Living Side by Side with a Virus: Characterization of Two Novel Plasmids from *Thermococcus prieurii*, a Host for the Spindle-Shaped Virus TPV1

Aurore Gorlas, Mart Krupovic, Patrick Forterre, Claire Geslin

Laboratory of Microbiology of Extreme Environments (LMEE), UMR 6197/CNRS/UBO IUEM, Plouzane, France; Unité Biologie Moléculaire du Gène chez les Extrémophiles, Département de Microbiologie, Institut Pasteur, Paris, France; Laboratoire de Biologie Moléculaire du Gène chez les Extrémophiles, Institut de Génétique et Microbiologie, CNRS-UMR 8621, Orsay, France

Microbial cells often serve as an evolutionary battlefield for different types of mobile genetic elements, such as viruses and plasmids. Here, we describe the isolation and characterization of two new archaeal plasmids which share the host with the spindle-shaped *Thermococcus prieurii* virus 1 (TPV1). The two plasmids, pTP1 and pTP2, were isolated from the hyperthermophilic archaean *Thermococcus prieurii* (phylum *Euryarchaeota*), a resident of a deep-sea hydrothermal vent located at the East Pacific Rise at 2,700-m depth (7°25′24″ S, 107°47′66″ W). pTP1 (3.1 kb) and pTP2 (2.0 kb) are among the smallest known plasmids of hyperthermophilic archaea, and both are predicted to replicate via the rolling-circle mechanism. The two plasmids and the virus TPV1 do not have a single gene in common and stably propagate in infected cells without any apparent antagonistic effect on each other. The compatibility of the three genetic elements and the high copy number of pTP1 and pTP2 plasmids (50 copies/cell) might be useful for developing new genetic tools for studying hyperthermophilic euryarchaea and their viruses.
globales (24), and Thermococcales (25). However, viruses have been isolated only from members of the latter order (26, 27). We have previously observed a close genetic relationship between a group of Thermococcales plasmids (7, 28) and the spindle-shaped Pyrococcus abyssi virus 1 (PAV1) (26, 29), with more than half of the viral genome being composed of genes that have homologues in plasmids (7). Consequently, it was proposed that PAV1-like viruses emerged as a result of recombination between a plasmid and a virus, which donated genetic determinants for genome propagation and virion formation. Indeed, the second recently isolated virus infecting Thermococcales, Thermococcus prieurii virus 1 (TPV1) (27), shares with PAV1 only two genes encoding putative structural proteins, but it does not carry the ones common to PAV1 and the plasmids.

TPV1 was originally isolated from T. prieurii (27), an endarchaeon residing in deep-sea hydrothermal vents of the East Pacific Rise (30), but was subsequently shown to infect a number of other Thermococcales species (31). Like PAV1, TPV1 possesses a spindle-shaped virion and contains a circular double-stranded DNA genome (27). The virus is not lytic, and its production is stimulated by UV irradiation. Interestingly, besides the TPV1 genome being composed of genes that have homologues in plasmids (7), plasmids (7, 28) and the spindle-shaped virus 1 (PAV1) (26, 29), with more than half of the viral genome being composed of genes that have homologues in plasmids (7). Consequently, it was proposed that PAV1-like viruses emerged as a result of recombination between a plasmid and a virus, which donated genetic determinants for genome propagation and virion formation. Indeed, the second recently isolated virus infecting Thermococcales, Thermococcus prieurii virus 1 (TPV1) (27), shares with PAV1 only two genes encoding putative structural proteins, but it does not carry the ones common to PAV1 and the plasmids.

Two Novel Plasmids of Thermococcus prieurii

**Materials and Methods**

**Strains and growth conditions.** Thermococcus prieurii strain Bio-pl-040512 (JCM16307; Japan Collection of Microorganisms designation) was isolated from a hydrothermal chimney sample collected from the East Pacific Rise at 2,700-m depth at the Sarah Spring area (7°25′24″N, 107°47′30″W). Cells were cultured in Rrovat medium as previously described (26, 27), with minor modifications as detailed below.

