Provirus Induction in Hyperthermophilic Archaea: Characterization of *Aeropyrum pernix* Spindle-Shaped Virus 1 and *Aeropyrum pernix* Ovoid Virus 1

Tomohiro Mochizuki,1 Yoshihiko Sako,2 and David Prangishvili1,*

Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Department of Microbiology, Institut Pasteur, 75015 Paris, France, and Laboratory of Marine Microbiology, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan

Received 15 April 2011/Accepted 28 June 2011

By *in silico* analysis, we have identified two putative proviruses in the genome of the hyperthermophilic archaeon *Aeropyrum pernix*, and under special conditions of *A. pernix* growth, we were able to induce their replication. Both viruses were isolated and characterized. Negatively stained virions of one virus appeared as pleomorphic spindle-shaped particles, 180 to 210 nm by 40 to 55 nm, with tails of heterogeneous lengths in the range of 0 to 300 nm. This virus was named *Aeropyrum pernix* spindle-shaped virus 1 (APSV1). Negatively stained virions of the other virus appeared as slightly irregular oval particles with one pointed end, while in cryo-electron micrographs, the virions had a regular oval shape and uniform size (70 by 55 nm). The virus was named *Aeropyrum pernix* ovoid virus 1 (APOV1). Both viruses have circular, double-stranded DNA genomes of 38,049 bp for APSV1 and 13,769 bp for APOV1. Similarities to proteins of other archaeal viruses were limited to the integrase and Dna1-like protein. We propose to classify APOV1 into the family Guttaviridae.

Sequencing of diverse bacterial and archaeal genomes has revealed the presence of integrated putative viral genomes, their fragments, and viral genes in cellular chromosomes (3, 8, 14–16, 28, 32). These results served as a basis for speculations on virus-host relationships and the nature of presumed proviruses. For example, *in silico* analysis allowed us to postulate infection of methanogenic euryarchaea by viruses related to bacterial or crenarchaeal icosahedral viruses (11) or head-and-tail viruses from the order *Caudovirales* (12). Traces of integration of viral genomes of members of the families *Fuselloviridae* and *Bicaudaviridae* were detected in the chromosomes of hyperthermophilic crenarchaea (8, 23, 25, 31, 32). However, until now, there have been no reports on the induction and isolation of putative integrated archaeal viruses indentified *in silico*. Moreover, it is unclear whether replication of all these putative proviruses can be induced at all.

To make matters more difficult, discovery of proviruses by *in silico* analysis is not straightforward and is often overlooked in primary genome annotations. For screening genome sequences for putative proviruses, conserved viral functions/genes are often used as markers and the surrounding regions are analyzed for their gene content. However, this approach implies that the provirus, in addition to the marker gene, carries other recognizable genes of viral origin.

In the case of archaeal genomes, integrase genes are often used as markers. On the basis of their sequences and functions, archaeal integrases can be classified into three major categories, the Xer type, the pNOB8 type, and the *Sulfolobus* spindle-shaped virus (SSV) type. Enzymes of the Xer type have bacterial homologues and were recently shown to be involved in resolving chromosome dimers during the cell cycles of *Pyrococcus* and *Sulfobolus* (4, 5). The other two enzymes are tyrosine recombinases, exclusive to archaea. Both are widely distributed in archaeal integrative extrachromosomal elements, such as viruses or plasmids, and both target tRNA sequences on cellular chromosomes. The two types are distinguished by the location of the target sequence. For enzymes of the pNOB8 type, the tRNA sequence acting as the attB and attP sites are located adjacent to the integrase gene, whereas for the enzymes of the SSV type, these sequences are located within the integrase gene itself. Consequently, after integration into the host chromosome, the integrase gene of the SSV type is split, and its two fragments are found at the boundaries of the integrated element, defined as intC and intN. All integrase genes from known crenarchaeal viruses are of the SSV type. This type of gene is also found on some archaeal plasmids (21, 34). The pNOB8-type integrase was found to be encoded on archaeal plasmids (6) and in putative proviruses of the *Euryarchaeota* (11).

