INTRODUCTION

RNA viruses evolve rapidly owing to their fast replication, large populations and extremely high mutation rates (Domingo & Holland, 1997). Genetic recombination, occurring when two virus strains simultaneously infect the same host cell, is also a major driving force for RNA virus diversity (Domingo & Holland, 1997). The recombination occurs via exchange of genetic material between homologous or non-homologous regions by crossing-over or through reassortment of entire genome segments (Domingo & Holland, 1997). The genetic plasticity allows RNA viruses to rapidly adapt to challenges posed by the environment, manifested e.g. by evasion of host immune response (Malim & Emerman, 2001), evolution of resistance to antiviral drugs (Nora et al., 2007), increase in virulence (Khatchikian et al., 1989) or host-range expansion (Brown, 1997; Duffy et al., 2006; Ferris et al., 2007; Gibbs & Weiller, 1999).

The currently classified dsRNA viruses belong to eight families, members of which infect a variety of hosts, including bacteria, protozoa, fungi, plants, invertebrates and vertebrates (Mertens, 2004). Despite the variability in host organisms and habitats, most dsRNA viruses have striking structural and functional similarities. Typically, the virion of a dsRNA virus is formed by concentric structural layers. The innermost capsid layers, containing RNA polymerization activity, are conserved among dsRNA viruses (Bamford et al., 2002; Luque et al., 2010), whereas the outermost layer shows greater genetic and structural diversity. The flexible composition of the outer layer allows the virus to adapt for transmission and initiation of infection in different host cells (Bamford et al., 2002; Poranen & Bamford, 2012b).

Bacteriophages of the family Cystoviridae have genomes composed of three dsRNA segments enclosed in polyhedral protein capsids which in turn are surrounded by protein–lipid envelope (Poranen & Bamford, 2012a). Owing to the dsRNA genome and the outer membrane, cystoviruses represent a unique group among known bacteriophages. The inner core of the virion is composed of 120 copies of the major capsid protein (Kistakiskis & Lang, 1987; Olkkonen & Bamford, 1987) arranged in a T=1 icosahedral lattice (Butcher et al., 1997; Huiskonen et al., 2006). This rare architecture of the innermost protein layer is also seen in eukaryotic dsRNA viruses, for example in the members of the family Reoviridae (Poranen & Bamford, 2012b).

In addition to Pseudomonas phage φ6 (Vidaver et al., 1973), the type virus of the Cystoviridae, several putative cystoviruses have been isolated (Mindich et al., 1999; O’Keefe et al., 2010; Silander et al., 2005; Qiao et al., 2010), but have not been formally classified by the International...
Committee on Taxonomy of Viruses. All isolates were obtained from plant debris in the USA, and they infect Gram-negative bacteria, mainly phytopathogenic Pseudomonas syringae strains. Complete genome sequences are available for four cystoviral isolates in addition to φ6 (Gottlieb et al., 1988, 2002; Hoogstraten et al., 2000; McGraw et al., 1986; Mindich et al., 1988; Qiao et al., 2010). Some cystoviruses are genetically remarkably similar to φ6, whereas others are only distantly related. Phage φ6 and its distant relative φ2954 (Mindich et al., 1999; Qiao et al., 2010) attach to a specific type IV pilus on the host bacterium, whereas the other distant relatives of φ6 (φ8, φ12 and φ13) bind directly to rough lipopolysaccharide on the host outer membrane (Mindich et al., 1999).

In cystoviruses, the general mechanism of recombination is template switching (Qiao et al., 1997). Exogenous genetic material, such as cellular transcripts or RNA strands from other viruses, may also be incorporated into cystovirus procapsids and recombine with endogenous viral RNA (Onodera et al., 2001). However, the probability of the packaging of two cystoviral RNA molecules of the same segment class into the same procapsid is very low owing to the strictly controlled packaging mechanism (Mindich, 1999), making homologous recombination an extremely rare event, both in vitro (Onodera et al., 2001) and in natural populations (Silander et al., 2005).

Reassortment of genome segments occurs frequently among cystoviruses in laboratory conditions as well as in nature (Onodera et al., 2001; Silander et al., 2005). Surprisingly, the rate of reassortment may vary among populations sampled from different geographical locations (O’Keefe et al., 2010). Additional genetic diversity is acquired via mutations occurring at the rate of ~10⁻⁶ per locus per generation (Burch et al., 2007; Chao et al., 2002; Ferris et al., 2007), although higher values have also been suggested, especially in terms of host-range mutations (Ferris et al., 2007).

Here, we present a putative new member of the Cystoviridae, which we name Pseudomonas phage φNN. φNN has a unique freshwater habitat among cystoviruses, demonstrating that this group is more diverse and widely distributed than previously known. Phylogenetic relationships in light of the putative new member are also discussed. φNN shares striking genetic similarity to φ6, whereas the other previously identified cystoviruses are more distantly related. There are several X-ray and electron microscopic structures available for various cystoviral proteins. Thus, it was feasible to build atomic-resolution models of the φNN protein in order to investigate its relationship with the known cystoviruses at the structural level also.

