

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Kejzar, Nejc; Laanto, Elina; Rissanen, Ilona; Abrishami, Vahid; Selvaraj, Muniyandi; Moineau, Sylvain; Ravantti, Janne; Sundberg, Lotta-Riina; Huiskonen, Juha T.

**Title:** Cryo-EM structure of ssDNA bacteriophage  $\Phi$ CjT23 provides insight into early virus evolution

**Year:** 2022

**Version:** Published version

**Copyright:** © The Author(s) 2022

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Kejzar, N., Laanto, E., Rissanen, I., Abrishami, V., Selvaraj, M., Moineau, S., Ravantti, J., Sundberg, L.-R., & Huiskonen, J. T. (2022). Cryo-EM structure of ssDNA bacteriophage  $\Phi$ CjT23 provides insight into early virus evolution. *Nature Communications*, 13, Article 7478. <https://doi.org/10.1038/s41467-022-35123-6>

# Cryo-EM structure of ssDNA bacteriophage $\Phi$ CjT23 provides insight into early virus evolution

Received: 9 June 2022

Accepted: 18 November 2022

Published online: 03 December 2022

 Check for updates

Nejc Kejzar<sup>1,5,7</sup>, Elina Laanto<sup>2,3,7</sup>, Ilona Rissanen<sup>1,7</sup>, Vahid Abrishami<sup>1</sup>, Muniyandi Selvaraj<sup>1,6</sup>, Sylvain Moineau<sup>4</sup>, Janne Ravantti<sup>2</sup>, Lotta-Riina Sundberg<sup>3</sup>✉ & Juha T. Huiskonen<sup>1</sup>✉

The origin of viruses remains an open question. While lack of detectable sequence similarity hampers the analysis of distantly related viruses, structural biology investigations of conserved capsid protein structures facilitate the study of distant evolutionary relationships. Here we characterize the lipid-containing ssDNA temperate bacteriophage  $\Phi$ CjT23, which infects *Flavobacterium* sp. (*Bacteroidetes*). We report  $\Phi$ CjT23-like sequences in the genome of strains belonging to several *Flavobacterium* species. The virion structure determined by cryogenic electron microscopy reveals similarities to members of the viral kingdom *Bamfordvirae* that currently consists solely of dsDNA viruses with a major capsid protein composed of two upright  $\beta$ -sandwiches. The minimalistic structure of  $\Phi$ CjT23 suggests that this phage serves as a model for the last common ancestor between ssDNA and dsDNA viruses in the *Bamfordvirae*. Both  $\Phi$ CjT23 and the related phage FLiP infect *Flavobacterium* species found in several environments, suggesting that these types of viruses have a global distribution and a shared evolutionary origin. Detailed comparisons to related, more complex viruses not only expand our knowledge about this group of viruses but also provide a rare glimpse into early virus evolution.

Several competing theories exist for the emergence of viruses<sup>1</sup>. In the absence of fossils, the study of virus evolution is limited to present-day viruses, which can nevertheless still provide valuable clues. From our exploration of the virosphere some common principles are emerging, even amongst seemingly unrelated viruses. While high-throughput genome sequencing of environmental samples has shown that a significant fraction of newly obtained viral sequences have no homology to any known viral genome<sup>2,3</sup>, comparative analysis of virus structures

allows structure-based phylogenetics. With this approach, several viruses have been grouped into a limited number of viral lineages, where viruses belonging to a certain lineage often infect hosts from entirely different domains of life<sup>4</sup>.

One viral lineage that has emerged as a result of comparative structural virology studies is the PRD1–adenovirus lineage, consisting of several dsDNA viruses infecting hosts ranging from humans<sup>5,6</sup> to eukaryotic unicellular algae<sup>7</sup> as well as bacteriophages (phages)

<sup>1</sup>Institute of Biotechnology, Helsinki Institute of Life Science HiLIFE, University of Helsinki, Helsinki, Finland. <sup>2</sup>Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland. <sup>3</sup>Department of Biological and Environmental Science, Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland. <sup>4</sup>Département de biochimie, de microbiologie, et de bio-informatique, Université Laval, Québec City, Québec, Canada. <sup>5</sup>Present address: Krupic Lab, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK. <sup>6</sup>Present address: The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK. <sup>7</sup>These authors contributed equally: Nejc Kejzar, Elina Laanto, Ilona Rissanen. ✉e-mail: [lotta-riina.sundberg@jyu.fi](mailto:lotta-riina.sundberg@jyu.fi); [juha.huiskonen@helsinki.fi](mailto:juha.huiskonen@helsinki.fi)

infecting Gram-negative<sup>8,9</sup> and Gram-positive<sup>10</sup> bacteria. These dsDNA viruses share the same major capsid protein (MCP) fold and have recently been classified as members of the kingdom *Bamfordvirae*<sup>11</sup>. PM2 (*Corticovirus* PM2, *Corticoviridae* family), a dsDNA phage-infecting strain of *Pseudoalteromonas*, is considered to be the prototypic member of this kingdom due to its minimalistic MCP fold<sup>12</sup>. The pseudo-hexameric MCP in *Bamfordvirae* has recently been proposed to be of cellular origin<sup>13</sup>. Earlier, we reported a phylogenetic connection between phage PM2 and the ssDNA phage FLiP (*Finnlakevirus* FLiP, *Finnlakeviridae*), which infects flavobacteria<sup>14</sup>. At the genomic level, the 9,174-nt-long circular FLiP genome and the 10,079-bp-long circular PM2 genome share only 43% identity. Structurally, however, the two share the same capsid organization, characterized by the triangulation ( $T$ ) number pseudo  $T = 21$  dextro. Furthermore, their MCPs share the same conserved double jelly-roll  $\beta$ -sandwich/ $\beta$ -barrel fold, a hallmark of *Bamfordvirae*<sup>12,14</sup>.

Despite extensive efforts to characterize the structural landscape of the *Bamfordvirae* viral kingdom<sup>6–10,15–21</sup>, structurally similar viruses with an alternative genome type still comprise only one representative, phage FLiP<sup>14</sup>. To increase our confidence in linking the evolutionary origins of ssDNA viruses and dsDNA viruses, more comparative structural studies between viruses harboring different genome types are required. Such analyses are crucial not only to better understand the constraints posed by the structural requirements for virion assembly but also to provide new insights into the early evolution of viruses.

In this study, we report a detailed characterization of  $\Phi$ CjT23, an icosahedral temperate ssDNA phage with an inner lipid membrane, isolated earlier from a strain of *Flavobacterium* (previously *Cytophaga*)<sup>22</sup>. The determined sequence of its relatively short (7642 nt) circular ssDNA genome shows little similarity to any known virus. The MCP sequence-based analysis, however, reveals an abundance of  $\Phi$ CjT23-like prophages in the genomes of several bacterial strains belonging to various *Flavobacterium* species. Cryogenic electron microscopy (cryo-EM) structure of the  $\Phi$ CjT23 virion, determined from purified viral particles shows a capsid organization with triangulation number pseudo  $T = 21$  dextro ( $h = 4$ ,  $k = 1$ ), similar to phages PM2 and FLiP<sup>12,14</sup>. The MCP has a minimalistic architecture in comparison to other homologous MCPs. The capsid vertex architecture is unique and comprises a flexible pentameric spike protein that includes a non-canonical penton base domain, differing from the specialized penton proteins of *Bamfordvirae* members. We conclude by deriving a possible scenario for the evolution of  $\Phi$ CjT23-like ssDNA and dsDNA viruses of *Bamfordvirae* from a shared ancestor with a primordial capsid.

## Results and discussion

### Temperate phage $\Phi$ CjT23 has a circular ssDNA genome

Phage  $\Phi$ CjT23 was investigated as part of our wider efforts to better understand the virosphere of lipid-containing bacteriophages. Phage  $\Phi$ CjT23 was purified from a lysate ( $\sim 1.1 \times 10^{13}$  PFU) using polyethylene glycol precipitation and a two-step gradient centrifugation resulting in  $\sim 6.6 \times 10^8$  PFU of purified virions (yield  $< 0.01\%$ ). The genome composition of  $\Phi$ CjT23 was identified to be ssDNA by degradation of its nucleic acids with DNase and S1 nuclease, while RNase was ineffective. Prior to sequencing, the complementary strand was synthesized with Klenow fragment. Sequencing revealed a circular genome of 7642 nt (30.5% GC) with 15 predicted open-reading frames (ORFs) (Fig. 1; Supplementary Fig. 1). BlastN search returned only one significant result; a highly similar nucleotide sequence detected in *Flavobacterium* sp. KBS0721 (73% query coverage, 89% identity). BlastP search received hits mainly to bacterial gene products of *Flavobacterium* species (Supplementary Data 1). ORF3 received several hits in the GenBank protein database to replication initiation proteins, including the replication initiation factor in *Cellulophaga* phages with ssDNA

genomes (phi18:4, phi12a:1, phi12:2). Indeed, also HHPred search resulted in a significant hit to replication initiation factor protein (rolling circle replication initiator protein from *Geobacillus stearothermophilus*, probability 99.97%,  $E$ -value  $4.9 \times 10^{-30}$ ). Results from the HHPred search for the ORF05 indicated homology with the MCP of *Flavobacterium* phage FLiP (probability 95.5%,  $E$ -value 1.2). In addition, hits showed similarity to the *Pseudoalteromonas* phage PM2 minor membrane proteins P3 (ORF6 in  $\Phi$ CjT23, probability 47.2%,  $E$ -value 28) and P6 (ORF11, probability 45.3%,  $E$ -value 37). While these hits were statistically insignificant, they hinted at homology between  $\Phi$ CjT23, FLiP, and PM2, which was confirmed by comparing the structures of the three phages (see below).

### $\Phi$ CjT23-like prophages detected in *Flavobacterium* genomes

BlastP searches of  $\Phi$ CjT23 ORFs resulted in several hits against *Flavobacterium* genomes suggesting the presence of integrated  $\Phi$ CjT23-like phages, or prophages, in bacterial genomes. In the set of 57 putative prophage sequences (ranging in size from 6021 to 8407 bp; Supplementary Fig. 1; Supplementary Data 2) detected with this method, a gene coding for a replication initiator protein was followed in proximity by a gene coding for a putative MCP. BlastP search among the detected sequences yielded few hits to FLiP genome (Supplementary Fig. 1). These hits were to gp16 which follows the predicted replication protein gp15 in FLiP genome. Similarly, as with  $\Phi$ CjT23 genome, some of the detected sequences received hits to ssDNA phages infecting *Cellulophaga* (Supplementary Fig. 1). Also, a putative *dif* site flanking both ends of these putative prophages was detected. A 28-nt long palindromic sequence was present both downstream and upstream of the prophage sequence, including motifs for *xerC* and *xerD* genes centered by a conserved hexanucleotide (Fig. 1, including also the surrounding six nucleotides). Several ssDNA phages, such as the filamentous *Vibrio* phage  $\Phi$ CTX, use the *dif* site for integration into the host chromosome<sup>23</sup>. These results indicate that  $\Phi$ CjT23-like phages use XerC/XerD mediated recombination for integration into *Flavobacterium* genomes.