Here we report the first example of induction of *in silico*-identified archaeal proviruses and describe two new viruses of *Aeropyrum pernix*.

**MATERIALS AND METHODS**

Search for a putative proviral sequence. Putative proviral sequences in the whole-genome sequence of *A. pernix* K1 were identified by searching for integrase and tRNA genes. The search for integrase genes was performed by BLASTp and tBLASTx analyses, using integrase gene sequences from archaea and their viruses as queries. The putative proviral region was determined as a region between two tRNA-related sequences flanking an integrase gene.

**Strain and medium.** *A. pernix* strain K1 was obtained from the DSMZ culture collection (DSM accession no. 11879). For cell growth, the following media were used: (i) JXTm (20); (ii) IST, consisting of 35 g of sea salts (Sigma) containing 1 g Na2S·9H2O·5H2O, 1 g yeast extract (Difco), and 1 g tryptone (Difco) per liter (18); and (iii) MBT, consisting of marine broth 2216 (Difco) containing addi-
MgCl₂, in a total volume of 500 μl. A brief, a virus suspension in 50 mM Tris-HCl (pH 7.5) buffer containing 10 mM sodium chloride.

Micrograph (Gatan) version 1.83.842. Recorded under low-dose conditions on a Gatan Ultrascan 4000 with Digital-2010F electron microscope (JEOL, Japan) operating at 200 kV. Images were transferred to a Gatan 626 DH cryoholder (Gatan) and examined on a JEOL Tools GmbH, Germany) and cryofixed in liquid ethane. The specimens were concentrated to 200 μl. Four microliters of the PCR product from the first amplification. Cell growth was monitored by transmission electron microscopy (TEM) after their concentration by ultracentrifugation at 48,000 rpm in 4-ml tubes using an SW60 rotor (Beckman) for 12 h or by polyethylene glycol (PEG) precipitation by addition of NaCl and PEG 6000 to final concentrations of 1 M and 10%, respectively; after incubation at 4°C overnight, viral particles were collected by centrifugation at 12,000 × g for 30 min at 4°C, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

DNA isolation. DNA was extracted from virions as described earlier (30). In brief, a virus suspension in 50 mM Tris-HCl (pH 7.5) buffer containing 10 mM sodium chloride.

For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.

Electron microscopy. For negative staining, samples were applied to glow-discharged carbon-coated copper grids, stained with 2% uranyl acetate for 10 s, and the particles were suspended in 20 mM Tris-acetate, pH 7.0, containing 3% sodium chloride.
suboptimal conditions. The virus was named *Aeropyrum pernix* spindle-shaped virus 1 (APSV1).

**Morphology of APSV1.** Virions of APSV1 are pleomorphic spindle-shaped particles, ~180 to 210 nm in length and ~40 to 55 nm in width (Fig. 2a). Part of the virions in the purified preparation carried a single tail or two tails protruding from pointed ends (Fig. 2b to d). Among the 444 analyzed particles, the tailed particles comprised about 25%. The majority of these tailed particles, about 67%, carried a single tail ranging in length from about 20 to 150 nm; a smaller proportion, 33%, was two tailed, with the total particle length reaching about 700 nm. Three short filaments were attached to one end of the particles, tailless as well as single tailed (Fig. 2c, inset). The virions have a tendency to attach to each other with their filament-carrying ends, sometimes forming rosette-like structures (Fig. 2e). The incubation of a purified viral preparation at different temperatures in the range of 75 to 90°C up to 7 days did not affect the ratios of the tailless, single-tailed, and two-tailed virions, either in Tris-HCl buffer at pH 7 or in the culture medium (3ST).

**Determination of the APSV1 genome sequence.** Nucleic acid extracted from the purified APSV1 virions was digested with several type II restriction enzymes, indicating that it was double-stranded DNA (dsDNA) (Fig. 3a). The pattern of digestion fragments showed stoichiometric distribution, suggesting that DNA originated from a single viral strain (Fig. 3a). Based on the lengths of the digestion fragments, the size of the viral DNA was estimated to be about 40 kb.