RESULTS

φNN virion

Bacteriophage φNN was isolated with its host Pseudomonas sp. B314 from Lake Vehkalampi in Central Finland (N 62° 14’ 42”, E 25° 42’ 54”). Virus production was most efficient when the host cells were infected in the middle of the exponential growth phase (~1.5 x 10⁹ c.f.u. ml⁻¹) with m.o.i 2–5. The duration of the latent period (~100 min; Fig. 1a) was similar to that described for bacteriophage φ6 (80–160 min, depending on the culture conditions; Vidaver et al., 1973), while the measured mean burst size (~80) was somewhat lower than in φ6 (125–400, depending on the culture conditions; Vidaver et al., 1973). However, the duration and extent of the bacterial lysis as well as the overall phage production in φNN infection varied somewhat between different cultures.

The chloroform sensitivity of φNN indicated a possible lipid component. Transmission electron microscope (TEM) imaging revealed that the size and shape of φNN resemble those of φ6 and other putative cystoviruses. The mature virion is tail-less and round; however, empty icosahedrally symmetrical structures analogous to the core particles of φ6 were also observed from purified φNN samples (Fig. 1b).

Based on SDS-PAGE, the protein profile of φNN was similar to that of φ6 (Fig. 1c). The proteins of φNN were named according to the corresponding proteins of φ6 (Fig. 1c, d, Table 1). Bands for the internally located RNA-dependent RNA polymerase, P2, and the external spike protein, P3, were assigned based on their stoichiometry. To confirm this interpretation, the electrophoretic analysis was repeated after treating the virions with butylated hydroxytoluene, which specifically removes the P3 spikes from the virion surface (Bamford et al., 1995; and data not shown). Furthermore, Triton X-100 solubilized/removed proteins P3, P6 and P9, which indicates their presence on the outer membrane (data not shown). The sensitivity of P3 and the resistance of P2 to these treatments support the interpretation made by stoichiometry.

Agarose gel electrophoresis (AGE) of dsRNA extracted from purified φNN virions revealed the presence of three dsRNA fragments, which were named according to their size as small (S), medium (M) and large (L). Mobility in agarose gel indicated that the sizes of the S- and L-segments resemble those of φ6, whereas the M-segment is smaller than that of φ6 (Fig. 1e).

Host interaction of φNN

Cross-infection studies of bacteriophages φNN and φ6 were conducted with their isolated hosts Pseudomonas sp. B314 and P. syringae pv. phaseolicola strain HB10Y (Vidaver et al., 1973), respectively. HB10Y has type IV pili and its outer membrane is covered with smooth lipopolysaccharide (Mindich et al., 1999). In spot assay, φNN and φ6 were incapable of infecting each other’s host strains (data not shown). TEM imaging of φNN infecting its host showed virions attached to long, thin protrusions (Fig. 1f), possibly indicating type IV pilus-mediated infection, previously shown for φ6 and φ2954 (Mindich et al., 1999; Qiao et al., 2010).
Fig. 1. \(^{\Phi}NN\) life cycle and the virion. (a) Turbidity (OD\(_{550}\)) of the infected (dashed line) and uninfected (solid line) \(^{\Phi}NN\) sp. B314 culture. The culture was infected with \(^{\Phi}NN\) virions at m.o.i. 2. Time 0 indicates the time of infection. (b) Transmission electron micrograph of bacteriophage \(^{\Phi}NN\). Purified virions were negatively stained in ammonium molybdate (pH 7.4) and imaged by a TEM at 80 kV. In addition to mature virions, empty core particles were also seen (box in upper right corner). Bar, 80 nm. (c) Protein profiles of \(^{\Phi}6\) and \(^{\Phi}NN\) by SDS-PAGE analysis. Purified virions were disrupted by boiling and loaded onto a 15 % polyacrylamide gel. The mobility of the \(^{\Phi}6\) proteins is indicated on the left and that of corresponding proteins of \(^{\Phi}NN\) on the right. Protein P6 of \(^{\Phi}6\) is not visible in the image but was seen on the gel. (d) Schematic representation of \(^{\Phi}6\) virion. Three dsRNA genome segments (S, M and L) are enclosed in an inner core (proteins P1, P2, P4 and P7), which is surrounded by an outer protein shell (protein P8). The outermost surface consists of a lipid envelope with embedded phage-encoded proteins and P3 spikes. (e) AGE analysis of dsRNA genomes of bacteriophages \(^{\Phi}6\) and \(^{\Phi}NN\). Extracted dsRNA was subjected to electrophoresis in 1 % agarose gel. The mobility of the \(^{\Phi}6\) segments L (6374 bp), M (4063 bp) and S (2948 bp) is indicated on the left and that of corresponding \(^{\Phi}NN\) segments on the right. (f) Transmission electron micrograph of \(^{\Phi}NN\) virions attaching to a pilus on \(^{\Phi}NN\) sp. B314 at 2 min post-infection. Bar, 400 nm.