### An icosahedral $T = 21$ protein shell encloses a lipid membrane and the genome in the $\Phi$ CjT23 virion

To address distant evolutionary relationships, we determined the structure of the  $\Phi$ CjT23 virion by cryo-EM (Fig. 2; Supplementary Fig. 2). Spherical particles approximately 60 nm in diameter and homogenous in shape were observed in the micrographs (Supplementary Fig. 2a). Occasionally, smaller particles were present, possibly corresponding to particles lacking the capsid and containing only the inner lipid envelope with the genome (Supplementary Fig. 2a, b). Single-particle analysis of complete particles resulted in an icosahedrally symmetric reconstruction with an average estimated resolution of 4.1 Å. Spike proteins extend from the 12 vertices, but the density corresponding to these was disordered in the icosahedrally averaged map (Fig. 2a; Supplementary Fig. 2c) due to the presence of a flexible hinge region in the spike. The basic capsid building block is the pseudo-hexameric MCP trimer (Fig. 2b). The observed 200 MCP trimers, occupying four unique positions in the asymmetric unit, follow icosahedral capsid organization characterized by a pseudo-triangulation number  $T = 21$ . Within the asymmetric unit, type 1 trimers circle the five-fold symmetry axes at the vertices, type 2 trimers reside on both sides of the icosahedral two-fold axis of symmetry, type 3 trimers reside on the icosahedral three-fold axis of symmetry (contributing one MCP per asymmetric unit) and type 4 trimers reside between the rest. This capsid organization is also found in dsDNA phage PM2 and ssDNA phage FLiP<sup>12,14</sup>. The dimensions of the faceted capsid are characterized by a face-to-face distance of  $\sim 55$  nm, an edge-to-edge distance of  $\sim 54$  nm, and a vertex-to-vertex distance of  $\sim 63$  nm. Inside, a lipid bilayer with a thickness of  $\sim 5$  nm is discernible (Fig. 2a; Supplementary Fig. 2c). Weak density hints at minor structural

components under trimers 2 and 4, possibly anchoring the lipid bilayer to the capsid (Fig. 2a).

**The  $\Phi$ CjT23 major capsid protein is a canonical upright double- $\beta$  sandwich**

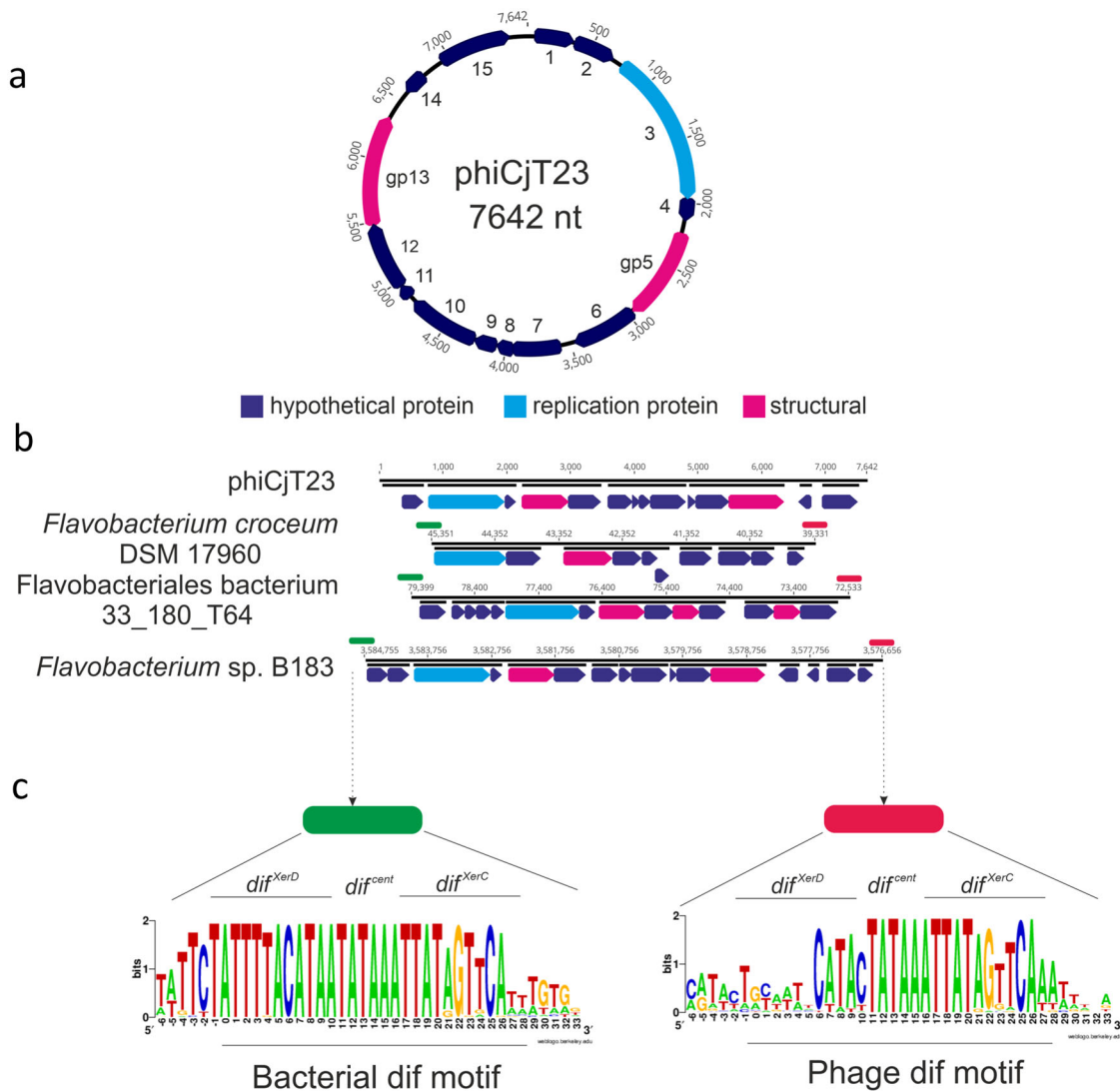
We determined the atomic model of  $\Phi$ CjT23 MCP to compare it with previously determined MCP structures. We used localized reconstruction<sup>24</sup> to determine a density map for each trimer type (see the “Methods” section; Supplementary Fig. 2d, e). This allowed refining local distortions in the capsid structure, improving the resolution of all trimers from 4.1 to 3.2 Å (FSC at 0.143 thresholds; Supplementary Fig. 3a). The improved map quality was sufficient to build atomic models of the trimers de novo (Supplementary Fig. 3b–d). The entire polypeptide sequence corresponding to ORF5 (M1–V239) fitted the density, confirming this ORF as the MCP (P5).

The MCP monomer is composed of two  $\beta$ -sandwiches, V1 (M1–T126) and V2 (N127–V239). Both domains follow a classic upright

jelly-roll topology<sup>5</sup>, each of which is composed of two antiparallel  $\beta$ -sheets composed of four strands, denoted B-I-D-G and C-H-E-F, respectively<sup>25</sup> (Fig. 2d; Supplementary Figs. 3, 4). The  $\beta$ -strands are connected by loops. The BC, DE, FG, and HI loops are facing the capsid exterior, whereas CD, EF, and GH loops, in addition to N- and C-termini, are facing the capsid interior. Unique to  $\Phi$ CjT23, the FG loops (FG1 and FG2) stand upright and seem to stabilize the trimer by interacting across subunit boundaries (Fig. 2e).  $\Phi$ CjT23 MCP also contains two  $\alpha$ -helices, a characteristic of all MCPs of the *Bamfordvirae*. These helices are embedded in the FG loop region in both V1 and V2 domains (termed FG1- $\alpha$  in V1 and FG2- $\alpha$  in V2). FG1- $\alpha$  is lodged between the V1 and V2 domains of each monomer, whereas FG2- $\alpha$  resides between V2 and V1 of the adjacent monomer of the same trimer (Fig. 2c).

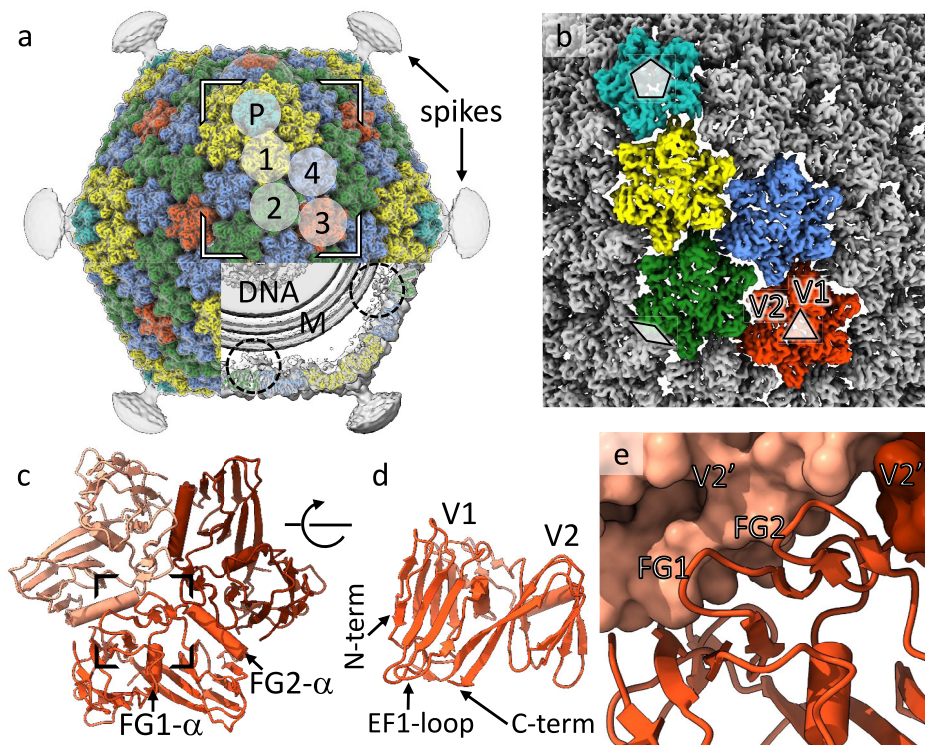
**A simple molecular switch at the MCP trimer–trimer interface**