In order to find out whether viral DNA corresponds to any of the putative proviruses of *A. pernix*, the sizes of fragments obtained by digestion of the viral DNA with BamHI, EcoRI, SalI, and XhoI (Fig. 3a) were compared to the calculated sizes of fragments of analyzed regions of the genome, shown in Fig. 1 (Fig. 3b). One of these regions, about 110 kb long, extended from about 55 kb upstream of APE_0818 until about 55 kb downstream of this gene. By comparing the lengths of theoretical and actual DNA fragments, the analyzed sequence could be trimmed down to about 40 kb, exactly to the segment corresponding to the putative “beta” provirus.

To verify that the sequence of the APSV1 genome is identical to the putative delta provirus of *A. pernix*, PCR analysis was performed. The primers V1F and V1R were designed based on the sequences bordering the beta provirus, facing outwards (Fig. 1). Using APSV1 DNA as a template, this

![FIG. 1. Scheme of the fragment of the *A. pernix* K1 genome containing three putative proviruses. Integrase genes (large black arrows), tRNA genes, and truncated tRNAs (gray) are indicated. The sites and orientations of the primer pairs which were used for sequence analysis are indicated with small arrows beneath the diagrams.](image1)

![FIG. 2. Transmission electron micrographs of APSV1 virions. Scale bars = 100 nm. Samples were negatively stained with 2% uranyl acetate.](image2)
primer pair produced a DNA fragment of about 700 bp, apparently from the excised and circularized genome fragment. The sequence of the first 547 bp was identical to an *A. pernix* sequence fragment from positions 580187 to 580733. The sequence of the last 61 bp was identical to an *A. pernix* sequence fragment from positions 542635 to 542696 and had an overlap of 50 bp with the first part (see Fig. S2 in supplemental material). The results confirmed that APSV1 has a circular genome derived from the excision and circularization of the provirus beta region of the *A. pernix* genome (Fig. 1). They also indicated that the viral genome contains only 50 bp from the tRNA 21-Val gene as a part of the integrase gene (Fig. 4). Moreover, based on these results and on the available genome sequence of *A. pernix* (9), the sequence of the APSV1 genome could be asserted.

Analysis of the APSV1 genome sequence. The genome of APSV1 is a circular dsDNA containing 38,049 bp. The GC content (56.51%) is in the range of that of the *A. pernix* chromosome (56.3%) but higher than that of the chromosomal regions bordering the provirus. The ORF of 1,566 bp coding for the integrase is split as a result of integration into the host chromosome in two parts. The first 118 bp, including the start codon, is present at the right border of the proviral region, and the last 1,498 bp, including the termination codon at the left border, has an overlap of 50 bp, as illustrated in Fig. 1.

In the genome, 53 ORFs encoding hypothetical proteins with 38 to 929 amino acids, which start from the ATG, GTG, and TTG codons, could be identified. In most cases, they were preceded by recognizable putative ribosome-binding sites (see Table S1 in the supplemental material). The majority of putative genes were present on one DNA strand, and only six putative genes were on the other. The genome map of APSV1 is illustrated in Fig. 4. The numbering of the nucleotides in the genome sequence starts from the start codon of the first ORF following the integrase gene. The ORFs are numbered correspondingly, and the number following the sequential number refers to the number of amino acids in the predicted protein.

Of the predicted ORFs, nine showed significant (E value < 0.001) sequence similarity with sequences in the public databases, and two of them had homologues in the genomes of other archael viruses (Table 2). Of note, the integrase gene of APSV1, ORF53-521, showed the highest sequence similarity to integrase genes in the chromosomes of *Sulfolobales* (i.e., *Sulfolobus islandicus*).