Table 1. Amino acid sequence similarities between ORFs of \(^{\Phi}NN\) and those of previously identified cystoviruses

<table>
<thead>
<tr>
<th>(^{\Phi}NN)</th>
<th>MW (kDa)</th>
<th>Function based on (^{\Phi}6) annotation</th>
<th>Amino acid similarity (%)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>P14</td>
<td>7</td>
<td>Non-structural, non-essential</td>
</tr>
<tr>
<td></td>
<td>P7</td>
<td>17</td>
<td>Assembly factor</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>75</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
<tr>
<td></td>
<td>P4</td>
<td>35</td>
<td>Packaging NTPase</td>
</tr>
<tr>
<td></td>
<td>P1</td>
<td>85</td>
<td>Major capsid protein, forms the T=1 shell</td>
</tr>
<tr>
<td>M</td>
<td>P10</td>
<td>6</td>
<td>Membrane protein, needed in cell lysis</td>
</tr>
<tr>
<td></td>
<td>P6</td>
<td>19</td>
<td>Fusogenic membrane protein, binds to P3</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>71</td>
<td>Spike protein, binds to the host receptor</td>
</tr>
<tr>
<td></td>
<td>P13</td>
<td>8</td>
<td>Minor membrane protein</td>
</tr>
<tr>
<td>S</td>
<td>P8</td>
<td>16</td>
<td>Forms the T=13 nucleocapsid shell</td>
</tr>
<tr>
<td></td>
<td>P12</td>
<td>20</td>
<td>Assembly factor in viral membrane morphogenesis</td>
</tr>
<tr>
<td></td>
<td>P9</td>
<td>10</td>
<td>Major membrane protein</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>24</td>
<td>Muralytic enzyme, needed in host entry and lysis</td>
</tr>
</tbody>
</table>

ORFs are listed in their genomic order.

*Corresponding cystoviral ORF was not detected. The most similar peptide/protein based on genome location and annotation for \(^{\Phi}NN\) P3 from \(^{\Phi}13\) is P3a (*), for \(^{\Phi}NN\) P3 from \(^{\Phi}12\), it is P3c (*), for \(^{\Phi}NN\) P3 from \(^{\Phi}8\), it is P3b (*), for \(^{\Phi}NN\) P13 from \(^{\Phi}8\), it is PF (*) and for \(^{\Phi}NN\) P5 from \(^{\Phi}6\), it is P5a (*), respectively.
The host bacterium of φNN was characterized by 16S rRNA sequence analysis and selected biophysical tests (conducted by DSMZ). According to these assays, the strain belongs to the genus *Pseudomonas*, and is presumably *P. syringae* or a closely related species.

**Sequencing and genome analysis of φNN**

The dsRNA genome of φNN was converted into complementary DNA (cDNA) fragments which were sequenced. Overlapping sites of the sequences were aligned to produce complete nucleotide sequences for the three genome segments (Fig. 2). The sizes of the segments are 2945 (S), 3814 (M) and 6503 (L) bp, comparable with the genome segments of φ6, which are 2948, 4063 and 6374 bp, respectively (Fig. 1e). The extreme 5′ end of the φNN L-segment could not be confirmed, and it may contain a few additional nucleotides. The GC-content is 53, 55 and 55 % for φNN S, M and L, respectively, compared with about 56 % GC-content of the φ6 genome (Mindich, 1988).

Thirteen potential ORFs were identified from the genome of φNN. The genetic organization of φNN resembles that of φ6 and other identified cystoviruses (Gottlieb et al., 1988, 2002; Hoogstraten et al., 2000; McGraw et al., 1986; Mindich et al., 1988, 2000, 2010), and putative functions could be assigned to the ORFs based on chromosomal position, size and sequence homology with φ6 (Fig. 2, Table 1).

Similarly to other cystoviruses, all the ORFs of the φNN genome are oriented in the same direction (Gottlieb et al., 1988, 2002; Hoogstraten et al., 2000; McGraw et al., 1986; Mindich et al., 1988; Qiao et al., 2000, 2010), and putative ORFs/genes of the M-segment was significantly lower (40–60 % similarity; Table 1). The highest similarities (89–98 %; Table 1) between translated ORFs/genes located in the S- and L-segments (highest in ORF 1/gene 1 and ORF 2/gene 2), whereas similarity in the ORFs/genes of the M-segment was significantly lower (40–69 %; Table 1), being lowest in the ORF/gene encoding the host attachment protein P3 (40 % similarity; Table 1). The sequence relatedness in ORFs predicted to encode nucleocapsid proteins (P1, P2, P4, P7, P8; 41–60 % similarity; Table 1) was also prominent between φNN and φ13. In general, putative ORFs of essential enzymes and proteins in the internal parts of the virion (Fig. 1d) shared greater amino acid sequence similarity between φNN and the more distant relatives, φ8, φ12, φ13 and φ2954 (11–60 %; Table 1), whereas the ORFs of the host recognition protein P3 (17–28 % similarity; Table 1) and the other membrane-associated proteins differed more.