How are the MCPs bound to each other and the spike penton domain in the pseudo  $T = 21$  capsid? The asymmetric unit contains 10 independent



**Fig. 1 | Genome of a temperate phage  $\phi$ CjT23 and related prophages found in *Flavobacterium* genomes. **a**  $\Phi$ CjT23 genome is a 7642-nt long circular single-stranded DNA molecule with 15 predicted open-reading frames (ORFs). The annotated replication initiation protein is indicated in light blue and the two identified structural proteins, major capsid protein (MCP, gp5) and spike protein (gp13), are in pink. Hypothetical proteins are indicated in dark blue. **b** Examples of putative prophages found in *Flavobacterium* genomes. The lengths of the sequences range between -6000 and 8000 bp. Shared genome synteny between the identified**

prophages included replication initiation protein near the start of the sequence, followed by a gene encoding a major capsid protein MCP (sequence similarity shared between the prophages). **c** Phages and bacterial dif-motifs were identified from the non-coding regions surrounding putative prophages implying XerC/XerD dependent integration. The 28-nt long dif locus (including six extra surrounding nucleotides) with XerC and XerD-binding sites, as well as the central hexanucleotide, are shown.



**Fig. 2 | Structure of the  $\Phi$ CjT23 bacteriophage capsid and major capsid protein trimer.** **a** Surface rendering of the  $\Phi$ CjT23 virion cryo-EM map at 4.1 Å resolution. The four different types of major capsid protein trimers are colored yellow (1), green (2), red (3) and blue (4). The penton base (P) is colored cyan with the remaining structural components colored gray. On top, a gray transparent surface, calculated from the cryo-EM map filtered to 12 Å resolution, is shown. This lower-resolution surface reveals disordered density attributed to flexible spikes which originate from the penton bases. A part of the density is removed, to reveal the lipid bilayer (M), DNA, and putative minor protein components (circles) bridging MCPs to the lipid bilayer. **b** A close-up of the high-resolution surface from the area indicated in **a** and rotated to face the viewer. The icosahedral axes of symmetry are

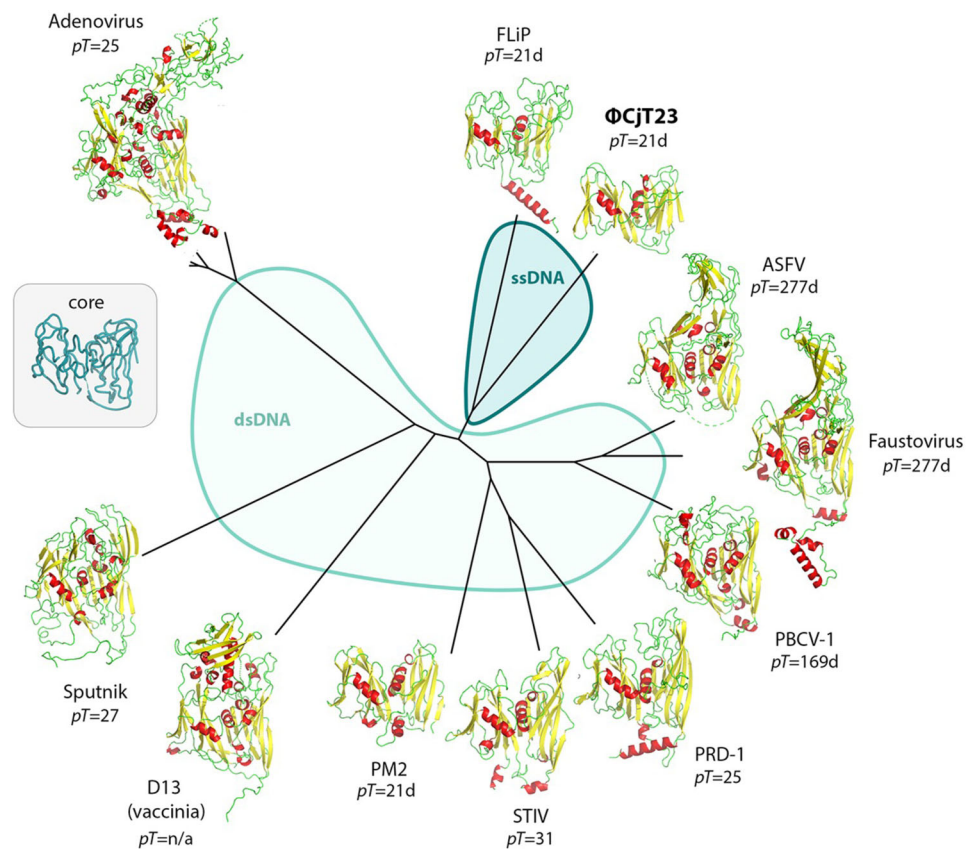
marked with a pentagon (five-fold axis), triangle (three-fold axis), and diamond (two-fold axis). Type 1 trimers are peripentonal, circling the five-folds. Type 2 trimers are adjacent to the two folds. Type 3 trimers locate on the three-folds. Type 4 trimers are farthest from any symmetry axes. **c** Model for one trimer (type 3) is shown as a ribbon. The three different subunits are shown in different shades. **d** The view in **c** is shown after rotating 90° around the horizontal axis as indicated. Only the frontmost monomer is shown. The two  $\beta$ -sandwich domains (V1 and V2) are labeled, in addition to the protein termini. **e** Closeup of the area indicated in **c** show how the two upright loops (FG1 and FG2) lock the monomer with the V2 domains (V2' and V2'') of its neighbors (shown as surfaces).

MCP monomers, contributing a total of 20  $\beta$ -sandwiches (V1 and V2). The penton domain of the spike presents one monomer per asymmetric unit and constitutes an additional building block (Fig. 2b). As these 21 building blocks are not identical in sequence (MCP V1, MCP V2, and penton domain of the spike protein), the triangulation is referred to as 'pseudo'. Furthermore, as these 21 building blocks are not related by icosahedral symmetry, they all exhibit a slightly different, or 'quasi-equivalent', environment, which in theory should be manifested as structural differences<sup>26</sup>. The overall structures of the individual chains were, however, highly similar (Supplementary Fig. 5a–c), showing a mean RMSD value of  $0.39 \pm 0.10$  Å (monomers from trimers 1, 2 and 4 relative to a monomer from trimer 3 with three identical chains). Notably, the structural motifs creating intra-trimer contacts (FG1-loop, FG2-loop, and FG2- $\alpha$ ) showed very little variation (Supplementary Fig. 5b). Interestingly, most variation existed in the EF1 and GH1 loops (Supplementary Fig. 5c). These loops create inter-trimer contacts, thus acting as molecular switches, and adopt slightly different conformations to accommodate each trimer in their local environment. This is the case, especially for loop EF1 which displayed three discrete conformations named 'outward', 'middle', and 'inward' (Supplementary Fig. 5d). The outward-facing EF1-loop is used by trimer 1 to reach under the spike penton domain, and analogously by trimer 2 to reach under trimer 1 (Supplementary Fig. 5e, f). The inward-facing EF1-loops, in turn, exist solely at the edges of the icosahedral face where the trimers meet the

differently tilted trimers of the adjacent face. Finally, the rest of the trimer–trimer interactions are mediated by EF1-loops in the middle position. These observations suggest that a simple structural element alone, such as a loop between two  $\beta$ -strands exhibiting only a few discrete conformational states, can provide an elegant solution for the 'quasi-equivalence' problem in the capsid assembly.

### Comparative analysis highlights the minimalistic nature of $\Phi$ CjT23 MCP

The MCP (P5) of  $\Phi$ CjT23 contains only 239 residues, making it the smallest known MCP with a double jelly-roll fold. The second smallest MCP (269 residues) is that of Pseudoalteromonas phage PM2, considered to be the prototypic member of the *Bamfordvirae*<sup>27</sup>. To compare the tertiary structure of the  $\Phi$ CjT23 MCP to all known double jelly-roll MCP structures, we constructed a structure-based phylogeny (Fig. 3)<sup>27</sup>.  $\Phi$ CjT23 MCP, in addition to MCPs of FLiP, PM2, and PRD1, lacks the elaborate decorations created by long loops and other structural elements on top of most other MCPs. The mean of RMSDs between the common core of the  $\Phi$ CjT23 P5 fold (183 residues) and all analyzed MCPs was 4.4 Å. The closest structural homolog was FLiP MCP with an RMSD of 3.0 Å. This phylogeny agrees with the genome type of these viruses as  $\Phi$ CjT23 and FLiP are the only known ssDNA viruses similar to the members of *Bamfordvirae*, which currently consist exclusively of dsDNA viruses.



**Fig. 3 | Structural phylogeny of major capsid proteins with double jelly-roll fold.** The two different parts of the phylogeny, corresponding to viruses harboring either an ssDNA (blue) or dsDNA (light blue) genome, are shown in different colors.

A generalized common core of the fold determined by the Homologous Structure Finder software and used in the comparisons is shown as a  $C_{\alpha}$ -trace in the inset. The T-number for the corresponding icosahedral capsid, when applicable, is given.

In all members of *Bamfordvirae*, the cores of the two  $\beta$ -sandwiches (V1 and V2) in the MCPs are similar, indicative of a gene duplication<sup>4,8,12</sup>. This similarity is especially evident in PM2,  $\Phi$ CjT23, and FLiP, where the structural alignment of V1 and V2 reports RMSD of 2.0, 2.3, and 3.4 Å, over 116, 88, and 105 matching residues, respectively. Further examination of the individual  $\beta$ -sandwich structures and their topologies highlights the minimalistic nature of  $\Phi$ CjT23 MCP (Supplementary Fig. 4). Even the simplest double jelly-roll MCPs reported earlier show some divergence between V1 and V2 in terms of additional structural elements. For instance, the MCPs of PM2, FLiP, and PRD1 have acquired an additional  $\alpha$ -helix, which is membrane-proximal, but can locate either in the N-terminus (PRD1), in a membrane-facing loop (GH-loop of V1 in PM2), or in the C-terminus (FLiP). In PRD1 this helix, in addition to the C-terminus, has been reported to act as a conformational switch. Such an additional helix is absent in  $\Phi$ CjT23, where the EF1-loop seems to perform the switching function alone. Taken together, these results suggest that V1 and V2 domains of  $\Phi$ CjT23 have diverged relatively little since the presumed gene duplication<sup>4,8,12</sup> and  $\Phi$ CjT23 MCP may thus serve as a model for an ancestral MCP present in the last common ancestor of  $\Phi$ CjT23 and members of *Bamfordvirae*.