The sequences of three consecutive genes, ORF38-469, ORF39-350, and ORF40-223, showed exceptionally high sequence similarity with the sequences in the public database (Table 2, and see Table S2 in supplemental material). Sequence analysis allowed us to predict that ORF38-470 is a methylase and that ORF40-223 is a DNA glycosylase, a member of a family of enzymes involved in DNA mismatch repair. The function of ORF39-350 could not be predicted by homology searches. These three genes are present in the same orientation as that of blocks in the genomes of two other hyperthermophilic members of the *Crenarchaeota*, *Pyrobaculum aerophilum* and *Thermoproteus neutrophilus*. In each case, the high similarity can be detected at both the amino acid and nucleotide sequence levels. ORF38-469 and ORF40-223 have homologues, although with lower sequence similarity, in genomes of different members of the three domains. ORF38-469 has homologues also in various bacterial and eukaryal viruses and one euryarchaeal virus.

The search for the origin of replication was performed by using Z-curve and by locating inverted-repeat sequences. The Z component of the Z-curve shows four characteristic peaks in the APSV1 genome, about 2 kb, 6 kb, 25 kb, and 34 kb away from nucleotide position 1. Inverted repeats were searched, and a region containing a perfect repeat of 18 bp, CTATCACCTATACCTAG(N)_{45}CTAGAGTATAGGTGATAG, was found, starting at nucleotide position 34121 and ending at position 34202. This region between ORF44-94 (negative strand) and ORF45-81 (positive strand) is a noncoding region and is likely to represent the origin of replication of the viral genome (Fig. 4). No homologue of ORF44-94 was found in
databases, and ORF45-81 showed low sequence similarity to the transcriptional regulator of *Brevibacterium* (E-value, 0.008).

**Morphology of APOV1.** After continuous culturing of *A. pernix* in 3ST or MBT medium without extensive aeration, we occasionally detected in the culture the presence of virus-like particles, different in morphology from virions of APSV1 (Fig. 5a). Their titer was comparable to that of APSV1 and was estimated to be about 10^7 particles/ml. These particles were oval and slightly pleomorphic and measured about 80 by 60 nm (Fig. 5b). They were often found in clusters and sometimes showed a concave morphology resembling that of beans (Fig. 5b, inset). No surface structures or tails could be observed. Cryo-electron microscopy analysis enabled better understanding of particle morphology. Particles that were embedded in vitreous ice were of regular ovoid shape and size, measuring 70 by 55 nm (Fig. 5c). Globular subunits about 3.5 nm in width could be distinguished as building blocks of the surface layer. The structure was clearly different from that of APSV1 virions, and we presumed that the particles represented virions of another induced virus of *A. pernix*, which we termed *Aeropyrum pernix* ovoid virus 1 (APOV1).

**Determination of the APOV1 genome sequence.** Unfortunately, we were unable to control the replication of the viruses and could not obtain a large-scale culture producing solely the virions of APOV1. In attempts to identify the origin of APOV1’s genome, it was important to discover an *A. pernix* strain, YKP1, which was producing only one type of virus, referred to here as *Aeropyrum pernix* ovoid virus 2 (APOV2), highly similar in morphology to virions of APOV1 (see Fig. S3 in the supplemental material). Virions of APOV2 were collected, and DNA was extracted from them. It could be digested with several types of type II restriction enzymes, confirming that it was dsDNA (data not shown). The sizes of DNA fragments did not clearly match the calculated sizes of fragments from the two potential proviruses shown in Fig. 1. Applying the SISPA method with random primers, we were able to PCR amplify 6 segments of the DNA, with the lengths ranging from 310 to 900 bp. Five of them revealed over 93% sequence similarity to different regions of the *A. pernix* K1 genome, but all were within positions 488171 to 491258 (Fig. 1). The results strongly suggested that the genome sequence of APOV2 is highly similar to the alpha provirus region (Fig. 1). This, in its turn, suggested that APOV1 represents an induced alpha provirus of *A. pernix* K1.

