles* **16:** 387–393.

Noble, R.T., and Fuhrman, J.A. (1998) Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat Microb Ecol* **14:** 113–118.

Petrus, A.K., Swithers, K.S., Ranjit, C., Wu, Si, Brewer, H.M., Gogarten, J.P., *et al*. (2012) Genes for the major structural components of *Thermotogales* species' togas revealed by proteomic and evolutionary analyses of OmpA and OmpB homologs. *PloS one* **7:** e40236.

Prieur, D., Erauso, G., Geslin, C., Lucas, S., Gaillard, M., Bidault, A., *et al*. (2004) Genetic elements of *Thermococcales*. *Biochem Soc Trans* **32:** 184–187.

Ram, G., Chen, J., Kumar, K., Ross, H.F., Ubeda, C., Damle, P.K., *et al*. (2012) *Staphylococcal* pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc Natl Acad Sci* **109:** 16300–16305.

Ravot, G., Magot, M., Fardeau, M.-L., Patel, B., Prensier, G., Egan, A., *et al*. (1995) *Thermotoga elfii* sp. nov., a novel thermophilic bacterium from an African oil-producing well. *Int J Syst Bacteriol* **45:** 308–314.

Six, E.W. (1975) The helper dependence of satellite bacteriophage P4: which gene functions of bacteriophage P2 are needed by P4? *Virology* **67:** 249–263.

Wang, I.-N., Smith, D.L., and Young, R. (2000) Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol* **54:** 799–825.

Wang, Y., and Zhang, X. (2010) Genome analysis of deep-sea thermophilic phage D6E. *Appl Environ Microbiol* **76:** 7861–7866.

Wang, Y., Duan, Z., Zhu, H., Guo, X., Wang, Z., Zhou, J., *et al*. (2007) A novel *Sulfolobus* non-conjugative extrachromosomal genetic element capable of integration into the host genome and spreading in the presence of a fusellovirus. *Virology* **363:** 124–133.

Williamson, S.J., Cary, S.C., Williamson, K.E., Helton, R.R., Bench, S.R., Winget, D., and Wommack, K.E. (2008) Lysogenic virus-host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *Int J Syst Evol Microbiol* **2:** 1112–1121.

Yasmin, A., Kenny, J.G., Shankar, J., Darby, A.C., Hall, N., Edwards, C., and Horsburgh, M.J. (2010) Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages. *J Bacteriol* **192:** 1122–1130.

Yoshida-Takashima, Y., Takaki, Y., Shimamura, S., Nunoura, T., and Takai, K. (2013) Genome sequence of a novel deep-sea vent epsilonproteobacterial phage provides new insight into the co-evolution of *Epsilonproteobacteria* and their phages. *Extremophiles* **17:** 405–419.

Young, R. (2013) Phage lysis: do we have the hole story yet? *Curr Opin Microbiol* **16:** 790–797.

Zhaxybayeva, O., Swithers, K.S., Foght, J., Green, A.G., Bruce, D., Detter, C., *et al*. (2012) Genome sequence of the mesophilic *Thermotogales* bacterium *Mesotoga prima* MesG1. Ag. 4.2 reveals the largest *Thermotogales* genome to date. *Genome Biol Evol* **4:** 812–820.

Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., *et al*. (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the *Thermotogales*. *Proc Natl Acad Sci* **106:** 5865–5870.

Figures

Fig. 1. Electron micrograph of two MPV1 particles negatively stained with 2% uranyl acetate. The viruses were cultivated at atmospheric pressure.

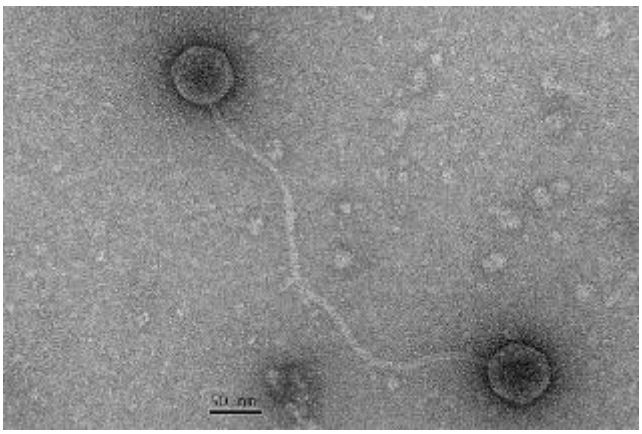


Fig. 2. Virus-host kinetics at atmospheric pressure monitored by flow cytometry. Typical growth curve of the host *M. piezophila* is shown as filled squares on continuous black line, and spontaneous production of MPV1 particles, reaching 1.11 x 10^8 viral particles per mL, is shown as open squares on dotted black line. Curves in grey correspond to the cultures grown with the addition of 5 μg/mL of mitomycin C (depicted with the arrow): filled triangles on continuous line show *M. piezophila* growth, while open triangles on dotted line refer to parallel MPV1 viral production, reaching ~3 x10^9 viral particles per mL.

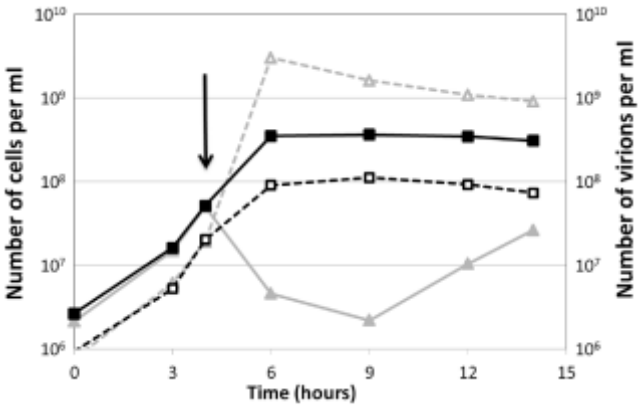


Fig. 3. Comparison of the MPV1 provirus in *M. piezophila* and viral DNA. The regions with significant pairwise BLASTN similarity scores are connected. Informative ORF annotations are shown. The ORF highlighted in green has a homolog in the pMP1 plasmid (see Fig. 4). The ORF marked with a star is disrupted in the linearized genome packaged in the virions. The black dots indicate the short sequences with significant sequence similarity to the regions of the pMP1 genome. The division of the virus into functional modules is approximate, since it is based only on gene annotations. The figure was produced using genoPlotR (Guy *et al.*, 2010).