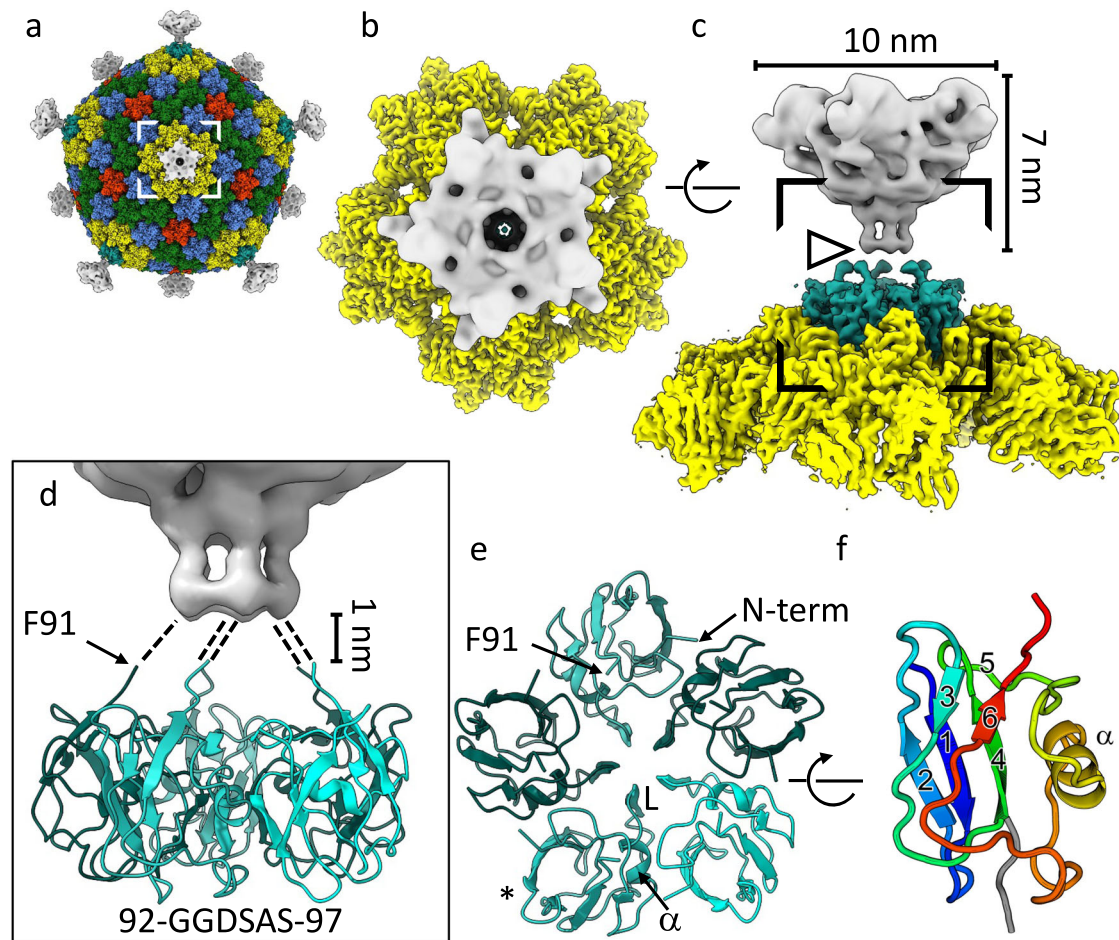
### The $\Phi$ CjT23 pentameric viral spike has a unique architecture

The clustering of  $\Phi$ CjT23 and FLiP MCPs in the structural phylogeny prompted us to investigate whether the structural homology also extends to the vertex complexes of these viruses. First, we analyzed the structure of the  $\Phi$ CjT23 spike. Localized reconstruction<sup>24</sup> combined with symmetry relaxation<sup>28</sup> allowed partial tackling with the observed high degree of spike flexibility by determining the orientations of individual spikes relative to the capsids (Fig. 4a,

Supplementary Fig. 2c, f; see the “Methods” section). This led to a 5.3-Å resolution structure of the capsid external part of the spike (Fig. 4a, b). Similar to the previously published low-resolution structure of the FLiP spike<sup>14</sup>, the  $\Phi$ CjT23 spike has a five-fold symmetry (Fig. 4b). At the resolutions reached so far, the protein folds remained elusive, impeding a more detailed comparison. The dimensions and shapes of the two spikes, however, appear different (Supplementary Fig. 2g)<sup>14</sup>.

In the absence of atomic models for the capsid-external parts of the spikes, we turned our focus to their capsid-internal parts instead. Following the approach applied to the MCPs, we calculated a localized reconstruction of the internal part of  $\Phi$ CjT23 spike (penton domain), in addition to the FLiP penton base from our previously published data<sup>14</sup>. The resolutions of the localized reconstructions for  $\Phi$ CjT23 and FLiP were 3.4 and 4.0 Å, respectively (Supplementary Fig. 6). This was sufficient to build the corresponding atomic models de novo. As expected, the structure of the FLiP penton revealed a pentamer formed by five copies of protein P12 (encoded by *orf12*) with the canonical  $\beta$ -sandwich fold, similar to penton base proteins of PM2 and PRD1 (Supplementary Fig. 7). This confirmed that the FLiP spike is formed by a separate protein forming a pentamer sitting atop the capsid integral penton base.

In contrast to FLiP, the entire spike of  $\Phi$ CjT23 is made by one protein, P13 (283 residues) (Fig. 1), which forms both the capsid-internal and capsid-external parts, analogously to the spike protein of phage PM2<sup>12</sup>. The capsid-integral part of the spike, corresponding to the P13 N-terminal penton base domain (residues M1–F91), sits in a slightly raised position relative to the type 1 MCP trimers circling it (Fig. 4c). The last residues visible in the cryo-EM density are P90 and F91, residing just prior to the disordered part in the map (Fig. 4d). The



**Fig. 4** | **ΦCjT23 vertex.** **a** A composite cryo-EM map of ΦCjT23 bacteriophage capsid combining localized reconstructions of the four MCP trimers (yellow, green, red, and blue), penton base (cyan), and spike (gray). **b** A close-up of one vertex seen from the top of the area indicated in **a**. **c** The same view as in **b**, rotated by 90° around the axis indicated. The dimensions of the spike, corresponding to the external domains of spike protein P13, are indicated. Note the discontinuation in the density (arrowhead). **d** A close-up of the area indicated in **c**. The model of the penton base part of the spike is shown in ribbon (cyan). The ordered part of the map ends around the area occupied by G92. Putative linker regions spanning the 1-nm gap between the penton base and the external part of the spike (gray) are

illustrated with dashed lines. The sequence of the putative linker is given. **e** A model of the pentamer formed by the penton base domain seen along a 5-fold axis of symmetry. Different monomers are colored with different shades of cyan. The positions of the N-terminus, C-terminal G92, α-helix at the monomer–monomer interface (α) and loop (L) closing the oculus in the middle of the pentamer are labeled. **f** The monomer is colored from blue (N-terminus) to red (C-terminal G92) and marked with an asterisk in **e** is shown rotated by 90° around the axis indicated. The β-strands are numbered 1–6. Strands 2–1–4 make a β-sheet. The α-helix, residing in the region connecting strands 5 and 6, is labeled.

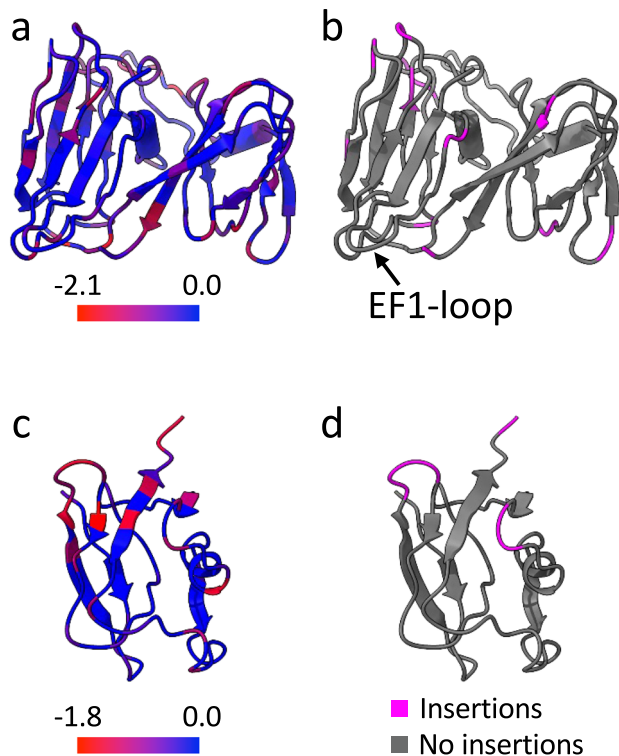
sequence immediately following P90 and F91 contains two glycine residues and constitutes a flexible hinge that allows a range of orientations of the external part of the spike relative to the capsid. The penton base domain is a small β-barrel (SBB) with a β-strand topology that differs from the topologies of better-characterized SBBs<sup>29</sup>, and from the canonical β-sandwich fold observed in pentons of other viruses such as FLiP and *Bamfordvirae* members. The ΦCjT23 spike SBB is formed by anti-parallel β-strands β1, β2, and β4, and additional β-strands β3, β5, and β6. The fold encloses a hydrophobic core, characteristic of SBBs<sup>29</sup>. The region connecting strands β5 and β6 include a short α-helix (S55–D64) and a long loop (L65–F78). The α-helix resides at the monomer–monomer interface. This is analogous to the MCPs where the α-helices reside between the V1 and V2 domains of the monomer, in addition to monomer–monomer interfaces. Part of the loop (L65–N71) faces toward the central cavity of the pentamer. Together the five loops close an oculus in the middle of the pentamer (Fig. 4e). Taken together, these results show that the ΦCjT23 spike has a unique architecture with a non-canonical penton domain fold.