**Structural modelling of φNN putative proteins**

Structures of all predicted φNN proteins were modelled and the models were quality controlled. The models for putative proteins of the capsid protein, P1, polymerase complex, P2, packaging NTPase, P4, and the nucleocapsid-associated lytic enzyme, P5 (Fig. 1d, Table 1), were also been demonstrated for other cystoviruses (Gottlieb et al., 2002; Hoogstraten et al., 2000; McGraw et al., 1986; Mindich et al., 1988; Qiao et al., 2000, 2010).

The nucleotide sequences of φNN genome segments were compared with previously identified cystoviruses for which complete genome sequences are available (φ6, φ8, φ12, φ13 and φ2954), and phylogenetic trees were reconstructed from these comparisons (Fig. 3). Each segment of φNN is closely related to the corresponding segment of φ6, the M-segments being the most divergent. Furthermore, the L-segments of φNN and φ13 are moderately related.

The high amino acid sequence relatedness between φNN and φ6 was evident when visualizing the relationships between φNN and the previously identified cystoviruses (Fig. S1, available in the online Supplementary Material). The host attachment protein P3 (40 % similarity; Table 1) was also prominent between φNN and φ13. In general, putative ORFs of essential enzymes and proteins in the internal parts of the virion (Fig. 1d) shared greater amino acid sequence similarity between φNN and the more distant relatives, φ8, φ12, φ13 and φ2954 (11–60 %; Table 1), whereas the ORFs of the host recognition protein P3 (17–28 % similarity; Table 1) and the other membrane-associated proteins differed more.

**Fig. 2.** Genetic maps of the φNN genome segments. ORFs, indicated by numbered boxes, were assigned based on similarity with φ6. The sizes of the S-, M- and L-segments are 2945, 3814 and 6503 bp, respectively. Proteins that form the nucleocapsid or are associated with it are indicated with dark grey and membrane-associated proteins with light grey.
considered to have a correct fold. The models of P1, P2, P4 and P5 all had ProSA-web Z-scores (−9.84, −10.57, −7.69 and −7.12, respectively) that are well within acceptable range and characteristic for native proteins (Figs S2–S5). These models were also mainly based on templates from homologous proteins of cystoviruses. The best models are presented, along with structural alignments to other known cystoviral structures, in Fig. 4, and additional details are provided in Figs S2–S5.

The model of the putative P1 major capsid protein of ϕNN is based mainly on P1 of ϕ6 (PDBid 4BTG; Nemecek et al., 2013; 98% sequence similarity). The Homology Structure Finder (HSF)-program (Ravantti et al., 2013) found 600 residues structurally equivalent between ϕNN and ϕ8 P1 (4BTP; El Omari et al., 2013a; 78% coverage of ϕNN P1), although sequence-level similarity is only 29%. The ϕNN P2 polymerase model (Fig. 4) is based on the ϕ6 structure (PDBid 1HI8; Butcher et al., 2001), and shares 574 equivalent residues (86% coverage of ϕNN P2) with the ϕ6 and ϕ12 P2 structures (Butcher et al., 2001; Ren et al., 2013). This is expected, as viral RNA polymerases are structurally highly conserved (Bruenn, 2003; Mönttinen et al., 2014; Poch et al., 1989). There are several cystoviral packaging NTPase P4 structures available in the Protein Data Bank (PDB) (PDBid 4BLO, 4BLP, 4BLQ and 1W44; El Omari et al., 2013b; Mancini et al., 2004; Fig. 4). The ϕNN P4 model is most similar to the ϕ6 P4 structure (Fig. 4), but interestingly the structural similarity to the ϕ8 P4 is higher than to the ϕ12 P4, although the ϕ12 P4 has higher sequence similarity to the putative ϕNN P4. These cystoviral NTPases, including the ϕNN model presented here, share a common catalytic core (172 equivalent residues by HSF-program), but they have different specificities and control mechanisms, which can be mapped onto the divergent N- and C-terminal domains (El Omari et al., 2013b).

Since the overall sequence similarity between ϕ6 and ϕNN is high (Table 1), it is not surprising that most structural models for ϕNN proteins are based on structures derived from ϕ6, the muralytic enzyme, P5, being no exception. The ϕNN P5 model (Fig. 4) covers all 160 residues solved by X-ray crystallography for ϕ6 P5 (PDBid 4DQ5).