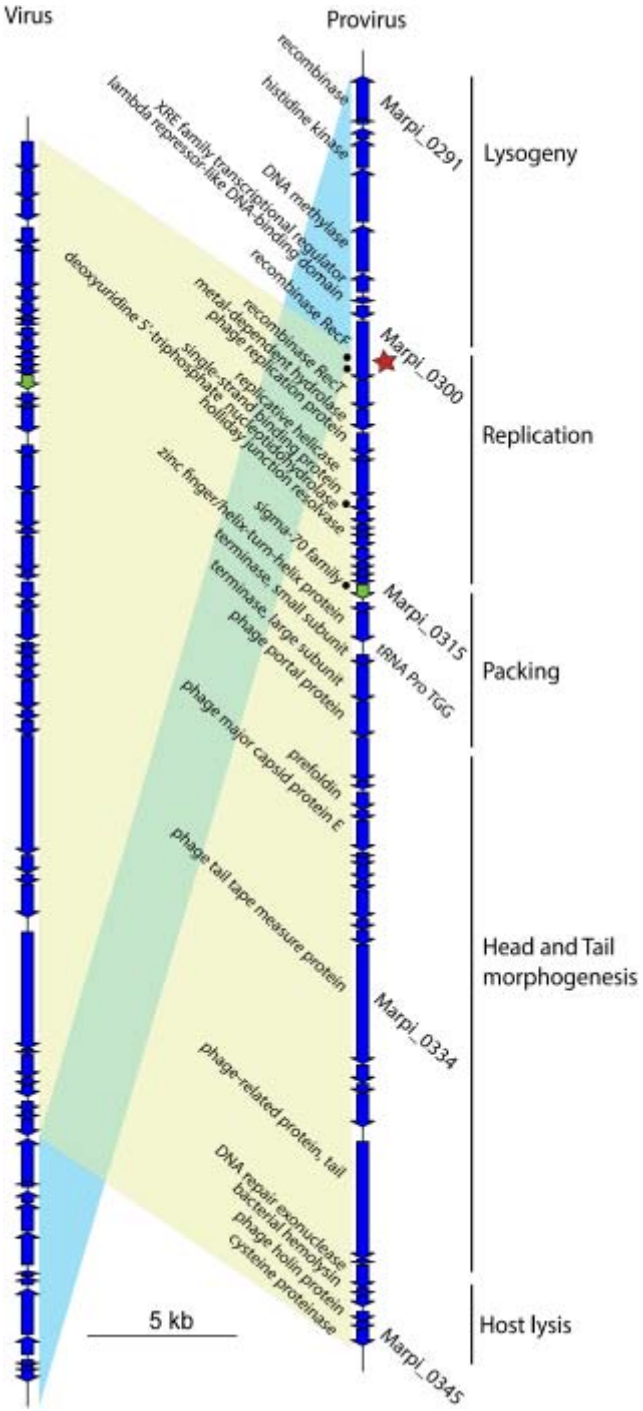


Fig. 4. Map of the pMP1 plasmid. Only informative ORF annotations are shown. The ORF highlighted in green has a homolog in the MPV1 genome (see Fig. 3). The black dots indicate the short sequences with significant similarity to the MPV1 genome. The figure was created in Geneious 7 (Biomatters Limited).

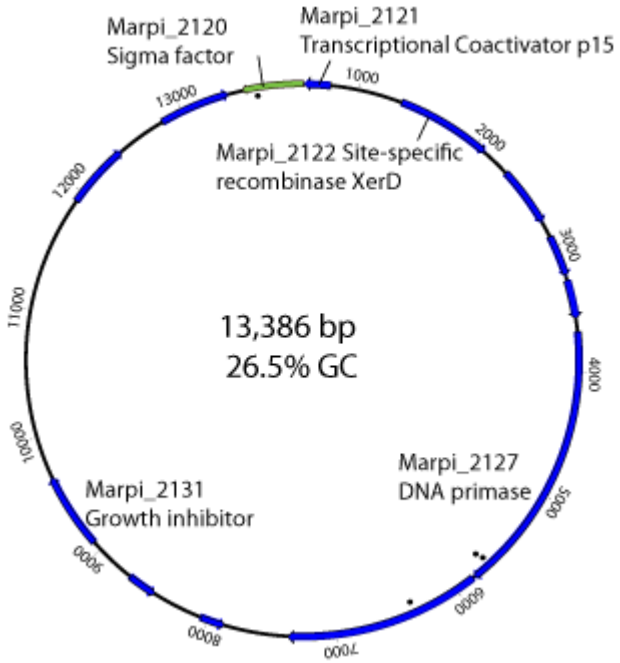


Fig. 5. MPV1 infectivity assays performed on *Thermosipho* sp. AT1244-VC14^T and checked by PCR experiments. Lanes 1, 2: PCR on total DNA of the uninfected *Thermosipho* strain targeting (1) MPV1 viral DNA or (2) pMP1 plasmid DNA. Lanes 3, 4: PCR on total DNA of the *Thermosipho* strain just after addition of MPV1 purified capsids targeting (3) MPV1 viral DNA or (4) pMP1 plasmid DNA. Lanes 5, 6: PCR on total DNA of *Thermosipho* strain 22 hours after addition of MPV1 purified capsids targeting (5) MPV1 viral DNA or (6) pMP1 plasmid DNA. Lanes 7, 9: PCR on total DNA of the *Thermosipho* strain after 3 subcultures targeting (7) MPV1 viral DNA or (9) pMP1 plasmid DNA. Lanes 10, 12: Negative controls targeting (10) MPV1 viral DNA or (12) pMP1 plasmid DNA. Lanes 11, 13: Positive controls targeting (11) MPV1 viral DNA or (13) pMP1 plasmid DNA. Lanes 8: Smart Ladder 1 kb (Eurogentec).