In order to verify that the sequence of APOV1 DNA is identical to that of the alpha provirus, PCR analysis was performed. The primers, V2F and V2R, were designed based on the sequences bordering the putative proviral region, facing outwards (Fig. 1). When APOV1 DNA served as a template, the PCR product of about 400 bp was produced. The sequence of the first 250 bp was identical to the alpha provirus sequence from positions 491733 to 491982, and the sequence of the last 189 bp was identical to the alpha provirus sequence from positions 478149 to 478337 (see Fig. S4 in the supplemental material). The sequences had identical sequences of 65 bp at two ends the of alpha provirus region (Fig. S4). The results confirmed the origin of APOV1 from the alpha

**TABLE 2. Significant database matches to APSV1 ORFs**

<table>
<thead>
<tr>
<th>ORF</th>
<th>Best match</th>
<th>Organism</th>
<th>No. of amino acids aligned</th>
<th>No. of identical/no. of positive amino acids</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-235</td>
<td>Hypothetical protein</td>
<td><em>Sulfolobus tokodaii</em></td>
<td>201</td>
<td>69/109</td>
<td>3e−19</td>
</tr>
<tr>
<td>12-290</td>
<td>DnaA-like protein</td>
<td>ASV1</td>
<td>210</td>
<td>64/113</td>
<td>9e−16</td>
</tr>
<tr>
<td>26-534</td>
<td>Hypothetical protein</td>
<td>ASV1</td>
<td>235</td>
<td>68/99</td>
<td>1e−12</td>
</tr>
<tr>
<td>31-111</td>
<td>Hypothetical protein</td>
<td><em>Methanocaldococcus vulcanius</em></td>
<td>85</td>
<td>49/64</td>
<td>6e−16</td>
</tr>
<tr>
<td>38-469</td>
<td>DNA-cytosine methyltransferase</td>
<td><em>Thermoproteus neutrophilus</em></td>
<td>429</td>
<td>132/216</td>
<td>4e−175</td>
</tr>
<tr>
<td>39-350</td>
<td>Hypothetical protein</td>
<td><em>Thermoproteus neutrophilus</em></td>
<td>349</td>
<td>199/255</td>
<td>2e−111</td>
</tr>
<tr>
<td>40-223</td>
<td>U/G and T/G mismatch-specific glycosylase</td>
<td><em>Pyrobaculum aerophilum</em></td>
<td>211</td>
<td>127/169</td>
<td>6e−74</td>
</tr>
<tr>
<td>43-534</td>
<td>Signal recognition particle</td>
<td><em>Saccoglossus kowalevskii</em></td>
<td>198</td>
<td>45/90</td>
<td>3e−04</td>
</tr>
<tr>
<td>53-521</td>
<td>Integrase family protein</td>
<td><em>Sulfolobus islandicus</em></td>
<td>429</td>
<td>132/216</td>
<td>4e−39</td>
</tr>
</tbody>
</table>

*FIG. 5.* Transmission electron micrographs of virions of APOV1 (a to c) with APSV1 (a). (a) Black arrowheads indicate APOV1, and white arrowheads indicate APSV1. (a and b) Samples were negatively stained with 2% uranyl acetate. (c) Sample embedded in vitreous ice. Scale bars, 1,000 nm (a), 500 nm and 100 nm in the insets (b), and 100 nm (c).
provirus region of the *A. pernix* chromosome by its excision and circularization, which also indicated that the circular viral genome contains only 65 bp from the tRNA20-Leu gene as a part of the integrase gene (Fig. 6). Based on these results and on the available genome sequence of *A. pernix* (9), the sequence of the APOV1 genome could be asserted.

**Genome of APOV1.** The genome of APOV1 is circular dsDNA containing 13,769 bp. The GC content (56.51%) was in the range of that of the whole host chromosome (56.3%). The integrase gene (ORF21-400) was split as a result of integration into the host chromosome in two parts; the first 298 bp, including the start codon, was located at the right border of the proviral region, and the last 967 bp, including the stop codon, was located on the left border, with an overlap of 65 bp.

Twenty-one ORFs starting from ATG, GTG, and TTG codons for hypothetical proteins with 56 to 678 amino acids could be identified on the genome. In most cases, they were preceded by recognizable putative ribosome-binding sites (see Table S3 in the supplemental material). The genome map of APOV1 is illustrated in Fig. 6. The numbering of the nucleotides in the genome sequence starts from the stop codon of the first ORF following the integrase. The majority of putative genes were located on one DNA strand; only seven were on the other. Five of these were clustered together, from ORF17-201 to ORF21-399 (Fig. 6).