As the spikes of *Bamfordvirae* members generally mediate host cell recognition<sup>30,31</sup>, we studied this further by testing the adsorption

and infectivity of ΦCjT23 against three host strains of FLiP (*Flavobacterium* sp. B330, B114, and B167), in addition to the type strain *F. johnsoniae* UW-101. ΦCjT23 infected strains B114 and B167, but not the other two strains tested (B330 and UW-101). Adsorption assays with ΦCjT23 showed similar adsorption patterns in all the strains (including the non-hosts B330 and UW-101) with approximately 35% of phages adsorbed after 5 min and the number remaining constant until the end of the experiment (15 min, Supplementary Fig. 8). However, adsorption to cells of *Flavobacterium* sp. B114 was not observed. It should be noted that because ΦCjT23 can infect this strain, the adsorption step likely requires different conditions than those used in our experiments. In conclusion, ΦCjT23 can adsorb to geographically distant *Flavobacterium* strains, including some of the host strains of its closest relative FLiP, in addition to infecting some of them.

#### Mapping of prophage sequence variation on the ΦCjT23 structures

Have ΦCjT23-like prophages detected in *Flavobacterium* genomes accumulated mutations that would be likely incompatible with capsid assembly? To address this question, we mapped the sequence



**Fig. 5 | Mapping of  $\Phi$ CjT23-like prophage sequence variation.** **a** Sequence conservation<sup>72</sup> (score -2.12–2.26) determined by multiple sequence alignment of  $\Phi$ CjT23-like prophage sequences ( $N = 21$ ) and  $\Phi$ CjT23 major capsid protein (MCP) mapped per residue on the  $\Phi$ CjT23 MCP structure. **b** The same view as in **a** with residues surrounding insertions highlighted in pink. **c** Sequence conservation (score -1.84–2.36) determined by multiple sequence alignment of  $\Phi$ CjT23-like prophage sequences ( $N = 8$ ) and  $\Phi$ CjT23 spike protein N-terminal domain. **d** The same view as in **c** with residues surrounding insertions highlighted.

variation in the homologous prophage ORFs on the structures of the  $\Phi$ CjT23 MCP and spike N-terminal penton domain (Fig. 5). In both cases, the secondary structure elements are relatively conserved with the greatest sequence variation localizing to loops (Fig. 5a, c). Furthermore, insertions that were present in some of the prophage sequences mapped exclusively to the loop areas (Fig. 5b, d). A notable exception is the EF1-loop that acts as a conformational switch; little sequence variation and no insertions were observed in this loop (Fig. 5a, b).

The presence of both MCP and spike protein sequences in  $\Phi$ CjT23-like prophages (Supplementary Fig. 1) suggests that these can in principle assemble particles with the same mechanism as  $\Phi$ CjT23 and the sequence variation present in the prophages is unlikely to impede particle assembly. It is possible that the last common ancestor of  $\Phi$ CjT23 and FLiP utilized the canonical penton fold, which itself has been proposed to be homologous to the MCP domain<sup>9</sup>, and that  $\Phi$ CjT23 has lost this structural component. Alternatively,  $\Phi$ CjT23 may be more similar to an ancestral virus that lacked the canonical penton base. Finally, we would like to note that particle assembly from  $\Phi$ CjT23-like prophages remains to be verified in future studies.

### Pseudo-hexameric MCPs constrain possible capsid architectures

Recent evidence supports a scenario, where the double jelly-roll (DJR) fold originated by gene duplication in a cellular organism predating viral DJR-MCPs, and trimerization led to the basic, hexagon-shaped building block of a primordial capsid<sup>13</sup>. What is the simplest possible viral capsid built of such pseudo-hexameric MCPs? Classification of different icosahedrally symmetric capsid types based on their

$T$ -number allows us to systematically consider different structural constraints (Fig. 6). From a geometric point of view, one constraint is that the MCP is a trimer and thus cannot occupy the two-fold symmetric position. This rules out several capsid types that would otherwise be possible ( $T = 4, 12, 16, 28, \dots$ ). Another constraint is the hexamer complexity ( $C^h$ )<sup>32</sup>. Capsids with hexameric or pseudo-hexameric building blocks with  $C^h$  larger than two ( $T = 12, 19, 27, \dots$ ) are under-represented in the observed capsid structures and have been proposed to be disfavored in evolution. A notable exception is the large satellite virus Sputnik, with a  $T = 27$  capsid and  $C^h = 4$  (ref. 17). Considering these constraints, the smallest allowed—and possibly favored— $T$ -numbers are thus 3, 7, 9, 13, 21, and 25. Of these,  $T = 21$  is the smallest that has so far been observed in three cases, namely  $\Phi$ CjT23 in this study, in addition to FLiP<sup>14</sup> and PM2<sup>16</sup>, whereas  $T = 25$  has been observed in PRD1<sup>33</sup>, Bam35<sup>34</sup>, and adenovirus<sup>35</sup>. The estimated volume of a  $T = 13$  shell is approximately 57% of the volume of the  $T = 21$  shell (Fig. 6). It is conceivable that among capsids built with a double jelly-roll fold,  $T = 13$  and smaller capsids would be too small to enclose a lipid bilayer and a genome with sufficient length to encode the two required structural proteins and other factors necessary for infection and assembly. Together, these considerations could explain why  $T = 21$  is the smallest  $T$  observed for these capsid types. Increased sampling of the virosphere may discover double jelly-roll fold-based capsids with smaller  $T$  than 21 but it remains to be seen if such capsids could encompass a lipid bilayer.

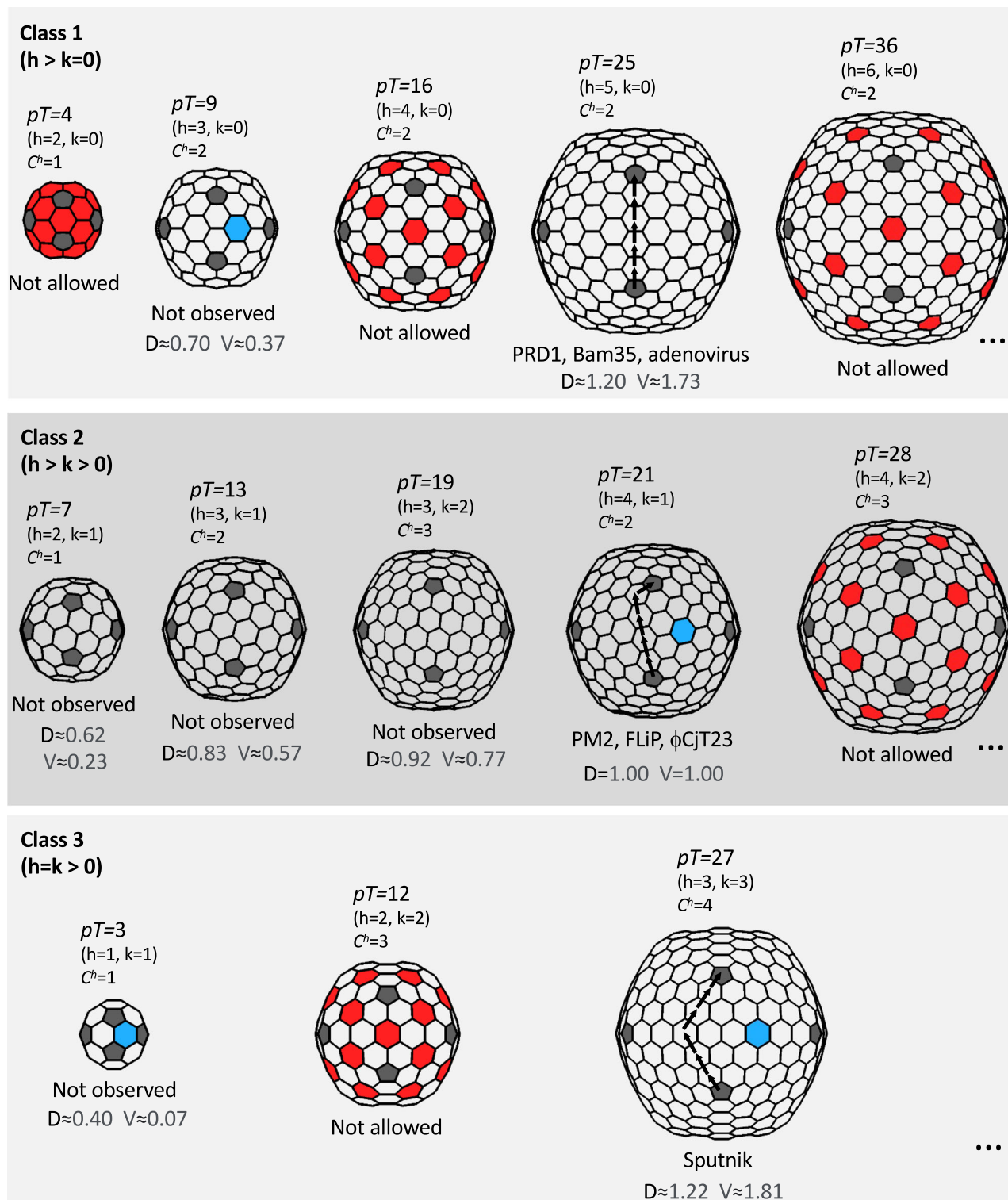
### Model for capsid assembly and spike swapping

Summarizing the structure of the prototypic  $\Phi$ CjT23 and analysis of  $\Phi$ CjT23-like elements in bacterial genomes allow us to propose a model for capsid assembly and evolutionary mechanisms for spike swapping (Fig. 7). First, a  $pT = 21$  capsid has a trimer at the three-fold axis of symmetry, an arrangement that allows MCP-driven capsid assembly to proceed from this point in a simple, symmetric fashion (Fig. 7a). Such an assembly pathway would not require minor proteins to modulate the conformation of the MCP or tape-measure proteins to define the dimensions of the capsid faces<sup>9</sup>. Instead, the only prerequisites would be a linkage to the membrane (which in turn provides a scaffold for different faces of the capsid to bind to as well as creates the required curvature to form a closed capsid) and a conformational switch in the MCP itself allowing it to lock in place in different positions. Second, pentameric spikes are required to fill the five-fold positions via interactions with peripentonal MCPs. These interactions ensure that the spikes, which presumably carry out essential receptor binding functions, are incorporated into the assembly.