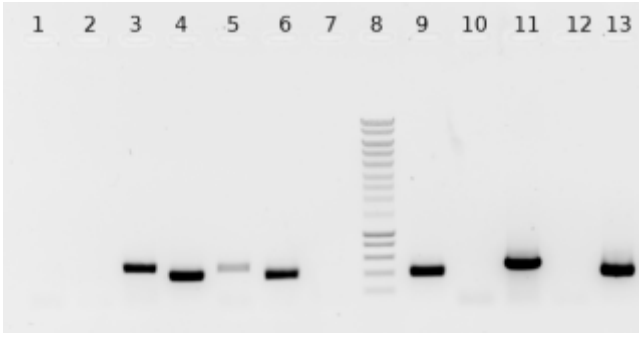
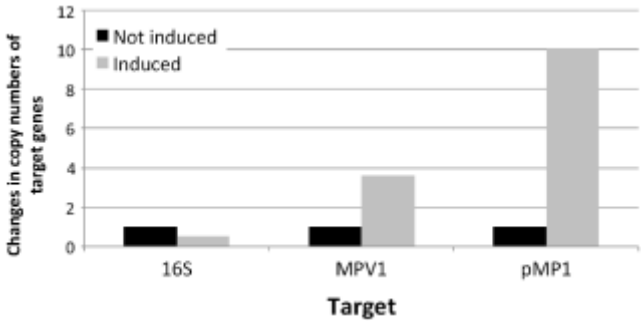


Fig. 6. qPCR assays to assess changes in copy number of *M. piezophila* chromosomal, proviral and plasmid DNA target genes in the presence or absence of mitomycin C induction (with 5 µg/mL of mitomycin C). Assays without mitomycin C (the black bars) have been set to a value of 1. The light grey bars represent the relative changes in copy number after the induction.



Supplementary experimental procedures

Screening of *Thermotogales* strains for mobile genetic elements. A total of 60 strains of *Thermotogales* were retrieved from the UBO culture collection "UBOCC" (www.univ-brest.fr/UBOCC) and from our laboratory cultures collections. These strains were isolated from deep-sea vents in the Lau Basin (latitude, 22°32'S; longitude, 176"43'W; depth, 1832 to 1887 m), the East Pacific Rise (latitude, 12°48721'N; longitude, 103°56351'W; depth, 2630 m) and the Middle Atlantic Ridge (latitude, 36°139N; longitude, 33°549W; depth 2775 m).

Culturing methods. *Thermotogales* strains were cultured in a modified Ravot medium which contained (per liter of distilled water) 0.2 g of NH_4Cl, 0.5 g of $MgCl_2.6H_2O$, 0.1 g of $CaCl_2.2H2O$, 0.5 g of KCl, 0.83 g of $NaCH_3COO.3H_20$, 2 g of yeast extract, 2 g of tryptone, 30 g of sea salt, 3.3 g of piperazine-*N*,*N*-bis(2-ethanesulfonic acid) (PIPES), 2 g de maltose and 0.001 g of resazurin. The pH was adjusted to 6 and the medium was sterilized by autoclaving. After the medium cooled down, the following solutions, separately sterilized by autoclaving, were added: 5 ml of a 7 % (wt/vol) KH_2PO_4 solution and 5 ml of 7 % (wt/vol) K_2HPO_4 solution. The medium was dispensed (50 ml) into 100-ml sterile vials and completed by adding 1% (wt/vol) elemental sulfur previously sterilized by steaming at 100°C for 1 h on three successive days. Anaerobiosis was obtained by applying vacuum to the medium and saturating it with N_2. Finally, the medium was reduced by adding a sterile solution of $Na_2S.9H_2O$ (final concentration of 0.05% [wt/vol]). The medium inoculated, to the final concentration of 1%, was incubated at 60°C, 65°C and 70°C depending to the strains. *Marinitoga piezophila* KA3 is piezophile as its optimum growth occurring at 40 MPa. However, after several subcultures at 0.1 MPa cells are adapted to atmospheric pressure. Most of the experiments were consequently performed with *M. piezophila* KA3 cultured at atmospheric pressure and 65°C with L-cystine added as electron acceptor. The high hydrostatic pressure experiments were performed in 5 ml syringes under 40 MPa pressure at 65 °C. The syringe was loaded anaerobically with 5 ml of reduced and modified Ravot medium and inoculated to a final concentration of 1% with a *M. piezophila* KA3 culture in late exponential phase that was grown under high pressure. The syringes were incubated at high pressure (40 MPa) and temperature (65°C) in HP/HT systems, custom-built by Top Industrie.

Primers design and PCR protocol. To confirm the nature of the DNA packaged in the viral capsids and to check potential chromosomal contamination, PCR amplifications were performed. Three primer sets were designed by using the Primer 3 software (Untergasser *et al*., 2007). The first set is specific to *Marinitoga piezophila* KA3: 16S rRNA gene (forward: 5'-ACACATGCAAGTCGAACGAG-3' and reverse: 5'- CACTGGAAACGGTGGCTAAT-3'), the second set is specific to the major capsid protein gene *Marinitoga piezophila* KA3 MPV1 provirus: (forward: 5'-AGGAGGACCTCAACCAACAA-3' and reverse: 5'-CAAGCACGAGATTTGAGTGG-3'). The last set is specific to "hypothetical protein" gene encoded on *Marinitoga piezophila* KA3 pMP1 plasmid: (forward: 5'-TATCTCGGGAACAGCTCCAA-3' and reverse: 5'-TGACAAAATGGAACATGTTTTTG- 3'). The PCR reactions were carried out in a volume of 50 µL containing 100 ng template, 0.4 µM of each primer, 0.8 µM dNTPs, 1.5 mM $MgCl_2$, 1 x buffer and 0.024 U polymerase (Taq Core, Qiagen). The products sizes were 118 bp for 16S rRNA gene, 198 bp for MPV1 major capsid protein gene and 198 bp for pMP1 hypothetical protein gene.

Construction of standard curves for *Marinitoga piezophila* KA3 chromosome, plasmid pMP1 and provirus MPV1 quantifications by QPCR. QPCR analyses were performed to assess the changes in the replication rates, after mitomycin C induction, of both plasmid (by

targeting a gene encoding a hypothetical protein specific to pMP1, Marpi_2124) and viral DNA (by targeting the gene encoding the MPV1 major capsid protein, Marpi_326). In the different experiments with or without induction, the replication rates of the chromosome (by targeting the 16S rRNA gene) were also monitored as a control.