Of the 21 predicted ORFs, 5 showed sequence similarity to sequences in the public database (Table 3). In two cases, the similarity was with genes of members of the *Fuselloviridae* encoding an integrase and a DnaA-like protein. In other cases, the similarities were with putative genes of hyperthermophilic crenarchaeotes.

We were unable to determine the origin of replication of APOV1 by using Z-curve and locating inverted-repeat sequences, the approach which was successful in the case of APSV1 (see above) and *Aeropyrum pernix* bacilliform virus 1 (APBV1) (18).

**DISCUSSION**

In our attempts to characterize viruses of the genus *Aeropyrum*, we searched for putative proviruses in the genome of the type strain of the genus *A. pernix* K1, having in mind to induce their replication. Earlier in silico analysis reported the presence of two integrase genes and a possible integrative element, APE1, in the genome of *A. pernix* (31, 32). In this study, we reanalyzed the *A. pernix* genome and revealed two putative proviruses, named alpha and beta (Fig. 1), bearing features typical of archaeal integrated elements and having an integrase gene flanked by tRNA-related sequences. Although the sizes do not exactly correspond to those of the putative integrative elements reported earlier, we presume that the alpha provirus corresponds to the integrative element APE1 (31).

Traditionally, to induce integrated viruses, cells are treated with stress-causing factors, such as UV irradiation or mitomycin C (17). Due to the specific growth conditions of *A. pernix* (90°C and difficulties in culturing on solid media), we avoided using such methods and attempted to induce virus replication by modifying growth conditions. Members of the genus *Aeropyrum* are known to be obligate aerobes (29). Therefore, it has been recommended that they be grown under extensive aeration, e.g., by vigorous shaking of cultures (9, 36). We hypothesized that for such an organism, a decrease in oxygen accessibility can be a stress factor. Indeed, limited oxygen supply in the course of culturing without shaking caused induction of provirus replication.

Two virus species with different morphotypes and genomes were found to be produced by the “stressed” cells of *A. pernix*. Their titer in the growth cultures reached about 10^7 particles/ml. One virus, named *Aeropyrum pernix* spindle-shaped virus 1

![FIG. 6. Genome map of APOV1. The integrase gene is white, and the attP site is indicated.](image)

<table>
<thead>
<tr>
<th>ORF</th>
<th>Best match</th>
<th>Organism</th>
<th>No. of amino acids aligned</th>
<th>No. of identical/no. of similar amino acids</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-81</td>
<td>CRISPR Csa2 family protein</td>
<td><em>Staphylothermus marinus</em></td>
<td>70</td>
<td>32/42</td>
<td>1e−05</td>
</tr>
<tr>
<td>10-284</td>
<td>Hypothetical Ape_0536.1</td>
<td><em>Aeropyrum pernix</em> K1</td>
<td>267</td>
<td>149/181</td>
<td>1e−73</td>
</tr>
<tr>
<td></td>
<td>Endo-alpha-mannosidase</td>
<td><em>Chryseobacterium gleum</em></td>
<td>258</td>
<td>59/112</td>
<td>7e−07</td>
</tr>
<tr>
<td>14-121</td>
<td>Hypothetical protein</td>
<td><em>Pyrobaculum aerophilum</em></td>
<td>120</td>
<td>53/67</td>
<td>6e−13</td>
</tr>
<tr>
<td>19-247</td>
<td>Hypothetical protein</td>
<td><em>SSV1</em></td>
<td>194</td>
<td>59/93</td>
<td>1e−08</td>
</tr>
<tr>
<td></td>
<td>DnaA-like protein</td>
<td><em>SSV7</em></td>
<td>134</td>
<td>36/58</td>
<td>0.003</td>
</tr>
<tr>
<td>21-399</td>
<td>Hypothetical protein</td>
<td><em>Archaeoglobus profundus</em></td>
<td>179</td>
<td>62/94</td>
<td>4e−20</td>
</tr>
<tr>
<td></td>
<td>SSV1 integrase homologue</td>
<td><em>Thermococcus kodakarenseis</em></td>
<td>263</td>
<td>78/123</td>
<td>1e−17</td>
</tr>
</tbody>
</table>
Fuselloviridae (APSV1), had a genome that originated from the beta provirus region of *A. pernix*, and the genome of the other, named *Aeropyrum pernix* ovoid virus 1 (APOV1), originated from the alpha provirus region. The possibility that we might observe stochastic variation in spontaneous induction could be excluded. First, we could never observe the presence of virus-like particles in cell cultures growing under standard conditions. Second, no extrachromosomal elements were detected in cells of *A. pernix* K1, which were grown under standard conditions (9). Third, no virion proteins were present in the proteome of *A. pernix* K1 grown under standard conditions (36). These data unambiguously indicate that we observed virus induction from the zero level to a titer of about 10^7 particles/ml. The results provide the first example of induction of *in silico* identified putative integrated viruses in the *Archaea*. Unfortunately, factors which could selectively trigger the induction of APSV1 or APOV1 could not be determined.