DISCUSSION

A freshwater bacteriophage, ϕNN, was isolated and shown to have a tripartite dsRNA genome and a lipid-containing virion. Further characterization revealed that the ϕNN virion is tail-less and contains an icosahedrally symmetrical core particle (Fig. 1b). Furthermore, the genome content and protein profile of ϕNN resemble those of ϕ6, the type member of the family Cystoviridae (Fig. 1c–e). Owing to these distinctive similarities, we propose that ϕNN is a new member of the Cystoviridae.

Bacteriophage ϕ6, isolated over 40 years ago, is still the only classified member of the Cystoviridae (Poranen & Barnford, 2012a). However, it has been shown that bacteriophages with a tripartite dsRNA genome can be readily isolated from agricultural plant species (Mendich et al., 1999; O’Keefe et al., 2010; Silander et al., 2005; Qiao et al., 2010). Although these additional viral isolates have not been specifically identified or classified, their high frequency in agricultural samples indicates that dsRNA bacteriophages are more common in terrestrial habitats than previously estimated. All the previous cystoviruses are
from legume samples from certain parts of the USA (Nebraska, New England, California, Connecticut). In contrast, \( \phi N N \) was isolated in Central Finland from a freshwater sample, demonstrating that cystoviruses are more widely distributed than previously known.

Despite the clear similarities between \( \phi N N \) and \( \phi 6 \), they did not infect each other’s natural host strains. According to 16S rRNA sequence analysis and several biophysical tests, the host strain of \( \phi N N \) belongs to the genus \( \textit{Pseudomonas} \) and is presumably \( \textit{P. syringae} \) or a closely related species. Electron microscope imaging (Fig. 1f) indicated that \( \phi N N \) likely uses type IV pilus as its primary receptor, similarly to \( \phi 6 \) and \( \phi 2954 \) (Mindich \textit{et al.}, 1999; Qiao \textit{et al.}, 2010). However, the sites recognized by \( \phi N N \) and \( \phi 6 \) are apparently non-conserved in their \( \textit{Pseudomonas} \) hosts.

The sizes and gene arrangements of \( \phi N N \) genome segments resemble closely those of \( \phi 6 \) (Figs 1e and 2). All in all, the genetic organization is relatively conserved within the identified cystoviruses, and the ORFs genes cluster into functional groups (Gottlieb \textit{et al.}, 1988, 2000; McGraw \textit{et al.}, 1986; Mindich \textit{et al.}, 1988; Qiao \textit{et al.}, 2000, 2010). For instance, the highly conserved L-segment includes ORFs genes encoding the components of the virion core containing the polymerase activity (Figs 1d and 2) whereas the M-segment, which varies the most among cystoviruses, contains ORFs genes encoding envelope proteins (Figs 1d and 2).

Phylogenetic trees of the genome segments indicate a close relationship between \( \phi N N \) and \( \phi 6 \) (Fig. 3). The sequence alignments reveal that the predominant amino acid sequence variation between \( \phi N N \) and \( \phi 6 \) is in the spike protein P3 (Fig. S1). This putative host attachment protein of \( \phi N N \) differs greatly from the corresponding proteins of other identified cystoviruses, the most similar one being the P3 of \( \phi 2954 \) (Table 1). Cystoviruses \( \phi 6 \), \( \phi N N \) and \( \phi 2954 \) use pilus-mediated infection (Mindich \textit{et al.}, 1999; Qiao \textit{et al.}, 2010; Fig. 1f) and have attachment complexes consisting of a single P3 polypeptide (Gottlieb \textit{et al.}, 1988; Qiao \textit{et al.}, 2010; Fig. 1c, Table 1), whereas other previously identified cystoviruses (\( \phi 8 \), \( \phi 12 \) and \( \phi 13 \)) interact directly with the layer of rough lipopolysaccharide and have a more complex heteromeric attachment.
apparatus with at least two or three polypeptides (Gottlieb et al., 2002; Hoogstraten et al., 2000; Mindich et al., 1999; Qiao et al., 2000). The similarity in infection mechanisms could explain the moderate sequence similarity between the host attachment proteins of $\phi_6$, $\phi_{NN}$ and $\phi_{2954}$.

The frequent genome segment reassortment of cystoviruses has been demonstrated in laboratory conditions and in nature (Onodera et al., 2001; Silander et al., 2005). Such a phenomenon could also explain the higher sequence variability between the M-segments of $\phi_{NN}$ and $\phi_6$ compared with the other two segments. Furthermore, strong selection pressure is presumably exerted on the M-segment-encoded proteins during co-evolution with the host, as demonstrated by the susceptibility of the host attachment protein P3 of $\phi_6$ to spontaneous mutations, resulting in phages with varying host range (Duffy et al., 2006; Ferris et al., 2007). Recombination by crossing-over with genetic material of the host cell or other viruses is an unlikely explanation for the observed sequence variation (Onodera et al., 2001) as homologous sequences for $\phi_{NN}$ were not found outside previously identified cystoviruses.