Primers sets are the same as those described above. The PCR products were obtained by conventional PCR and the products sizes were 118 bp for 16S rRNA gene, 198 bp for MPV1 major capsid protein gene and 198 bp for pMP1 hypothetical protein gene. The reactions were performed at 60°C in a volume of 50µL containing 100 ng of *Marinitoga piezophila* KA3 total DNA, 0.4 µM of each primer, 0.8 µM dNTPs, 1.5 mM $MgCl_2$, 1 x buffer and 2 U Pfu polymerase. The PCR products were deposited on agarose gel, purified by "GeneJet Gel Extraction Kit" (ThermoScientific) and cloned in pUD plasmid used to transform recombinant *E. coli* DH5α strains that were cultivated in LB medium at 37°C with addition of 100 µg/mL of ampicillin. The cloned plasmids were extracted using "GeneJET Plasmid Midiprep" (ThermoScientific) kit and purified as previously described. The concentration of the three different plasmids was measured using a NanoDrop ND-1000 and the corresponding copies numbers were calculated using the following equation: DNA copy = 6.02 x 1023 x DNA concentrations / 660 x DNA length (pUD + insert). 10-fold serial dilutions were performed for the 3 types of cloned plasmids pUD, ranging from 2×10^4 to 2×10^{10} copies/µL to construct the standard curves of genes targeting the chromosome and the provirus and ranging from 2×10^5 to 2×10^{11} copies/µL, for the gene targeting the cccDNA. The R^2 of standard curves obtained were up to 0.997 and efficiency of the reactions up to 97%.

Preparation of template DNA for QPCR analysis. Two cultures of *Marinitoga piezophila* KA3 were established using the same inoculum. The first one was induced by addition of mitomycin C (5 µg/mL) in the early exponential growth phase whereas the second culture was not induced. The two cultures were incubated for 80 minutes after induction and then total DNA extractions were performed. The concentration of the extracted DNA was measured by using a Nanodrop ND-1000 and diluted in order to reach 0.2 ng/µL.

QPCR assays by using Sybr Green I dye. The QPCR assays were performed using a StepOnePlus Instrument (Applied Biosystem). The QPCR reactions were performed in a volume of 25µL containing 1ng template, 12.5 µL of SybrGreen Supermix (Taq Core, Qiagen) and 0.9 µM of each primers The thermal cycling protocol was as follows : 10 min at 95°C then 40 cycles of 15 s at 95°C, 1 min at 60°C and 10 s at 72°C.

Sequencing and assembly of bacteriovirus and plasmid DNA. The provirus was detected using Prophinder (Lima-Mendez *et al*., 2008). Homologs of the predicted proteins in both MPV1 bacteriovirus and pMP1 plasmid genomes were extracted from the RefSeq and viral RefSeq databases using BLASTP (Altschul *et al*., 1997) with an E-value cutoff of 10^{-4}. For ORFs with no or few matches in RefSeq database, the *nr* database was also searched. Matches spanning at least 70% of a query were retained. For datasets with > 100 hits, only one representative per genus was kept. Homologs were aligned using ClustalW 2 (Larkin et al., 2007), and phylogenetic trees were reconstructed in FastTree (Price *et al*., 2009) under JTT+CAT model. Trans-membrane helices, helix-turn-helix motifs, coiled coils and isoelectric point for each ORF were predicted using programs from the EMBOSS package (Rice et al., 2000). ORFs were assigned to gene families using Superfamily (Gough *et al*., 2001) and Conserved Domains (Marchler-Bauer *et al*., 2009) databases. Predicted ORFs of the MPV1 genome were also screened for the features of the genes involved in known lysis pathways of double-stranded bacterioviral DNA of Gram-negative bacteria (Wang *et al*., 2000; Young, 2013). Analysis of amino acid composition and predictions of protein charge

and transmembrane helices were performed using the PEPSTATS, CHARGE and TMAP programs of the EMBOSS package (Rice *et al*., 2000). Protein secondary structure was predicted using JPred 3 (Cole *et al*., 2008).

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Cole C, Barber JD, Barton GJ. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **36:** 197–201.

Gough J, Karplus K, Hughey R, Chothia C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* **313:** 903–919.

Larkin M, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–2948.

Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. (2008). Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics* **24:** 863–865.

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, *et al.* (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* **37:** 205–210.

Price MN, Dehal PS, Arkin AP. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26:** 1641–1650.

Rice P, Longden I, Bleasby A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet* **16:** 276–277.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35:** 71–74.

Wang I-N, Smith DL, Young R. (2000). Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol* **54:** 799–825.

Young R. (2013). Phage lysis: do we have the hole story yet? *Curr Opin Microbiol* **16:** 790–797.

Supporting information

Figures and tables

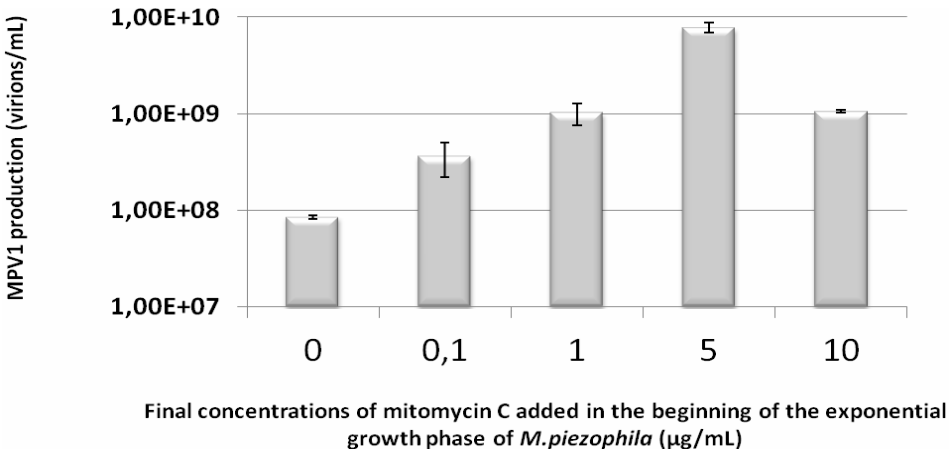


Fig. S1. Induction assays of MPV1 production by addition of several mitomycin C concentrations. Several final concentrations of mitomycin C (0.1 µg/mL, 1 µg/mL, 5 µg/mL and 10 µg/mL) were added to cultures of *M. piezophila* at mid-log growth phase. Aliquots were collected in duplicate to determine viral production by epifluorescence microscopy. The high of bars correspond to the averages of the replicates. Error bars show one standard deviation.