The proposed capsid assembly mechanism applies to both  $\Phi$ CjT23 and PM2. An alternative scenario, where the assembly starts from the vertices, or a special vertex that is structurally distinct from the rest, is also plausible. However, the unique fold of the  $\Phi$ CjT23 spike penton domain suggests that this structural element is not central to the assembly of these relatively simple  $T = 21$  capsids. Other viruses utilizing the double jelly-roll fold have evolved complex capsid architectures requiring minor proteins and more elaborate spike assemblies. For instance, the spikes of adenovirus<sup>36</sup> and FLiP, are formed by a protein separate from the penton base. Such separate spikes may be more readily swapped to other proteins than those spikes which are formed of a single protein (Fig. 7b). These viruses may have gained entry to new hosts by this proposed mechanism of spike swapping, thereby increasing their fitness. This notion is supported by the plethora of different spike architectures observed in members of *Bamfordvirae*. PRD1 and its close relative PR772 present even more complex vertex architectures. In these viruses, a trimeric spike binds to the capsid by forming a heteropentameric penton base with a penton protein. A monomeric second spike is attached to this penton<sup>37,38</sup>.

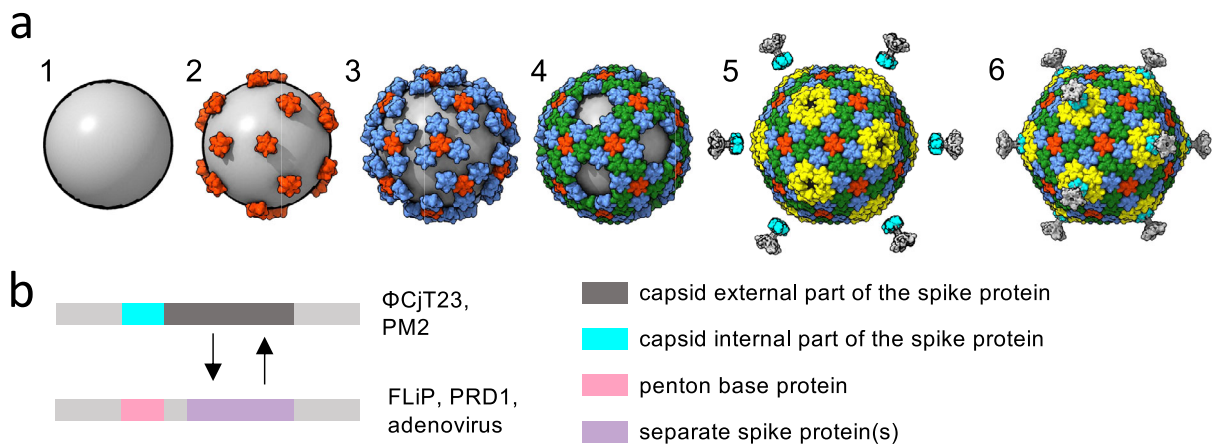
In conclusion, this study, together with our earlier findings on FLiP<sup>14</sup>, solidifies a link between dsDNA and ssDNA viruses with capsids





**Fig. 6 | Icosahedral virus capsid architectures.** A schematic diagram is shown for different types of icosahedrally symmetric capsids built of 12 pentagons (gray) and  $N = 60 \times (T-1)/6$  hexagons (white). The triangulation ( $T$ ) number of each capsid is defined by two lattice indexes  $h$  and  $k$  as  $T = h^2 + h \times k + k^2$ . In the members of the PRD1-adenovirus lineage (*Bamfordvirae*), the allowed hexagon positions are occupied by pseudo-hexameric trimers, hence the  $T$ -number is indicated being pseudo ( $p$ ). Different capsid types are classified in classes 1, 2, and 3 (ref. 32). Only the first members of the series (...) are shown in each class. The lattice axes are illustrated for one capsid in each class with arrows. Hexagons that are located on the icosahedral two-fold axes of symmetry, and that thus cannot be occupied by a

pseudo-hexameric, trimeric major capsid protein (MCP), are colored in red (corresponding capsids are labeled as “not allowed”). For each capsid, the value of hexamer complexity ( $C^h$ ) is given<sup>32</sup>. Some names of the characterized viruses with a certain capsid type ( $pT = 21, 25, 27$ ) are given. Capsid geometries with no characterized representatives are labeled as unobserved. Note that the lattice in class 2 capsids is handed and only the right-handed (dextro) organization, which is observed in PM2, FLiP, and  $\phi$ CjT23, is shown. Allowed capsid geometries ( $pT = 3, pT = 9, pT = 21$  and  $pT = 27$ ) where a hexagon occupies the three-fold axes of symmetry are indicated by coloring one such hexagon blue.  $D$  and  $V$  denote the relative diameter and volume, respectively, relative to the  $pT = 21$  shell.



**Fig. 7 | Models for capsid assembly and spike swapping.** **a** A hypothetical model for the assembly of  $\Phi$ CjT23 and PM2-like viruses with  $T=21$  capsids. First, a membrane vesicle wrapping around the genome assembles to create a lipid core (1). Major capsid protein (MCP) trimers attach to the lipid core via minor membrane proteins (2). More MCPs can be incorporated in a symmetrical fashion via MCP–MCP interactions (3,4) until a nearly closed shell is formed (5). Finally, spikes

are incorporated into the capsid (6). **b** The genome region encoding for the spikes is depicted in gray. Genetic changes (insertions and deletions) resulting in swapping first the complete spike (such as those observed in  $\Phi$ CjT23 and PM2) with a penton base and then additional spike proteins being incorporated (such as those observed in FLiP and adenovirus) are depicted.

utilizing trimeric building blocks with the double jelly-roll fold, a hallmark of the *Bamfordvirae* viral kingdom. No structures of RNA viruses with similar capsid protein folds have been reported so far. It is thus possible that the last common ancestor of  $\Phi$ CjT23 and members of the *Bamfordvirae* arose from DNA-based genetic elements, separately from RNA viruses, and co-evolved with cellular life leading to increasingly elaborate variations of the same basic structural principles. Both phages FLiP and  $\Phi$ CjT23 parasitize *Flavobacterium* species, which are commonly found in both aquatic and terrestrial environments<sup>39–41</sup> as well as in ancient permafrost<sup>42</sup> and glacier samples<sup>43</sup>. Furthermore, our empirical data show that  $\Phi$ CjT23 adsorbs to several *Flavobacterium* strains, some of which are hosts to FLiP. As ssDNA phages are more prevalent in the environment than often acknowledged<sup>44–46</sup>, further studies are required to establish if similar types of membrane-containing ssDNA phages infect members of other bacterial genera. Understanding the minimal requirements for capsid assembly from the simplest possible structural elements in this group of viruses may open avenues for the engineering of membrane-containing virus-like particles for different applications.

## Methods

### Bacteria and phage

Bacterial strain *Flavobacterium johnsoniae* HER1326 (UW101-36) and phage  $\Phi$ CjT23 (HER326) were obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses ([www.phage.ulaval.ca](http://www.phage.ulaval.ca)). *Flavobacterium* sp strain B183 was originally isolated from a freshwater sample in Finland<sup>47</sup>. Bacteria were grown at 25 °C in constant agitation at 150 rpm while high tryptone cytophaga (HTC) medium and 20% Lysogenic Broth (LB) were used as medium. For phage propagation, HTC medium was supplemented with 3 mM MgCl<sub>2</sub> and 10 mM CaCl<sub>2</sub>. For solid medium, 1% agar was used while 0.5% Gelrite was used when phage plaque formation was required. The host range of  $\Phi$ CjT23 was tested against the known hosts of the ssDNA phage FLiP (*Flavobacterium* sp strains B114, B330, and B167) in addition to *F. johnsoniae* UW101.

### Phage adsorption assay

Adsorption of  $\Phi$ CjT23 to the cells of five *Flavobacterium* strains (the host strain HER1326, *F. johnsoniae* UW101, and *Flavobacterium* sp. strains B114, B167, and B330) was tested by adapting an earlier

method<sup>48</sup>. In brief, bacteria were grown to mid-log phase and diluted to OD<sub>595</sub> ~ 0.1 with HTC (supplemented with 3 mM MgCl<sub>2</sub> and 10 mM CaCl<sub>2</sub>). These were divided into triplicates (9 ml in 125 ml Erlenmeyer flasks) and 1 ml of  $\Phi$ CjT23 ( $\sim 2.4 \times 10^5$  PFU) was added. A growth medium without bacteria was used as control. Samples of 50  $\mu$ l were taken into pre-chilled tubes containing 950  $\mu$ l medium and filtered (0.45  $\mu$ m). From HER1326 samples were taken at time points 2, 5, 7, 10, and 15 min and from others at 5, 10, and 15 min, and 100  $\mu$ l of the sample was plated with 500  $\mu$ l of -1/5 diluted HER1326 on HTC Gelrite plates.

### Purification of virions

Phages were propagated by transferring the o/n grown bacterial host HER1326 to a fresh HTC medium (supplemented with 3 mM MgCl<sub>2</sub> and 10 mM CaCl<sub>2</sub>) at a 1:10 ratio and adding a fresh lysate of  $\Phi$ CjT23 at ratios between 1:100 and 1:200. After six hours of incubation (25 °C, 150 rpm), cells were pelleted by centrifugation (Sorvall RC-6+ F9 6 $\times$ 1000 rotor, 5400 $\times$ g, 30 min, 4 °C) and the supernatant was filtered through bottle top filter (0.22  $\mu$ m, Nalgene). Phages were precipitated by dissolving NaCl (0.5 M) and PEG6000 (10%) to the supernatant and pelleting (Sorvall RC-6+, F9 6 $\times$ 1000 rotor, 5400 $\times$ g, 45 min, 4 °C). The resulting pellet was dissolved in 0.02 M Tris–HCl, 3 mM MgCl<sub>2</sub>, and 10 mM CaCl<sub>2</sub> pH 7.2. Rate zonal and equilibrium centrifugations were used for subsequent purifications. First, phage suspension was layered on top of a 5–20% sucrose (w/vol, in Tris–HCl pH 7.2 with 3 mM MgCl<sub>2</sub> and 10 mM CaCl<sub>2</sub>) and centrifuged (Beckman Optima L-K90, SW-28 rotor, 80,000 $\times$ g, 45 min, 15 °C). The light-scattering zone was collected and further purified in 20–70% sucrose gradient (w/vol, in Tris–HCl pH 7.2 with 3 mM MgCl<sub>2</sub> and 10 mM CaCl<sub>2</sub>) by centrifugation (Beckman Optima L-K90, SW-41 rotor, 111,000 $\times$ g for 18 h, 4 °C). The light-scattering zones were collected and pelleted by centrifugation (Beckman Optima L-K90, 70 Ti rotor, 76,000 $\times$ g for 3 h, 4 °C).