In contrast to protein components of the outer membrane, the internal major structural proteins and the key viral enzymes, encoded by the S- and L-segments (Table 1), are relatively conserved, indicating that host–virus interactions are not significantly driving their evolution. The similarity of these proteins appears to be even higher at structural level (Fig. 4), underlining how RNA viruses undergo rapid evolution affecting the primary sequence yet conserving the structures needed to maintain a viable virus.

Structural variation in the outermost layer of the virion is commonly observed in animal viruses. For instance, the outermost layer of different reoviruses consists of a variety of unrelated proteins (Poranen & Bamford, 2012b). Significant variation is also seen in the surface glycoproteins (haemagglutinin and neuraminase) of influenza viruses. Since these proteins are the major target of neutralizing antibodies of the host immune system, their rapid evolution is presumed to be at least partly driven by the selection pressure caused by the host’s antibody-mediated immunity system (Webster et al., 1992). However, antibody-mediated selection pressure cannot be the driving force for the variation detected here in the external layers of bacterial viruses. Apparently, changes in the outer layer accumulate to broaden the host range of the virus, whereas protein–protein associations inside the virion are highly coupled, so that even minor structural changes in any of the internal proteins may disturb virion assembly. Furthermore, essential viral enzymes, which likely are more sensitive to variations, are typically located in the interior of the virion.

Even though $\phi_{NN}$ and $\phi_6$ were isolated from highly diverse ecological habitats on different continents at a >40-year interval, these two phages are significantly similar. Their resemblance reflects the conservation of key viral elements, while simultaneously the inherent viral ability to broaden the host range leads to ecological plasticity.

**METHODS**

**Bacteriophages, bacterial strains and plasmids.** $\phi_{NN}$ and its host bacterium *Pseudomonas* sp. B314 were isolated in this study. *Pseudomonas* phage $\phi_6$ and its natural host *P. syringae* pv. phaseolicola strain HB10Y (Vidaver et al., 1973) were obtained from the laboratory of D. H. Bamford (University of Helsinki). A high-copy-number plasmid, pUC18 (Yanisch-Perron et al., 1985), was used for the cloning of cDNA copies of the $\phi_{NN}$ genome. Recombinant plasmids were propagated in *Escherichia coli* DH5a (Sambrook et al., 1989).

Bacteria were grown in Luria–Bertani (LB) medium (Sinclair et al., 1976). Selective plates contained 200 µg ampicillin ml$^{-1}$ in LB agar and were supplemented with 0.1 mM IPTG and 50 µg X-Gal ml$^{-1}$ for blue–white screening (Maniatis et al., 1982).

**Characterization of the host bacterium.** The 16S rRNA gene of *Pseudomonas* sp. B314 was amplified by PCR using primers 1D1 and rD1 (Weisburg et al., 1991) and DreamTaq DNA polymerase (Thermo Scientific). Bacterial genome extracted with the GeneJet Genomic DNA Purification kit (Thermo Scientific) was used as a template. The 16S rRNA-specific DNA was purified with the QIAquick PCR Purification kit (Qiagen), and its nucleotide sequence was determined using the BigDye Terminator v3.1 kit and ABI Prism Genetic Analyzer 3100 (Life Technologies). The microbiological characterization of *Pseudomonas* sp. B314 strain was commercially conducted by DSMZ.

**Isolation, growth and purification of viruses.** *Pseudomonas* sp. B314-specific viruses were isolated from a filtered water sample (Micropore, 0.45 µm filter) using enrichment culture containing the host bacteria in 20 % (v/v) LB. After incubation at room temperature (RT) for 2 days with agitation (110 r.p.m.), samples of the enrichment culture were mixed with 0.7 % (w/v) soft agar in 20 % (v/v) LB and plated on the same medium supplemented with 1 % (w/v) agar. The plates were incubated at RT for 2–3 days, then independent plaques were picked, suspended in 20 % LB medium and stored at 4 °C. Virus stocks were prepared by collecting the top agar from plates showing semi-confluent lysis of the bacterial lawn by the phage as previously described for phage $\phi_6$ (Bamford et al., 1995).