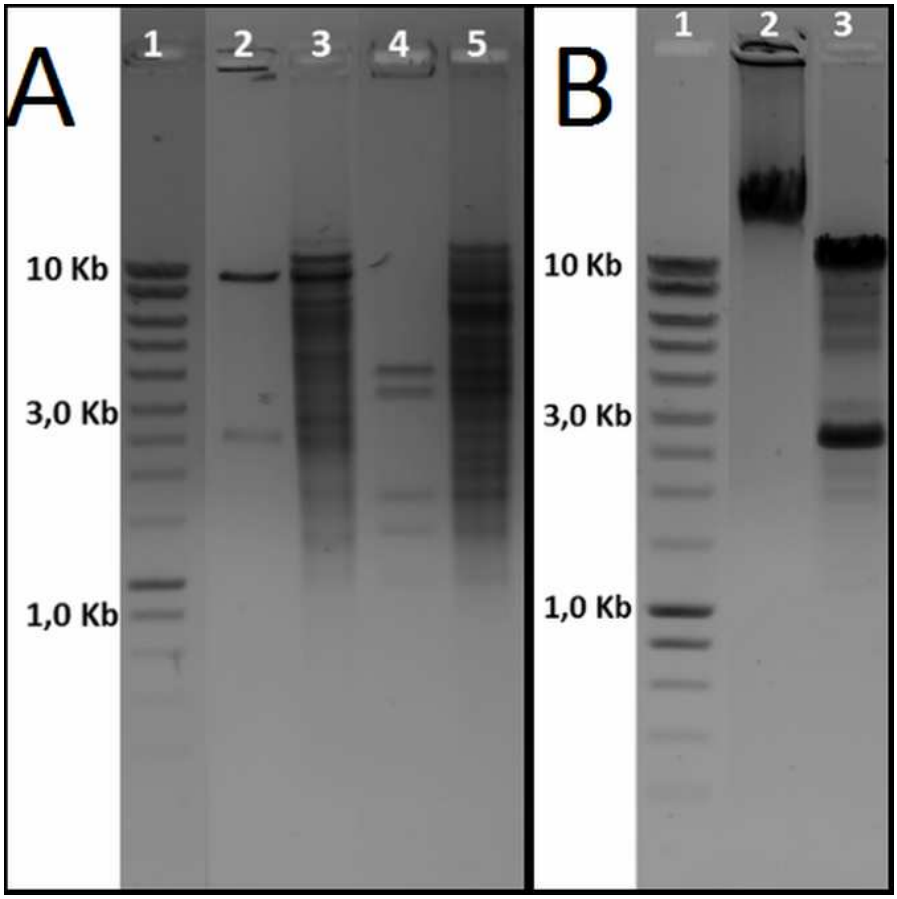


Fig. S2. RFLP analyses of *M. piezophila* total DNA, plasmid DNA (pMP1) and viral DNA. **A.** RFLP analyses of both *M. piezophila* total DNA and plasmid DNA (pMP1).1: DNA Smart Ladder; 2: plasmid DNA restricted by HhaI; 3: total DNA restricted by HhaI; 4: plasmid DNA restricted by HindIII; 5: total DNA restricted by HindIII.

B. RFLP analyses of the total DNA extracted from the purified viral capsids.1: DNA Smart Ladder; 2: Packaged DNA uncut; 3: Packaged DNA restricted by HhaI. The two thickest bands (of ~10,000 bp and 2,600 bp) correspond to the plasmid DNA restriction pattern (see lane 2 on the panel A). The weaker bands correspond to the viral sequence, since several of them are identical to those obtained after an *in silico* restriction of the *M. piezophila* provirus (not shown).

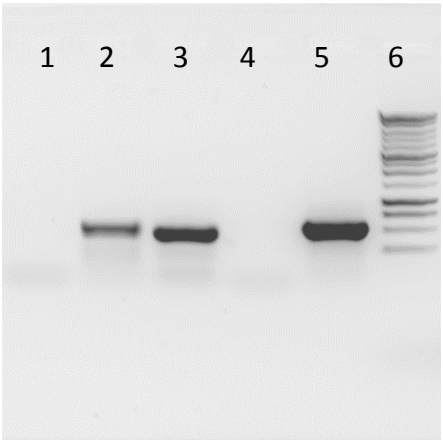


Fig. S3. PCR analysis of the total DNA extracted from purified MPV1 viral capsids from an induced culture. 1: Packaged DNA amplified by 16S rRNA primers; 2: Packaged DNA amplified by pMP1 plasmid primers; 3: Packaged DNA amplified by MPV1 proviral primers; 4: 16S rRNA negative control; 5: 16S rRNA positive control; 6: DNA Smart Ladder.

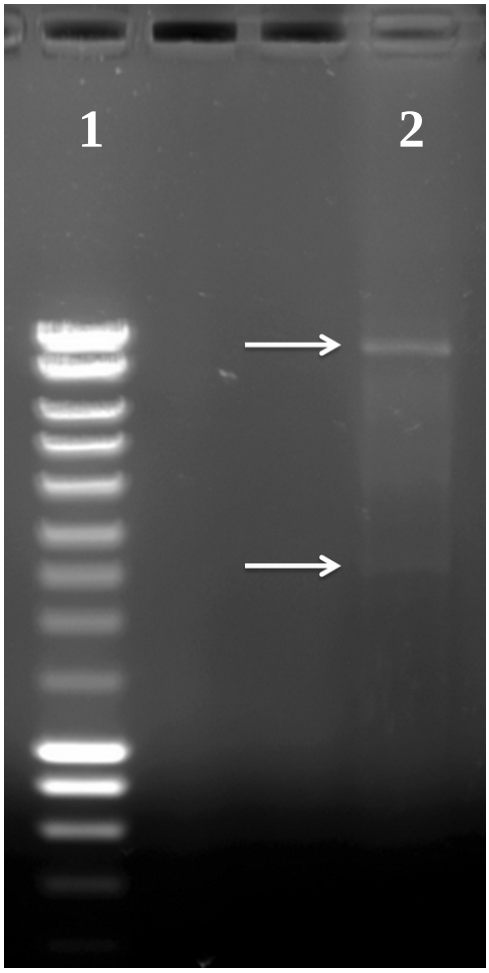


Fig. S4. Agarose gel electrophoresis of plasmid DNA extracted by alkaline lysis from *Thermosipho* sp. strain AT1244-VC14, three subcultures after being infected by MPV1 virions. 1: DNA Smart Ladder; 2: plasmid DNA restricted by HhaI. The two bands indicated by arrows correspond to pMP1 restriction pattern obtained by HhaI (Figure S2 panel A, lane 2).