*Aeropyrum pernix* spindle-shaped virus 1. The spindle-shaped morphotype of APSV1 is common for archaeal viruses infecting diverse hyperthermophilic and hyperhalophilic hosts, including members of the genera *Sulfolobus*, *Acidianus*, *Pyrococcus*, and *Haloarcula* (22, 23), which are abundant in archaeon-rich geothermal and hypersaline environments (24, 27, 33, 37). The known species have been classified in the families *Fuselloviridae* and *Bicaudaviridae* and in the genus *Salterprovirus*, while some remain unclassified. Morphologically, APSV1 is closest to the unclassified virus *Sulfolobus tengchongensis* spindle-shaped virus 1 (STSV1) (35). The spindle-shaped virions of the two viruses carry fibers which can stick to each other to form rosette-like structures (Fig. 2d). Virions of both viruses can adopt a tadpole shape, resulting from elongation of the virion from one of the pointed ends; the protrusions from both ends were seldom able to be observed and resembled those of the *Acidianus* two-tailed virus (ATV) (7). In the last case, the tails protrude from both pointed ends of the spindle-shaped immature virion after its release from a host cell, specifically at temperatures above 75°C, similar to those of the natural environment (7). A population of ATV virions often reveals tail polymorphism, reflecting different stages of extracellular morphogenesis. Whether the tail polymorphisms of APSV1 and STSV1 have similar origins is still unclear, since we were unable to affect tail lengths in a purified virion preparation.

The genome of APSV1 was identified in two steps. The first step was a comparison of the lengths of restriction digestion fragments of the virus DNA with those estimated theoretically from the nucleotide sequence of the provirus region (beta region in Fig. 1). The second step was aimed at determining the positions of the two ends of the identified proviral sequence in the viral genome and confirming the circularization of the excised proviral sequence into a 38,049-bp-long viral genome. As with genomes of all crenarchaeal viruses with a spindle-shaped morphotype, the genome of APSV1 is circular. In the proviral state, the integrase gene is split in two parts, at the position of the 30th amino acid from the N terminus. Thus, the enzyme is a member of the SSV-type integrase family of archaeal viruses, which are split as a result of integration into the host chromosome (31). However, the amino acid sequence of the APSV1 integrase gene shows limited similarity to those of the integrases from archaeal viruses; the sequence similarity is higher with some cellular integrases (Table 2). However, these presumably cellular enzymes could also have a viral origin.

The presence of the cluster of three putative genes, ORF38-469 (for putative methylase), ORF39-350 (unknown function), and ORF40-223 (for a DNA mismatch repair DNA glycosylase), in genomes of the crenarchaeon *P. aerophilum* and *T. neutrophilus* is remarkable. The extremely high similarity even at the nucleotide sequence level suggests a recent horizontal transfer of this cluster among hyperthermophilic archaeal species which may share a habitat. The fact that these genes are transferred and maintained together suggests related functions. Examples of simultaneous transfer of functionally related genes as a single unit are known from restriction-modification systems between *Bacteria* and *Archaea* through their viruses (13).