The medium contained, per liter of distilled water, the following: 1 g NaH2Cit, 0.2 g MgCl2·6H2O, 0.1 g CaCl2·2H2O, 0.1 g KCl, 0.83 g CH3COONa·H2O, 20 g NaCl, 5 g yeast extract, 5 g tryptone, 3.45 g D-piperazine-N,N′-bis(2-ethanesulfonic acid) (PIPES), and 0.001 g resazurin. The pH was adjusted to 7 before autoclaving. After autoclaving, 5 ml of 6% (wt/vol) K2HPO4 solution and 5 ml of 6% (wt/vol) KH2PO4 solution, separately sterilized, were added. The medium was dispensed (50 ml) into 100-ml sterile vials and completed by adding 1% (wt/vol) previously sterilized elemental sulfur. Anaerobiosis was obtained by applying vacuum to the medium and saturating it with N2. Finally, a sterile solution of Na2S·9H2O (final concentration, 0.05% [wt/vol]) was added to reduce the medium. The medium, inoculated to a final concentration of 1%, was incubated at 85°C.

**DNA extraction and sequencing.** Total DNA from T. prieurii was prepared as previously described (26, 32). Covalently closed circular DNA was extracted from cells in exponential growth phase by the alkaline lysis method (26, 33). Plasmids designated here pTP1 and pTP2 were completely digested with restriction endonucleases HindIII and Smal, respectively, and all of the fragments obtained were cloned in the corresponding plasmid vector. Sequencing reactions were carried out with the BigDye Terminator kit (Applied Biosystems) and analyzed at the Plateforme Biogenouest (Roscoff, France; http://www.sb-roscoff.fr/plateformes-techniques/génomique-sbr.html) on an ABI Prism 3100 genetic analyzer. Each insert was sequenced from both ends using the M13 forward and M13 reverse primers. The sequences were trimmed and assembled using the SeqMan Lasergene 8.0 program (DNASTAR, Inc., Madison, WI) with both strands completely sequenced and with a 3-fold coverage.

**Sequence analysis and annotation.** Glimmer (34), GenMark (35), and RBS finder (36) were used to identify open reading frames (ORFs). The amino acid sequence of each ORF was searched against the NCBI nonredundant protein database (February 2013) using PSI-BLAST (37) and compared to the database of known protein structures using HHpred (38). Membrane-spanning regions were predicted using the TMHMM program (39). The K/Kc ratio, namely, the ratio of the number of non-synonymous substitutions per nonsynonymous site (Ks) to the number of non-synonymous substitutions per synonymous site (Ks) (40), was estimated using a web-based Kc/Kc calculation tool at http://services.cbu.ubn.no/tools/kaks.

**Plasmid copy number determination.** Real-time PCR was used to determine the copy number of plasmid genomes in the host cell using an ABI Prism 7500 sequence detection system (Applied Biosystems). The method is based on the difference in threshold cycle values (ΔΔCt) of amplicons targeting the plasmids and an amplicon targeting the single-copy gene of 16S rRNA carried by the host chromosome and used as an internal standard. Total and plasmid DNAs were obtained as described previously (27). Primers were designed with Primer Express software, version 3.0. The following primer pairs were used for gene amplification: for the partial 16S rRNA gene, forward primer 5′-CCGTGGCCTTATTGGAATCA-3′ and reverse primer 5′-ACTTCCAGGCTGCCCTTTACA-3′; for pTP1 DNA, forward primer 5′-CCGGTACGGAATGCTGAT-3′ and reverse primer 5′-GCGGGGAAACACTAATGAC-3′; and for pTP2 DNA, forward primer 5′-CGAAAGATTGGCCAAAGAGGT-3′ and reverse primer 5′-TCACGGCGCAACAGAGCGA-3′. In all three cases, the fragments were identically sized.

Total DNA was mixed with 12.5 μl SYBR green mix for quantitative PCR (qPCR), 100 nM forward primer, and 100 nM reverse primer. Samples were incubated at 95°C for 10 min, followed by 60 cycles of amplification (at 95°C for 15′ and at 60°C for 1 min), and then cooled to 4°C. A standard curve was obtained using 10-fold serial dilutions of DNA, ranging from 1 to 10−6. Samples were tested in triplicate. The PCR products amplified had identical melting points (about 86°C) and were analyzed by agarose gel electrophoresis.

ΔΔCt used to calculate the relative plasmid copy number, is the difference in the mean Ct (threshold cycle) value of the amplicon targeting single-copy 16S rRNA genes and the mean C(t) value of the amplicon(s) targeting the plasmid DNAs.

**Nucleotide sequence accession numbers.** The complete sequences of pTP1 and pTP2 were deposited in GenBank under accession numbers KC617920 and KC617921, respectively.