For large-scale liquid culture, host cells grown (26 °C, 220 r.p.m.) to exponential growth phase (~1.5 $\times$ 10$^9$ cf.u. ml$^{-1}$) were infected by $\phi_{NN}$ or $\phi_6$ at m.o.i. 2–5. Culturing was continued (26 °C, 110 r.p.m.) in the presence of 5 µg DNase 1 ml$^{-1}$ (Sigma-Aldrich) until lysis. Cell debris was removed by centrifugation (Sorvall SLA3000 rotor, 10,800 g, 15 min, 4 °C) and virions were concentrated by precipitation with 10 % (w/v) polyethylene glycol 6000 and 0.5 M NaCl (Bamford et al., 1995). Phage precipitate was collected by centrifugation (Sorvall SLA3000, 10,800 g, 20 min) and suspended in 20 mM potassium phosphate buffer (PPB, pH 7.2). Aggregates were removed from the concentrated phage preparation by centrifugation (Sorvall SS-34, 3000 g, 5 min, 4 °C) and the clear supernatant was layered on top of a 5–20 % (w/v) sucrose gradient (in 20 mM PPB, pH 7.2) for subsequent centrifugation (Beckman Optima L-90K ultracentrifuge, SW28 rotor, 76,000 g, 40 min, 15 °C). The light-scattering zone, containing the phage, was collected and subjected to equilibrium centrifugation in a 20–70 % (w/v) sucrose gradient (in 20 mM PPB, pH 7.2) and stored at ~80 °C. The non-ionic detergent Triton X-100 was used to remove the viral membrane and membrane-associated proteins from the virions (Bamford & Palva, 1980). The protein profiles of the purified virions were analysed by SDS-PAGE (Laemmli, 1970) in 15 % polyacrylamide gel and subsequently scanned using a Bio-Rad imaging device.
The analysis of the burst size was carried out as described by Dennehy & Turner (2004). The unbound viruses were removed by centrifugation of the culture and resuspension of the cells into fresh medium at 30 min post-infection.

**Host-range investigation and chloroform-sensitivity assay.** *Pseudomonas* strains were grown in LB overnight, after which 200 μl each culture was mixed with 3 ml molten soft agar and laid onto an LB plate. Ten microlitres of viral dilutions was spotted onto the solidified plate. After overnight incubation at RT, the plates were examined for bacterial lysis.

For the chloroform-sensitivity assay, a drop of chloroform (~50 μl) was added to 100 μl virus stock and the solution was mixed with vigorous shaking. After 15 min, 5–10 μl of the aqueous phase or an untreated sample of the phage stock was spotted onto a lawn of the host bacterium.

**Electron microscopy.** The morphology of virions and their attachment to host cells were examined by negative-stain transmission electron microscopy. Purified viral particles or samples of infected host culture were deposited onto coated copper grids for 1–3 min. The grids were subsequently stained with ammonium molybdate (pH 7.4) for 30 s–1 min and examined by means of a JEOL JEM-1200EX TEM at an accelerating voltage of 80 kV.

**Isolation of viral dsRNA and cDNA synthesis.** Genomic dsRNA (~3 μg) was denatured by boiling for 5 min in 15 % (v/v) DMSO and cooled in ice-water slurry. Poly(A) tails were added to the 3' ends of the RNA strands using *E. coli* poly(A) polymerase (Life Technologies). The RNA was then purified by phenol/chloroform extraction, precipitated with ammonium acetate and ethanol, and resuspended in 10 μl nuclease-free H₂O.

The poly(A)-tailed RNA was combined with sequence-specific primers (0.5 μg μg⁻¹ RNA) and heated to 96 °C for 3 min. The sample was cooled on ice before the cDNA was synthesized with the Universal RiboClone cDNA Synthesis kit (Promega) according to manufacturer’s instructions. The resulting cDNA was again purified, precipitated and resuspended in 10 μl nuclease-free H₂O.

Plasmid pUC18 was digested with *Smal* (Thermo Scientific), dephosphorylated with FasTAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), extracted from agarose gel after electrophoresis, and purified using the Qiagen Gel Extraction kit. The cDNA fragments of the dsRNA genome (~100–400 ng) were blunt-end ligated into the *Smal* site of pUC18 (~50 ng) using T4 DNA ligase (400 U; New England Biolabs) in 1 x ligation buffer containing 20 % (w/v) polyethylene glycol 6000. After overnight incubation at 16 °C, the resulting DNA molecules were transformed into competent *E. coli* DH5α cells using a heat-shock method. The transformants were plated onto selective LB plates containing IPTG and X-Gal, and cultivated at 37 °C overnight. Plasmid-DNA was purified from selected clones by using the Qiagen Miniprep kit and digested with EcoRI and HindIII (Thermo Scientific) to verify the presence of an insert.

Both DNA strands of the cloned cDNA fragments were sequenced at least three times using the BigDye Terminator v3.1 Cycle Sequencing kit and an ABI Prism Genetic Analyzer 3100 (Life Technologies). First, M13/pUC primers were used, but later the sequencing reactions were performed with specific oligonucleotides designed based on sequence information already revealed.