### Analysis of purified virions

Titers (PFU/ml) of the phage suspensions after each purification step were determined. The protein concentrations of the pelleted light-scattering bands from rate zonal and equilibrium centrifugations were determined using the Bradford assay<sup>49</sup>. Structural proteins of the purified virions from the rate of zonal centrifugation were resolved by 17% Tricine SDS–PAGE. The gel was first dyed with Coomassie brilliant blue to detect the proteins after which Sudan Black dye was used for

indicating putative lipid structures. Purified virions of phage PRD1 were used for comparison.

### Phage genome isolation, sequencing, and bioinformatics

Phage genomic DNA was extracted from a filtered phage lysate using a  $ZnCl_2$  precipitation method<sup>50</sup> with slight modifications. The complementary strand was synthesized using hexamer primers and Klenow fragments. Sequencing was done with Illumina MiSeq sequencer using a paired-end (600 bp) kit at the Institute of Biotechnology, University of Helsinki. The assembled genome was further sequenced using Sanger sequencing to confirm uncertain bases. Putative open reading frames were predicted using Glimmer<sup>51</sup> and GeneMarkS<sup>52</sup>. The NCBI programs BLASTp and DELTA-BLAST against the nonredundant GenBank protein database were used for identifying homologous proteins to the translated ORF sequences<sup>53</sup>. In addition, HHPred search<sup>54</sup> was employed also for predictions of transmembrane segments.

### Bacterial genome sequencing and analysis

*Flavobacterium* sp B183 Genomic DNA was purified using a QIAGEN Genomic-tip 20/G kit, according to manufacturer instructions. Single-molecule real-time sequencing was performed on a PacBio RSII sequencer (G enome Qu ebec Innovation Centre) and assembly was done using the HGAP workflow<sup>55</sup>.

### Prophage identification

BLASTp search for most of the open reading frames predicted in  $\Phi$ CjT23 genome received several hits to *Flavobacterium* genomes. The resulting hits and their surrounding genome regions were investigated in more detail. In the case of hits to both replication initiation factor containing protein and a putative major capsid protein (hits to ORF that received a hit to FLiP MCP in HHPred search), the regions were subjected to a detailed analysis for identification of integration sequences. WebLogo (available at [weblogo.berkeley.edu](http://weblogo.berkeley.edu)) was used to create sequence logos for the identified conserved areas at both of the prophage genomes ends. Selected 38 prophage sequences filling the above-mentioned criteria were aligned using MAFFT<sup>56</sup>. Some of the prophage sequences were similar (above ~80% identity) thus one sequence was selected as a representative of the group for further analysis (Supplementary Data 2). Easyfig<sup>57</sup> and BLASTp algorithm were used to create a comparison between the prophage sequences (Supplementary Fig. 1).

### Cryogenic electron microscopy of virions

Copper grids coated with holey carbon foil (Quantifoil Cu 1.2/1.3, 300 mesh) were glow-discharged for 30 s at 0.45 Torr. A 3- $\mu$ L aliquot of the purified virions was applied to the grids and after 30 s blotted for 4.5 s and plunge-frozen in liquid ethane with a vitrification apparatus (Vitrobot Mark IV, Thermo Fisher Scientific) equilibrated to 6 °C and 100% relative humidity. The grids were imaged using Talos Arctic transmission electron microscope, operated at 200 kV, and equipped with a Falcon III direct electron detector (Thermo Fisher Scientific) in counting mode. Data were collected using EPU software ver. 1.20.3.10 (Thermo Fisher Scientific). A total of 1023 movies were collected (30 frames per movie) with an electron dose of  $0.5 \text{ e}^-/\text{Å}^2/\text{frame}$ , resulting in a total dose of  $\sim 15 \text{ e}^-/\text{Å}^2$ . A nominal magnification of 120,000 $\times$  was used, which resulted in a calibrated pixel size of 1.24 Å. A second data set (852 movies) was collected at a 20° stage tilt angle using the same data acquisition parameters as for the first dataset. The two datasets were combined for processing. Data acquisition parameters are summarized in Supplementary Table 1.

### Single-particle data processing and refinement

Data processing was performed using RELION 3.0<sup>58</sup> as part of Scipion software framework (version 3.0)<sup>59</sup>. Movie frames were aligned with

MotionCorr<sup>60</sup>. CTF was estimated with GCTF<sup>61</sup> (first data set) and CTFFIND4<sup>62</sup> (second data set). The defocus values ranged from 0.1 to 2.2  $\mu$ m. The particles were picked with ETHAN picker<sup>63</sup>, using a search radius of 240 pixels (first data set), or CrYOLO<sup>64</sup> using a manually picked set of 301 particles as training data (second data set). Particles (total of 8862) were extracted in a box of 800  $\times$  800 pixels. A subset of particles (6501) was selected for further processing after 2D and 3D classification. The previously determined structure of the FLiP virion (EMD-3771), filtered to 25 Å resolution to reduce template bias, was used as an initial volume. The final map was reconstructed by running RELION's Bayesian particle polishing, CTF refinement (per-particle defocus and beam tilt), and gold standard 3D auto-refinement applying icosahedral symmetry. A mask defining the capsid shell with a 12-pixel soft edge was created and the resolution was estimated using Fourier shell correlation (FSC) with 0.143 thresholds.

### Localized reconstructions of the major capsid protein trimers

To further improve the resolution of the icosahedrally symmetric reconstruction of the capsid, the localized reconstruction method was applied on the four different types of MCP trimers (1–4) to take into consideration the defocus gradient and possible flexibility of the capsid<sup>24,65</sup>. The positions of the trimers were manually defined in UCSF Chimera by placing a marker in the center of each trimer in one asymmetric unit. A total of 390,060 sub-particles (60 for each capsid) were created for trimers 1–3. Trimer 3 position, which is on the icosahedral 3-fold axis of symmetry, yielded 130,020 sub-particles (20 for each capsid). The sub-particles were aligned by placing the local three-fold symmetry axis of each trimer on the Z-axis. Sub-particles were extracted in a box of 200  $\times$  200 pixels and subjected to one round of 3D classification without alignment, before the gold standard, local 3D auto-refinement in RELION. In local refinement, shifts of  $\pm 3$  pixels were considered with an initial sampling of 1 pixel. Initial angular sampling in the local alignment was 0.9°. Three-fold symmetry was applied on trimer 3. No symmetry was applied on trimers 1, 2 and 4. A mask with a 3-pixel soft edge was created for each trimer localized reconstruction and the resolution was estimated as above.

### Localized reconstruction of the penton

Similar to the trimer localized reconstructions, we separately reconstructed the penton base. A total of 78,012 penton sub-particles (12 for each capsid) were defined, aligned in C5 symmetry convention, extracted, and refined as above. Five-fold symmetry was applied. The map was masked, and its resolution was estimated as above.

### Localized reconstruction of the spike

To facilitate localized reconstruction of the spikes, the contribution of the icosahedral capsid was subtracted from the particle images. A mask defining the capsid and its interior was created by a thresholding operation, applied on a lowpass filtered (15 Å) version of the capsid reconstruction, followed by dilation and erode operations to fill holes in the mask. Partial signal subtraction of the capsid was performed in RELION<sup>66</sup> considering the CTF of each particle. Sub-particles (78,012) were defined at a radius of 315 Å as above and extracted from the capsid-subtracted images. To find the tilt direction of each spike, particles were aligned using “*relax\_sym C5*”<sup>28</sup> option in RELION 3D classification. Only the five directions related to C5 symmetry were considered. Other angular and translational parameters were fixed. No symmetry was applied, and two classes were used. Spikes in one class were severely tilted and excluded from further processing whereas spikes in the other class were only moderately tilted and were included. To reconstruct an initial map of the moderately tilted spike, RELION 3D auto-refinement was used with 3.7° initial angular sampling and default search range. No symmetry was applied. To re-align the spike C5 symmetry axis, another local alignment was carried out using 7.5° initial angular sampling and custom angular search range sigma of

10° (option *sigma\_ang*). To compensate for possibly large rotations of the spike, shifts of ±10 pixels were allowed. To calculate the final reconstruction of the spike, the sub-particle alignment parameters were adjusted by applying the required rotation to align their C5 symmetry axis on the Z-axis. This was performed using protocol ‘*xmipp – align volume and particles*’ and volume alignment of the initial asymmetric spike reconstruction on the initial C5 spike reconstruction. The adjusted spike orientations were considered as priors in the final round of RELION local auto-refinement using 3.7° initial angular sampling and default search range. This workflow improved the resolution of the spike from 7.2 Å (tilted spike, no symmetry) to 5.3 Å (aligned spike, C5 symmetry) and B-factor from 420 to 219 Å<sup>2</sup>, respectively.

### Composite model of the capsid

Protocol ‘*localrec – stitch subvolumes*’ in Scipion plugin LocalRec 3.0.2 was used to combine the localized reconstructions of trimers 1–4, penton base and the spike into a composite map of the icosahedral capsid<sup>65</sup>. A spherical mask with a diameter of 80 pixels was used for the trimers and the penton. To define the boundary between the high-resolution penton map and the low-resolution spike map, a spherical mask defining the spike (diameter 100 pixels) was first subtracted from the spherical mask of the penton after an appropriate shift reflecting the different centering of the two localized reconstructions relative to the capsid. This ensured that signals for the shaft and linker regions of the spike were maximally retained in the composite map. Prior to running the protocol, the density values in the low-resolution spike map were scaled by a factor of two which was determined by trial and error to allow their visualization as a single isosurface.

### Localized reconstruction of the FLiP penton

Localized reconstruction of the FLiP penton was calculated similarly to the ΦCjT23 penton (see above). Particles (2351) were extracted from a previously published data set of 235 micrographs. A total of 28,212 penton sub-particles were defined at a radius of 300 Å and extracted in a box of 200 × 200 pixels. One round of local RELION 3D auto-refinement with an initial angular sampling of 1.8° and a custom search range of 2° (option *sigma\_ang*) resulted in a FLiP penton map with a resolution of 4.0 Å. Data acquisition parameters are summarized in Supplementary Table 2.