**RESULTS**

Isolation and characterization of plasmids pTP1 and pTP2. Agarose gel electrophoresis of the extrachromosomal DNA extracted from TPV1-infected T. prieurii cells revealed that in addition to the viral genome, two smaller elements were present (27). The latter were isolated and digested with several type II restriction endonucleases (REases), including HindIII, EcoRI, EcoRV, and Smal. The analysis showed that the elements represent covalently closed circular DNA molecules of approximately 2 and 3 kb. The DNA was susceptible to digestion with both methylation-insensitive and methylation-sensitive (e.g., Smal) REases.

The complete nucleotide sequences of both DNA molecules were subsequently determined. The sizes were found to be consistent with those defined using restriction analysis: the larger element, designated pTP1 (for T. prieurii plasmid 1), consists of 3,126 bp, while the smaller one, pTP2, is 2,038 bp long. Six-frame in silico translation of pTP1 and pTP2 sequences suggested that each plasmid contains 5 open reading frames (ORFs) (Fig. 1). The G+C content of pTP1 and pTP2 was found to be 42.5 and 41.7%, respectively. These values are considerably lower than those of TPV1 (49.9%) and the T. prieurii chromosome (53.6%) (27) but...
are similar to the G+C content of other small plasmids of *Thermococcales*, namely, pTN1 (3.6 kb; 45.5% G+C) from *T. nautilus* (41), pGT5 (3.4 kb; 43.5% G+C) from *Pyrococcus abyssi* (42), and pRT1 (3.4 kb; 43% G+C) from *Pyrococcus* sp. strain JT1 (43). pTN1 and pGT5 encode homologous replication proteins (Rep)s and replicate via a rolling-circle (RC) mechanism (41, 42). The predicted Rep protein of pRT1, on the other hand, does not display the canonical order of RC-Rep motifs (44) and the protein is not related to the RC-Rep of pTN1 and pGT5 (43, 45). The mechanism of pRT1 replication thus remains unresolved.

**pTP1 is a unique recombinant variant related to the pTN1-like rolling-circle plasmids.** Three of the five predicted pTP1 ORFs (Table 1) display sequence similarity to the RC-Rep, Rep74, from plasmid pTN1 (Fig. 2A), but only one of them is predicted to encode a full-length Rep74-like protein, which we designate RepTP1. All three conserved motifs typical of RC-Rep proteins of the superfamily II (SFII; with one catalytic Tyr residue in motif 3) (44) are readily identifiable in RepTP1 (Fig. 2B). The two other ORFs, ORF1 and ORF3, match the central and 3′-distal regions, respectively, of the pTN1 rep74 (Fig. 2A and Table 1). The similarity between ORF3 and rep74 extends throughout the region encompassing the three signature motifs. However, there is a deletion of motif 1 in ORF3, while the catalytic Tyr in motif 3 is changed to a Ser residue. Interestingly, ORF1 and ORF3 are separated by a repTP1 gene and are in the opposite orientation with respect to repTP1. These observations suggest that ORF1 and ORF3 once constituted a single ORF encoding an RC-Rep of the ancestral pTP1. The gene was subsequently split in the course of recombination between two ancestral pTP1 plasmids, leading to integration of a new RC-Rep gene copy into the preexisting one (Fig. 2A). Notably, the inverse and nested orientation of the repTP1 gene suggests that the incoming rep was provided in trans rather than emerging in the course of duplication of the original gene copy.

The scenario described above predicts that the incoming and the resident RC-Rep gene copies were identical. To verify this hypothesis, we compared the sequences of RepTP1 and ORF3. The two turned out to be identical at the protein level over the 142-amino-acid (aa) region (note that identity abruptly stops at the catalytic Tyr residue in motif 3; Fig. 2C). This observation argues against the possibility that the two Rep genes (fragmented and full length) are derived from heterologous plasmids. Strikingly, however, when the corresponding nucleotide sequences were compared, we found that there is only 75% identity (322/428) between them (Fig. 2C). Closer examination of the pairwise nucleotide sequence alignment revealed that all mutations are silent, with the vast majority of them occurring at the third position of the codons; 81 out of 142 codons had mutations in the third position. In addition, 11 codons (6 for Ser and 5 for Leu) also displayed changes in other positions (Fig. 2C). This observation is consistent with the fact that Ser and Leu represent two of the three amino acids (the third one is Arg) for which the genetic code is most degenerated: each can be encoded by six different codons. Since ORF3 (and ORF1) apparently represents a fragment of a disrupted Rep gene, we expected that it will be evolving under neutral (in the case of the inactivated gene) or positive (in the case of an interfering gene product) selection. Indeed, the result presented above clearly indicates that ORF3 evolves under strong purifying selection. Indeed, the *Ks/Ka* ratio for the ORF3-RepTP1 gene pair was estimated to be 0.036 (*Ks/Ka* ratios below 1 indicate purifying selection [40]). Thus, despite the lack of motifs 1 and 3, the product of ORF3 is likely to be a functional player in pTP1 propagation, and biochemical studies are needed to reveal its exact role(s). As mentioned above, ORF1 shares sequence similarity with the 3′-distal region of rep74 from pTN1 but not with the equivalent region of repTP1. Consistent with this, the C-terminal regions of RepTP1 and Rep74 are not homologous. Thus, it appears that the original Rep gene of pTP1, which was subsequently disrupted by introduction of repTP1, was more similar to rep74 of pTN1 than the current repTP1 gene is.