**Bioinformatical analysis.** A nucleotide-based sequence similarity search was performed for the 16S rRNA gene of *Pseudomonas* sp. B314, using the Basic Local Alignment Search Tool (http://www.blast.ncbi.nlm.nih.gov/Blatt.cgi). The genome of φNN was screened using the programs NCBI ORF finder (http://www.ncbi.nlm.nih.gov/orffinder.cgi), Genemark (Besemer et al., 2001) and VectorNTI (Life Technologies) for possible protein-encoding ORFs. Each ORF was visually inspected and compared with those of existing putative cystoviruses to confirm its location within the genome segments. The GenBank accession numbers for the nucleotide sequences of φNN genome segments S, M and L are KJ957166, KJ957165 and KJ957164, respectively. φNN was compared with those previously identified cystoviruses for which complete genome sequences are available. These are φβ (GenBank accession numbers for S, M and L, respectively: NC_003714.1, NC_003716.1, NC_003715.1), φβ (NC_003301.1, NC_003300.1, NC_003299.1), φ2 (NC_004174.1, NC_004175.1, NC_004173.1), φ13 (NC_004170.1, NC_004171.1, NC_004172.1) and φ2956 (NC_012093, NC_012092, NC_012091.2). The genome segments were aligned separately per segment type (S, M and L) using Clustal Omega (Sievers et al., 2011) and visualized using Dendroscope3 (Huson & Scornavacca, 2012).

The ORFs were translated into amino acid sequences with the Emboss Transeq program (Rice et al., 2000). Statistics for putative proteins were calculated using Emboss Pepstat (Rice et al., 2000). Each putative protein was compared against all other putative cystoviral proteins using Emboss Needle (Rice et al., 2000). The most similar proteins to φNN ORFs from each phage were selected and their similarity was visualized (Fig. S1) using Circos (Krzywinski et al., 2009).

An online version of the i-TASSER program (Roy et al., 2010; Roy & Zhang, 2012) was used to model all predicted φNN proteins. Only those putative proteins that had a TM-score of the model >0.5 or had a known cystoviral structure available for comparison (i.e. putative protein P5, TM-score 0.49) were used. The model quality was assessed using ProSA-web (Sippl, 1993; Wiederstein & Sippl, 2007). The models were aligned and their equivalent residues detected by the HSF-program (Ravantti et al., 2013). The cystoviral structures were downloaded from the PDB and visualized using PyMOL (www.pymol.org).

**ACKNOWLEDGEMENTS**

This study was supported by the Academy of Finland Centre of Excellence Programme in Virus Research (1129684, 2006–2011 to J.K. H.B.), the Center of Excellence Programme in Biological Interactions (252411, 2012–2017 to J.K. H.B.), the Academy of Finland grants 251106 (to J.K. H.B.), 250113, 256069, 283192 and 272507 (to M. M. P.), and a grant from the Finnish Cultural Foundation (00110600 to S. M.). We thank the Academy of Finland (grants 271413 and 272507) orfig.cgi), Genemark (Besemer et al., 2001) and VectorNTI (Life Technologies) for possible protein-encoding ORFs. Each ORF was visually inspected and compared with those of existing putative cystoviruses to confirm its location within the genome segments. The GenBank accession numbers for the nucleotide sequences of φNN genome segments S, M and L are KJ957166, KJ957165 and KJ957164, respectively. φNN was compared with those previously identified cystoviruses for which complete genome sequences are available. These are φβ (GenBank accession numbers for S, M and L, respectively: NC_003714.1, NC_003716.1, NC_003715.1), φβ (NC_003301.1, NC_003300.1, NC_003299.1), φ2 (NC_004174.1, NC_004175.1, NC_004173.1), φ13 (NC_004170.1, NC_004171.1, NC_004172.1) and φ2954 (NC_012093, NC_012092, NC_012091.2). The genome segments were aligned separately per segment type (S, M and L) using Clustal Omega (Sievers et al., 2011) and visualized using Dendroscope3 (Huson & Scornavacca, 2012).

The ORFs were translated into amino acid sequences with the Emboss Transeq program (Rice et al., 2000). Statistics for putative proteins were calculated using Emboss Pepstat (Rice et al., 2000). Each putative protein was compared against all other putative cystoviral proteins using Emboss Needle (Rice et al., 2000). The most similar proteins to φNN ORFs from each phage were selected and their similarity was visualized (Fig. S1) using Circos (Krzywinski et al., 2009).

An online version of the i-TASSER program (Roy et al., 2010; Roy & Zhang, 2012) was used to model all predicted φNN proteins. Only those putative proteins that had a TM-score of the model >0.5 or had a known cystoviral structure available for comparison (i.e. putative protein P5, TM-score 0.49) were used. The model quality was assessed using ProSA-web (Sippl, 1993; Wiederstein & Sippl, 2007). The models were aligned and their equivalent residues detected by the HSF-program (Ravantti et al., 2013). The cystoviral structures were downloaded from the PDB and visualized using PyMOL (www.pymol.org).

**REFERENCES**