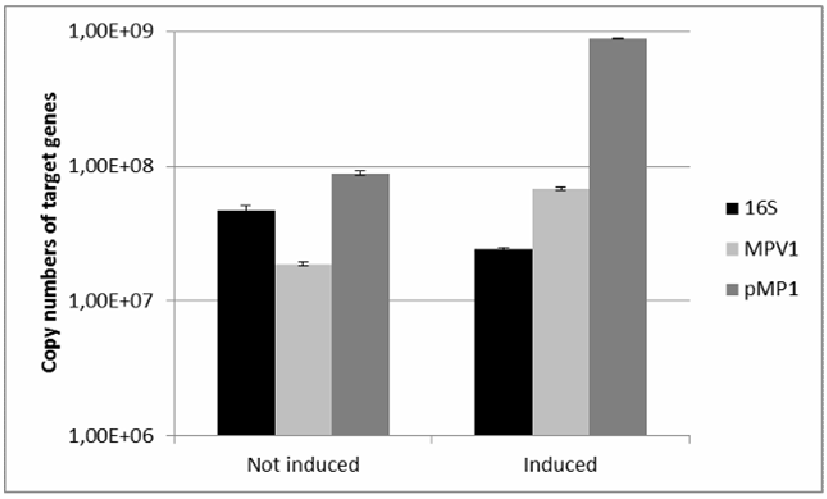


Fig. S5. QPCR assays. Copy number of target genes in the presence or absence of mitomycin C induction. The black bars represent the number of copies of the 16S rRNA chromosomal gene. The light grey bars represent the number of copies of the MPV1 viral gene. The dark grey bars represent the number of copies of the pMP1 plasmid gene.

Table S1 Comparison of viral production at 0.1 and 40 MPa monitored by flow cytometry.

Time	Bacterial counts (cells/mL)		Viral counts (virions/mL)	
	0.1 MPa	40 MPa	0.1 MPa	40 MPa
t = 0h	2.63×10^6	1.41×10^6	9.22×10^5	5.79×10^5
t = 4h	5.13×10^7	4.31×10^7	2.01×10^7	9.14×10^6
t = 9h	3.63×10^8	3.74×10^8	1.10×10^8	3.39×10^7

Table S2. Annotation of the MPV1 provirus.

			BLAST analyses			Protein predictions							Phylogenetic analyses[j]	
ORF[a]	length[b]	annotation[c]	BlastP match in nr[d]	phage[e]	*Thermotogae*[f]	Superfamily	CD Acc.[h]	CD name	Tmap	HTH	CC[i]	pI	Group containing sister taxa	Sister taxa members
291	1,470	Site-specific recombinases, DNA invertase Pin homologs	WP 006569700 site-specific recombinase, DNA invertase Pin *Thermoanaerobacter siderophilus*	x	Tmel 0397 (Tmel 1486)	Resolvase-like	cd00338	Ser_Recombinase			1	9.7682	*Enterococcus* phage	*Enterococcus* phage phiFL2A, *Enterococcus* phage phiFL1A, *Enterococcus* phage phiFL3A
292	162	hypothetical protein	No match									4.3022	*	*
293	372	hypothetical protein	YP002728383 hypothetical protein SULAZ 0390 *Sulfurihydrogenibium azorense* Az-Fu1				pfam14213	DUF4325				10.3206	*Sulfurihydrogenibium*	*Sulfurihydrogenibium azorense*
294	966	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase./STAS domain.	YP004027214 ATP-binding region ATPase domain-containing protein *Caldicellulo siruptor kristjanssonii*	x		SpoIIaa-like	pfam13466	STAS_2	1			4.8478	**	**
295	1,800	T5orf172 domain.	YP065302 hypothetical protein DP1566 *Desulfotalea psychrophila*	x			smart00974	T5orf172	2		5	9.3993	*Cystobacter*	*Cystobacter fuscus*
296	1,557	Adenine specific DNA methylase Mod	WP008474462 DNA methylase N-4/N-6 domain protein *Nitrolancetus hollandicus*	x		S-adenosyl-L-methionine-dependent methyltransferases	pfam01555	N6_N4_Mtase				9.5555	Cyanobacteria	*Nodularia spumigena*, *Synechococcus* sp. PCC 7502
297	621	SOS-response transcriptional repressors (RecA-mediated autopeptidases)	WP006874092 hypothetical protein *Anaerotruncus colihominis*	x	(Tmel 1479, H17ap60334 04802)	LexA/Signal peptidase	cd00093	HTH_XRE		1		5.8055	Viruses and Firmicutes	*Staphylococcus aureus*, *Enterococcus dispar*, *Staphylococcus* phage 47, *Staphylococcus* phage 2638A, *Oenococcus oeni*, *Staphylococcus* phage 29, *Staphylococcus* phage 37, *Macrococcus caseolyticus*, *Staphylococcus* phage phi2958PVL, *Staphylococcus* phage SAP-26, *Staphylococcus* phage TEM123, *Staphylococcus* phage phiNM
298	210	Predicted transcriptional regulators	WP008783175 DNA-binding protein *Bifidobacterium* sp. 12 1 47BFAA	x	(Tmel 1478, H17ap60334 04807)	lambda repressor-like DNA-binding domains	cd00093	HTH_XRE		1		10.8154	Firmicutes	*Clostridium difficile*, *Ruminococcus*
299	450	hypothetical protein, contains possible packing site	No match		(H17ap60334 04812)						1	9.873	*Thermosipho*	*Thermosipho africanus*
300	2,046	hypothetical protein	WP003364581 recombinase RecF *Clostridium botulinum*	x		P-loop containing nucleoside triphosphate hydrolases	pfam13476	AAA_23			4	5.4559	Clostridiaceae	*Clostridium botulinum*, *Alkaliphilus metalliredigens*
301	369	Recombinational DNA repair protein (RecE pathway)	WP003407752 recombinase RecT *Clostridium butyricum*	x	(H17ap60334 04827, Tmel 1472)		pfam03837	RecT			1	5.1063	Clostridiales	*Clostridium sticklandii*, *Peptostreptococcus anaerobius*, *Finegoldia magna*
302	729	Metal-dependent hydrolases of the beta-lactamase superfamily I	YP003781312 phage-like hydrolase *Clostridium ljungdahlii*	x		Metallo-hydrolase/oxidoreductase	cl00446	Lactamase_B superfamily				7.7311	Clostridiales	*Clostridium pasteurianum*
303	768	phage conserved hypothetical protein, C-terminal domain	ADA79920 putative phage replication protein *Lactobacillus* phage phiPYB5	x		"Winged helix" DNA-binding domain[g]	pfam09524	Phg_2220_C				9.4949	*Carnobacterium*	*Carnobacterium* sp. 17-4
304	255	hypothetical protein	No match									8.8039	*	*
305	1,275	Replicative DNA helicase	YP001568017 replicative DNA helicase *Petrotoga mobilis*	x	Theba 0328, Pmob 0972, Marpi 1890, Kole 1515, Theth 0376, Kole 1515, THA 1353, THA 826, Tlet 0531, Fnod 0667, Ferpe 0238	P-loop containing nucleoside triphosphate hydrolases	cd00984	DnaB_C			1	6.0647	*Petrotoga*	*Petrotoga mobilis*
306	468	single stranded DNA-binding protein (ssb)	WP016288767 single-stranded DNA-binding protein *Lachnospiraceae* bacterium 3- 1	x	Marpi 0443, Theba 0951, Theth 1460, Kole 0314, CTN 0058, Pmob 0559, Tlet 0829, Tmel 0637	Nucleic acid-binding proteins	cd04496	SSB_OBF				4.4113	*Thermotogae*	*Marinitoga*, *Thermosipho*, *Thermotoga*, *Kosmotoga*, Mesotoga
307	420	deoxyuridine 5''-triphosphate nucleotidohydrolase (dut)	XP005142568: deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial *Melopsittacus undulatus*	x		dUTPase-like	cd07557	trimeric_dUTPase	1			9.8363	Synergistes	*Aminobacterium colombiense*