**pTP2 is unrelated to other plasmids of Thermococcales.** With 2,038 bp, pTP2 is the smallest known plasmid of hyperthermophilic archaea. Sequence analysis revealed that it is not related to pTP1 or, in fact, to any other plasmid of *Thermococcales*; not a single gene is shared between pTP2 and the other plasmids. Nevertheless, similar to pTP1, pTP2 is likely to replicate using the RC mechanism. ORF3 of pTP2 encodes a putative protein (RepTP2) of 252 aa which shares significant sequence similarity with RC-Rep encoded by diverse archaeal MGE (Table 2), including *Archaeoglobus profundus* plasmid pG5 (24), putative *Methanococcus voltae* A3 provirus (MVV) (46), and a group of pleomorphic haloarchaeal viruses (47). Sequence analysis of RepTP2 showed that, unlike RepTP1, it possesses two conserved Tyr residues.

### Table 1 Annotation of plasmid pTP1

<table>
<thead>
<tr>
<th>ORF</th>
<th>Position (strand)</th>
<th>Start/stop codon</th>
<th>Length (aa)</th>
<th>Function/feature</th>
<th>Homologue (GenBank accession no.)</th>
<th>Identity (%)</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5...178 (−)</td>
<td>CTT/TGA</td>
<td>58</td>
<td>RC-Rep fragment</td>
<td><em>Thermococcus nautilus</em> plasmid pTN1, Rep74 (YP_001351689)</td>
<td>31/37 (41%); 3e−03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(repTP1)</td>
<td>158...1975 (+)</td>
<td>606</td>
<td>RC-Rep</td>
<td><em>Thermococcus nautilus</em> plasmid pTN1, Rep74 (YP_001351689)</td>
<td>172/488 (35%); 4e−81</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1896...2468 (−)</td>
<td>ATG/TAA</td>
<td>191</td>
<td>RC-Rep fragment</td>
<td><em>Thermococcus nautilus</em> plasmid pTN1, Rep74 (YP_001351689)</td>
<td>64/179 (36%); 5e−22</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2646...3008 (−)</td>
<td>GTG/TAG</td>
<td>121</td>
<td>TMD* (2×)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2977...3400 (−)</td>
<td>GTG/TAA</td>
<td>38</td>
<td>TMD† (1×)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* TMD, transmembrane domain. The numbers in parentheses denote the number of predicted TMDs.
dues in motif 3 (Fig. 3) and should be considered a member of SFI of RC-Reps. RepTP2 is considerably shorter than its homologues in other archaeal MGE (252 aa versus 430 to 700 aa; Fig. 3). The size difference is mainly due to the lack of an N-terminal domain (H11011 200 aa) in RepTP2. Notably, the N-terminal extensions in other RepTP2-like proteins are very divergent, to the point that N-terminal regions of Reps from haloarchaeal MGE do not share identifiable similarity with the corresponding sequences from methanococcal or methanosarcinal RC-Reps. We hypothesize that an ORF preceding and overlapping the repTP2 gene encodes a functional counterpart of the N-terminal domains observed in other RepTP2-like proteins. However, the validity of this prediction remains to be verified by biochemical studies of the pTP2 proteins.