Although virions of APSV1 strongly resemble in their morphology the virions of STSV1, the lack of any significant similarity at the genomic level precludes classification of the two viruses in the same taxonomical unit. Additional information, mainly on the tertiary structures of virion proteins, would be required for revealing evolutionary relationships between these two unclassified viruses. According to the recently suggested scheme for the nomenclature of viruses of *Bacteria* and *Archaea* (10), the full name of the virus would be vA_Ape”Un”_APSV1 (“Un” is an proposed abbreviation for unclassified family).

*Aeropyrum pernix* ovoid virus 1. On negative-contrast electron micrographs, the virions of APOV appeared as pleomorphic ovoid particles, sometimes resembling droplets (Fig. 5). Cryo-electron microscopy, avoiding osmotic artifacts caused by negative staining, enabled better understanding of the virion morphology. The virions embedded in amorphous ice were regular ovals of uniform size, 70 by 55 nm. They bear clear resemblance to virions of the *Sulfolobus newzealandicus* droplet-shaped virus (SNDV), a member of the family *Guttaviridae* (2). SNDV appears as pleomorphic, mostly droplet-shaped particles on negative-contrast electron micrographs, and its morphological similarity to APOV1 virions is more pronounced when cryo-electron micrographs are compared. The major structural difference between ovoid virions of APOV1 and SNDV appears to be an absence of any fibers attached to the former virion, whereas they are abundant at one end of the SNDV virion. Additionally, SNDV virions are about 1.5 times larger than APOV1 virions.

For the analysis of APOV1 DNA, it was helpful to isolate the *A. pernix* strain YKP1, which replicated only one type of virus morphologically highly similar to APOV1 (see Fig. S3 in the supplemental material). PCR products produced by random primers from the DNA of this APOV1-like virus were highly similar in sequence to fragments from the putative provirus alpha region of the *A. pernix* K1 genome, strongly suggesting that APOV1 represents an induced alpha provirus (Fig. 1). This suggestion was verified by determining the locations where two boundaries of the putative provirus would circularize the viral genome, in the course of excision. The results allowed us to conclude that APOV1 represents an induced alpha provirus and to determine that its genome sequence constitutes 13,769 bp.

Based on the available sequence data, it is not possible to unequivocally assign APOV1 to any known viral family. How-
ever, morphologically, the virion shows pronounced similarity to the virion of SNDV, the sole member of the family Guttavirusidae. There is no possibility to claim further similarities between the viruses. The available information on the SNDV is restricted to the original description of the virion and a superficial characterization of the genome, which was reported to be a circular dsDNA containing about 20 kb (2). SNDV has not been characterized further since the original description, and it is impossible to do this at present due to the absence of the virus in strain collections. Considering the similarities of the virion morphotypes of APOV1 and SNDV, as well as the fact that the genomes of both viruses are circular, double-stranded DNA molecules of comparable sizes, we propose to assign APOV1 to the family Guttavirusidae and, moreover, to consider it the new type species of the family. Such taxonomical classification will enable researchers to populate the new type species of the family. According to the recently suggested scheme for the nomenclature of viruses, the full name of the virus would be vA_ApeU_APOV1 ("U" is the abbreviation for Guttaviridae). 

Conclusions. It has been over 15 years since the genomes of the Archaea began to be sequenced, and presently, over 100 archaeal genome sequences are available, many of which contain putative proviral sequences. Nonetheless, no attempts have been made to induce the replication of these cryptic archaeal viruses. The present study is the first to attempt this, demonstrating how easily, using a simple technique, new viruses can be isolated and described from well-known organisms, and may serve as an example for further analogous studies.

ACKNOWLEDGMENTS

We thank Patrick Forterre and Mart Krupovic for helpful discussions, Yutaka Kawarabayasi for helping us with viral genome sequences, and Gerard Pehau-Arnaudet for taking cryo-electron micrographs. T.M. was supported by a grant of the Gouvernement Français (dossier no. 2008661). This work was partially supported by a grant from the Agence Nationale de la Recherche, France (Program Blanc).

REFERENCES