308	201	hypothetical protein	No match									9.8649	*	*
309	249	hypothetical protein	No match				cl06547	Shugoshin_N superfamily	2		1	10.4978	*	*
310	366	Holliday junction resolvase	YP003183942 endodeoxyribonuclease RusA *Alicyclobacillus acidocaldarius*	x		Holliday junction resolvase RusA	pfam05866	RusA				10.1709	Various Bacteria	*Lactobacillus mali, Leptotrichia hofstadii*
311	465	hypothetical protein	No match									4.55	*	*
312	273	hypothetical protein	No match									4.5088	*	*
313	276	hypothetical protein	No match								1	5.6601	*	*
314	231	hypothetical protein	No match			Fzo-like conserved region[g]					1	8.979	*	*
315	483	hypothetical protein	YP005093729 sigma-70 family RNA polymerase sigma factor *Marinitoga piezophila* KA3 Marpi 2120 (Plasmid)		Marpi 2120, (Tmel 1467, H17ap60334 04862)	Sigma3 and sigma4 domains of RNA polymerase sigma factors	TIGR02937	sigma70-ECF		1		9.4095	*Marinitoga piezophila* (paralog on MP plasmid 01)	*Marinitoga piezophila* KA3
316	327	hypothetical protein	No match		Tnap 1136							4.303	*Thermotoga*	*Thermotoga naphthophila*
317	1,029	putative zinc finger/helix-turn-helix protein, YgiT family	YP003841324 helix-turn-helix domain-containing protein *Caldicellulosiruptor obsidiansis*	x	Pmob 1215 (best match), Tnap 1135	lambda repressor-like DNA-binding domains	cd00093	HTH_XRE	2	1	1	6.6932	*Petrotoga*	*Petrotoga mobilis*
318	75	tRNA Pro TGG												
319	381	hypothetical protein	WP004096102 hypothetical protein *Acetonema longum*, Among Phi blast matches: WP016077038.1 phage terminase small subunit *Lactobacillus* phage AQ113			Homeodomain-like	cl17451	HTH_23 superfamily		1		9.9187	**	**
320	1,248	phage terminase, large subunit, PBSX family	YP002804814 phage terminase large subunit, pbsx family *Clostridium botulinum*	x			cl12054	Terminase_3 superfamily				7.066	Clostridiales	*Clostridium difficile, Clostridium botulinum*
321	1,251	hypothetical protein	WP021769687 hypothetical protein *Leptotrichia* sp. oral taxon 22 second hit: WP006773947 phage portal protein *Clostridium hathewayi*	x			pfam05133	Phage_prot_Gp6	1		1	5.0359	**	**
322	1,506	hypothetical protein	WP003432484 hypothetical protein *Clostridium butyricum*	x			cl01681	DUF2115 superfamily	1		2	9.4814	**	**
323	282	hypothetical protein	No match									9.4531	*	*
324	603	hypothetical protein	No match	(x)		Prefoldin[g]					2	4.3817	*	*
325	354	hypothetical protein	WP017687985 hypothetical protein *Paenibacillus* sp. PAMC 26794	x			pfam02924	HDPD				4.7569	*Paenibacillus*	*Paenibacillus lactis*
326	1,020	Phage major capsid protein E.	WP004445976 capsid protein *Clostridium botulinum*	x			pfam03864	Phage_cap_E				8.7531	Viruses and Firmicutes	*Eubacteriaceae bacterium, Clostridium lentocellum* , *Paenisporosarcina* sp. HGH0030, *Peptoniphilus rhinitidis, Enterococcus* phage phiEf11, *Lactobacillus* prophage Lj965, *Oenococcus oeni, Paenibacillus lactis, Weissella confusa*
327	183	hypothetical protein	No match									10.0744	*	*
328	321	hypothetical protein	YP001320641 hypothetical protein Amet 2832 *Alkaliphilus metalliredigens*									4.38	Firmicutes	*Thermosediminibacter oceani, Alkaliphilus metalliredigens, Thermoanaerobacter ethanolicus, Clostridium perfringens*
329	474	hypothetical protein	WP006306523 hypothetical protein *Centipeda periodontii*	x								10.7346	Centipeda	*Centipeda periodontii*
330	387	hypothetical protein	No match			TRAP-like[g]						4.4597	*	*
331	966	hypothetical protein	WP002670872 hypothetical protein *Treponema denticola*		(Tmel 1452, H17ap60334 04952)				1			4.3066	*Thermosipho*	*Thermosipho melanesiensis* , *Thermosipho africanus*
332	351	hypothetical protein	No match		(Tmel 1451, H17ap60334 04957)							4.7969	*Thermosipho*	*Thermosipho* melanesiensis, *Thermosipho africanus*
333	498	hypothetical protein	No match		(Tmel 1450, H17ap60334 04962)							9.0927	*Thermosipho*	*Thermosipho melanesiensis*
334	4,080	phage tail tape measure protein, TP901 family, core region	YP001285821 putative tail tape measure protein *Geobacillus* virus E2	x	(Tmel 1449, H17ap60334 04967)	ARM repeat	pfam10145	PhageMin_Tail	9		2	7.0184	*Thermosipho*	*Thermosipho melanesiensis*
335	615	hypothetical protein	No match		Theba 1821							9.6954	*	*