Besides repTP2, putative functions could be inferred for two other pTP2 ORFs (Table 2). HHpred analysis suggests that ORF1 and ORF5 encode DNA-binding proteins that possibly are involved in transcription regulation. The product of ORF1 is predicted to possess a winged helix-turn-helix motif, while the one of ORF5 belongs to a family of SpoVT/AbrB-like transcriptional regulators displaying the swapped-hairpin fold (Table 2). Notably, an SpoVT/AbrB-like protein is also encoded by TPV1 (27); however, it does not share significant sequence similarity with the pTP2 protein.

pTP1 and pTP2 stably propagate in TPV1-infected cells. We have previously shown that in infected T. prieurii cells, the TPV1 genome is present at 20 copies per chromosome (27). Interestingly, qPCR analysis showed that pTP1 and pTP2 are considerably more abundant than TPV1 genome in infected cells; both plasmids were found to be present in 50 copies per chromosome. This
number is similar to those reported for numerous bacterial RC plasmids (48) but is much higher than the copy number estimated for the pTN1-based Escherichia coli-T. kodakaraensis shuttle vector pLC70, which was found to be present in just 3 copies per chromosome (49).

To gain insights into the possible interaction between the two plasmids and TPV1, T. prieurii, with a doubling time of 23 min, was subjected to serial passages in liquid cultures (8 times, corresponding to approximately 187 generations), and the resultant culture was streaked on a plate and examined for the presence of the three extrachromosomal elements. Both plasmids and the virus TPV1 were stably maintained in T. prieurii cells during the course of the experiment, with all colony clones analyzed testing positive for the presence of all three MGE. However, unlike the case of crenarchaeal fuselloviruses, which can package and transposition was catalyzed by RepTP1 itself. Another unexpected finding has been noticed previously (41). Thus, it is possible that, akin to the homologous RC-Rep gene. What the mechanism of such integration could have been is currently unclear. Remarkably, a relationship between RepTP1-like RC-Reps and bacterial transposases has been noticed previously (41). Thus, it is possible that, akin to the transposition reaction of rolling-circle transposons, the integration was catalyzed by RepTP1 itself. Another unexpected finding was that ORF3, the larger fragment of the disrupted RC-Rep of pTP1, is evolving under strong purifying selection (Fig. 2A and C), consistent with this, neither pTP1 nor pTP2 was observed in DNA preparations extracted directly from TPV1 viral particles (27).

**DISCUSSION**

In this study, we have characterized two cryptic plasmids of T. prieurii. Our results indicate that the two plasmids and the spindle-shaped virus TPV1 stably propagate within T. prieurii cells. Such stable coexistence is reminiscent of that described for certain pRN-like plasmids and fuselloviruses of Sulfolobus (19). However, comparative genome analysis did not reveal a single gene that would be shared between either of the two plasmids and the virus. This is in contrast to the extensive gene content overlap between the medium-sized pTN2-like plasmids of Thermococcales and PAV1 (7), the only other virus known to infect Thermococcales (29, 31); however, none of these PAV1 plasmid genes was found in TPV1. Nevertheless, MGE other than pTN2-like plasmids appear to have contributed to shaping of the TPV1 genome (27).

Similar to many other medium-sized euryarchaeal viruses and plasmids (50), TPV1 encodes a replicative minichromosome maintenance (MCM) helicase and is likely to replicate via a theta mechanism (27). pTP1 and pTP2, on the other hand, apparently utilize the rolling-circle mode of replication. The two plasmids encode proteins belonging to two distinct superfamilies (SF) of RC-Reps. Whereas SFII Reps were previously found in plasmids of Thermococcales (41, 42), MGE encoding SFII RC-Reps have not been observed in this euryarchaeal order. Our analysis has uncovered a unique recombination event which occurred in the evolution of pTP1. As a result of this recombination, an ancestral copy of the RC-Rep gene has been disrupted by the incoming copy of a homologous RC-Rep gene. What the mechanism of such integration could have been is currently unclear. Remarkably, a relationship between RepTP1-like RC-Reps and bacterial transposases has been noticed previously (41). Thus, it is possible that, akin to the transposition reaction of rolling-circle transposons, the integration was catalyzed by RepTP1 itself. Another unexpected finding was that ORF3, the larger fragment of the disrupted RC-Rep of pTP1, is evolving under strong purifying selection (Fig. 2A and C), indicating that its product retained some function in plasmid reproduction. Biochemical studies will be crucial for rationalizing these perplexing observations.

Although constantly growing, the genetic toolkit for Thermococcales is still limited. Especially modest is the choice of replicative expression vectors (51, 52). Indeed, all of them are currently
ACKNOWLEDGMENTS

This work was supported by the ANR project Thermovésicules (ANR-12-BSV3-0023-01) (to P.F. and C.G.), a MENRT grant (to A.G.), and the Foundation for Research on Biodiversity (DIVIVIR 2009–2012).

REFERENCES