336	339	hypothetical protein	No match		(Tmel 1448, H17ap60334 04972)						4.5118	*Thermosipho*	*Thermosipho melanesiensis*, *Thermosipho africanus*
337	1,131	hypothetical protein	No match		(Tmel 1447, H17ap60334 04977)	Concanavalin A-like lectins/glucanases[g]					4.6324	*Thermosipho*	*Thermosipho africanus*
338	4,011	hypothetical protein	WP016610917 phage minor structural protein *Enterococcus casseliflavus*	x	(Tmel 1446, H17ap60334 04982)	Fibronectin type III	cd00063	FN3		2	4.4423	**	**
339	162	hypothetical protein	No match			tRNA-binding arm[g]				1	9.0385	*	*
340	768	DNA repair exonuclease	WP018671617 hypothetical protein *Brevibacillus laterosporus*		(Tmel 1444) THA 1537, THA 1538	Metallo-dependent phosphatases	cd00838	MPP_superfamily			9.3745	**	**
341	261	hypothetical protein	No match			Bacterial hemolysins	pfam11853	DUF3373		1	6.5311	*	*
342	378	hypothetical protein	No match		(Tmel 1439)	Duffy binding domain-like[g]					9.1701	*Thermosipho*	*Thermosipho melanesiensis*
343	330	Phage holin protein (Holin LLH).	WP016226830 LL-H family phage holin *Lachnospiraceae* bacterium 10-1	x			cl09890	Holin_LLH superfamily	1	1	6.2626	*Streptococcus* phage (Viruses)	*Streptococcus* phage SM1
344	300	hypothetical protein	No match						1		4.861	*	*
345	564	Orthopoxvirus protein of unknown function (DUF830).	YP002454578 hypothetical protein BC G9842 A0016 *Bacillus cereus* G9842	x		Cysteine proteinases					5.6907	**	**

a) ORF is number in the genome sequence, i.e. 291 corresponds to Maripi_0291.

b) Length in bp.

c) Annotation in genome.

d) Best BlastP match outside the Thermotogales except for the ORF with match to the plasmid ORF which had no significant matches outside the Thermotogales. We also list additional matches where these added some information to the possible function of the protein.

e) Presence in other phage.

f) Presence in other Thermotogae genomes. If this was in another predicted prophage the ORF id is listed in parenthesis.

g) Weak HMM library classifications, with an E-value greater than 0.0001.

h) CD accession.

i) Number of Coiled Coils.

j) * indicates that no phylogenetic analysis could be done as there were no matches in Genbank. ** indicates that no phylogenetic analysis could be done as there were to few hits or the sequences were too divergent to align confidently.

Table S3. Summary of the similarity searches, protein properties' predictions and phylogenetic analyses of pMP1.

			BLAST analyses	Protein predictions							Phylogenetic analyses[h]	
ORF[a]	length[b]	annotation[c]	BlastP match in nr[d]	Superfamily	CD Acc.[f]	CD name	Tmap	HTH	CC[g]	pI	Group containing sister taxa	Sister taxa members
2120	480	Sigma-70,_region_4.	YP_005096064 hypothetical protein Marpi_0315 *Marinitoga piezophila* KA3 (Prophage)	Sigma3 and sigma4 domains of RNA polymerase sigma factors	cl17315	LuxR_C_like superfamily			0.886	9.8069	*Marinitoga*	*Marinitoga piezophila* KA3
2121	204	Transcriptional_Coactivator_p15_(PC4)	YP_008403878 hypothetical protein JCM7686_1010 *Paracoccus aminophilus*	ssDNA-binding transcriptional regulator domain	cl03492	PC4 superfamily				9.6314	Proteobacteria	Proteobacteria
2122	750	Site-specific_recombinase_XerD	YP_001739834 integrase family protein *Thermotoga* sp. R Q2	DNA breaking-rejoining enzymes	cl00213	DNA_BRE_C superfamily				10.3024	*Thermotoga*	*Thermotoga* spp.
2124	480	hypothetical_protein	No match							4.657	*	*
2125	330	hypothetical_protein	No match							5.2917	*	*
2126	294	hypothetical_protein	No match						0.623	5.7185	*	*
2127	2148	DNA_primase_(bacterial_type)	YP_004026647 DnaB domain containing protein *Caldicellulosiruptor kristjanssonii*	Zinc beta-ribbon						7.0274	**	**
2128	1527	hypothetical_protein	No match	Sigma2 domain of RNA polymerase sigma factors[e]	cd01029	TOPRIM_primases			0.884	9.9326	*	*
2129	198	hypothetical_protein	No match				1			10.3545	*	*
2130	231	hypothetical_protein	No match				1			10.1286	*	*
2131	603	Growth_inhibitor	YP_002335318 hypothetical protein THA_1536 *Thermosipho africanus* TCF52B	Cell growth inhibitor/plasmid maintenance toxic component						9.9314	Various Bacteria	*Desulfotomaculum nigrificans*, *Natranaerobius thermophilus*, *Thermosipho africanus_*, *Kosmotoga olearia*
2132	522	hypothetical_protein	No match	Ribbon-helix-helix	pfam02452	PemK		2		10.2669	*	*
2133	552	hypothetical_protein	YP_001410905 hypothetical protein Fnod_1402 *Fervidobacterium nodosum* Rt17-B1				2			9.467	*Fervidobacterium*	*Fervidobacterium* spp.

a) ORF is number in the genome sequence, i.e. 2120 corresponds to Maripi_2120.

b) Length in bp.

c) Annotation in genome.

d) Best BlastP match.

e) Weak HMM library classifications, with an E-value greater than 0.0001.

f) CD accession.

g) Coiled Coils probablility.

h) * indicates that no phylogenetic analysis could be done as there were no matches in Genbank. ** indicates that no phylogenetic analysis could be done as there were to few hits or the sequences were too divergent to align confidently.