### Atomic model building and refinement

To identify ORFs encoding ΦCjT23 MCP, penton, and FLiP penton, polyalanine models of each target protein were built into their respective cryo-EM reconstructions in Coot<sup>67</sup>. Secondary structure elements found in the polyalanine models were compared to the secondary structure elements predicted for all ΦCjT23 and FLiP ORFs using JPred<sup>68</sup>, and sequences from ORFs with similar profiles were built into the cryo-EM reconstructions in Coot in order to validate matches between sequence and density. Following ORF identification, model building, and structure refinement were performed using Coot and PHENIX<sup>69</sup> in consecutive cycles with secondary structure and geometry restrained. Finally, the four ΦCjT23 MCP trimer atomic models and the ΦCjT23 penton atomic model were rigid-body fitted back in their original positions of the capsid to create an atomic model of the asymmetric unit of the entire capsid (only one chain from trimer 3 was kept). Refinement statistics of all atomic models are presented in Supplementary Table 1. All figures were rendered in UCSF ChimeraX<sup>70</sup> and topology diagrams were generated with PDBsum<sup>71</sup>.

### Structural phylogeny

Structural alignments of ΦCjT23 MCP and other viral MCPs with a double β-sandwich fold and a phylogenetic tree were calculated using Homologous Structure Finder software 1.1<sup>27</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The sequence data generated in this study have been deposited in the GenBank genetic sequence database under the accession codes ON067806 (ΦCjT23 genomic sequence) and CP097434 (*Flavobacterium* sp. B183 whole genome sequence). The cryo-EM data have been deposited in the Electron Microscopy Data Bank (EMDB) under the accession codes EMD-15042, EMD-15044, EMD-15045, EMD-15046, EMD-15047, EMD-15048, EMD-15049, EMD-15050, EMD-15051 (cryo-EM density maps) and in the Protein Data Bank (PDB) under the accession codes 7ZZZ, 8A01, 8A02, 8A03, 8A04, 8A05, 8A06 (atomic models). All other data generated or analyzed during this study are included in this published article (and its supplementary information files). Source data are provided with this paper.

### References

- Krupovic, M., Dolja, V. V. & Koonin, E. V. Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat. Rev. Microbiol.* **17**, 449–458 (2019).
- Mokili, J. L., Rohwer, F. & Dutilh, B. E. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* **2**, 63–77 (2012).
- Bin Jang, H. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
- Abrescia, N. G. A., Bamford, D. H., Grimes, J. M. & Stuart, D. I. Structure unifies the viral universe. *Annu Rev. Biochem.* **81**, 795–822 (2012).
- Roberts, M. M., White, J. L., Grutter, M. G. & Burnett, R. M. 3-Dimensional structure of the adenovirus major coat protein hexon. *Science* **232**, 1148–1151 (1986).
- Bahar, M. W., Graham, S. C., Stuart, D. I. & Grimes, J. M. Insights into the evolution of a complex virus from the crystal structure of vaccinia virus D13. *Structure* **19**, 1011–1020 (2011).
- De Castro, C. et al. Structure of the chlorovirus PBCV-1 major capsid glycoprotein determined by combining crystallographic and carbohydrate molecular modeling approaches. *Proc. Natl Acad. Sci. USA* **115**, E44–E52 (2018).
- Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98**, 825–833 (1999).
- Abrescia, N. G. A. et al. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**, 68–74 (2004).
- Laurinmäki, P. A., Huiskonen, J. T., Bamford, D. H. & Butcher, S. J. Membrane proteins modulate the bilayer curvature in the bacterial virus Bam35. *Structure* **13**, 1819–1828 (2005).
- Adams, M. J. et al. 50 years of the International Committee on Taxonomy of Viruses: progress and prospects. *Arch. Virol.* **162**, 1441–1446 (2017).
- Abrescia, N. G. A. et al. Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Mol. Cell* **31**, 749–761 (2008).
- Krupovic, M., Makarova, K. S. & Koonin, E. V. Cellular homologs of the double jelly-roll major capsid proteins clarify the origins of an ancient virus kingdom. *Proc. Natl Acad. Sci. USA* **119**, e2120620119 (2022).
- Laanto, E. et al. Virus found in a boreal lake links ssDNA and dsDNA viruses. *Proc. Natl Acad. Sci. USA* **114**, 8378–8383 (2017).
- San Martin, C. et al. Minor proteins, mobile arms and membrane-capsid interactions in the bacteriophage PRD1 capsid. *Nat. Struct. Biol.* **9**, 756–763 (2002).
- Huiskonen, J. T., Kivela, H. M., Bamford, D. H. & Butcher, S. J. The PM2 virion has a novel organization with an internal membrane and

- pentameric receptor binding spikes. *Nat. Struct. Mol. Biol.* **11**, 850–856 (2004).
17. Zhang, X. et al. Structure of Sputnik, a virophage, at 3.5-Å resolution. *Proc. Natl Acad. Sci. USA* **109**, 18431–18436 (2012).
18. Aalto, A. P. et al. Snapshot of virus evolution in hypersaline environments from the characterization of a membrane-containing *Salisaeta* icosahedral phage 1. *Proc. Natl Acad. Sci. USA* **109**, 7079–7084 (2012).
19. Veesler, D. et al. Atomic structure of the 75 MDa extremophile *Sulfolobus* turreted icosahedral virus determined by CryoEM and X-ray crystallography. *Proc. Natl Acad. Sci. USA* **110**, 5504–5509 (2013).
20. Klose, T. et al. Structure of faustovirus, a large dsDNA virus. *Proc. Natl Acad. Sci. USA* **113**, 6206–6211 (2016).
21. Wang, N. et al. Architecture of African swine fever virus and implications for viral assembly. *Science* **366**, 640–644 (2019).
22. Richter, C. A. & Pate, J. L. Temperate phages and bacteriocins of the gliding bacterium *Cytophaga johnsonae*. *J. Gen. Microbiol.* **134**, 253–262 (1988).
23. Huber, K. E. & Waldor, M. K. Filamentous phage integration requires the host recombinases XerC and XerD. *Nature* **417**, 656–659 (2002).
24. Ilca, S. L. et al. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat. Commun.* **6**, 8843 (2015).
25. Rossmann, M. G. et al. Structural comparisons of some small spherical plant viruses. *J. Mol. Biol.* **165**, 711–736 (1983).
26. Caspar, D. L. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
27. Ravanti, J., Bamford, D. & Stuart, D. I. Automatic comparison and classification of protein structures. *J. Struct. Biol.* **183**, 47–56 (2013).
28. Ilca, S. L. et al. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. *Nature* **570**, 252–256 (2019).
29. Youkharibache, P. et al. The small beta-Barrel domain: a survey-based structural analysis. *Structure* **27**, 6–26 (2019).
30. Bewley, M. C., Springer, K., Zhang, Y. B., Freimuth, P. & Flanagan, J. M. Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. *Science* **286**, 1579–1583 (1999).
31. Xu, L., Benson, S. D., Butcher, S. J., Bamford, D. H. & Burnett, R. M. The receptor binding protein P2 of PRD1, a virus targeting antibiotic-resistant bacteria, has a novel fold suggesting multiple functions. *Structure* **11**, 309–322 (2003).
32. Mannige, R. V. & Brooks, C. L. 3rd Periodic table of virus capsids: implications for natural selection and design. *PLoS ONE* **5**, e9423 (2010).
33. Butcher, S. J., Bamford, D. H. & Fuller, S. D. DNA packaging orders the membrane of bacteriophage PRD1. *EMBO J.* **14**, 6078–6086 (1995).
34. Laurinmaki, P. A., Huiskonen, J. T., Bamford, D. H. & Butcher, S. J. Membrane proteins modulate the bilayer curvature in the bacterial virus Bam35. *Structure* **13**, 1819–1828 (2005).
35. Stewart, P. L., Burnett, R. M., Cyrklaff, M. & Fuller, S. D. Image reconstruction reveals the complex molecular organization of adenovirus. *Cell* **67**, 145–154 (1991).
36. van Raaij, M. J., Mitraki, A., Lavigne, G. & Cusack, S. A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**, 935–938 (1999).
37. Reddy, H. K., Carroni, M., Hajdu, J. & Svenda, M. Electron cryo-microscopy of bacteriophage PR772 reveals the elusive vertex complex and the capsid architecture. *Elife* **8**, e48496 (2019).
38. Huiskonen, J. T., Manole, V. & Butcher, S. J. Tale of two spikes in bacteriophage PRD1. *Proc. Natl Acad. Sci. USA* **104**, 6666–6671 (2007).
39. Eiler, A. & Bertilsson, S. *Flavobacteria* blooms in four eutrophic lakes: linking population dynamics of freshwater bacterioplankton to resource availability. *Appl. Environ. Microbiol.* **73**, 3511–3518 (2007).
40. Zhu, F., Wang, S. & Zhou, P. *Flavobacterium xinjiangense* sp. nov. and *Flavobacterium omnivorum* sp. nov., novel psychrophiles from the China No. 1 glacier. *Int. J. Syst. Evol. Microbiol.* **53**, 853–857 (2003).
41. Bernardet, J. F. & Bowman, J. P. The genus *Flavobacterium*. In: *The Prokaryotes* (eds Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K. H. & Stackebrandt, E.) (Springer, 2006).
42. Steven, B. et al. Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. *FEMS Microbiol. Ecol.* **59**, 513–523 (2007).
43. Zhong, Z. P. et al. Glacier ice archives nearly 15,000-year-old microbes and phages. *Microbiome* **9**, 160 (2021).
44. Labonte, J. M. & Suttle, C. A. Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J.* **7**, 2169–2177 (2013).
45. Szekeley, A. J. & Breitbart, M. Single-stranded DNA phages: from early molecular biology tools to recent revolutions in environmental microbiology. *FEMS Microbiol. Lett.* **363**, fnw027 (2016).
46. Brum, J. R., Schenck, R. O. & Sullivan, M. B. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *ISME J.* **7**, 1738–1751 (2013).
47. Laanto, E., Sundberg, L. R. & Bamford, J. K. Phage specificity of the freshwater fish pathogen *Flavobacterium columnare*. *Appl. Environ. Microbiol.* **77**, 7868–7872 (2011).
48. Kropinski, A. M. Measurement of the rate of attachment of bacteriophage to cells. *Methods Mol. Biol.* **501**, 151–155 (2009).
49. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
50. Santos, M. A. An improved method for the small scale preparation of bacteriophage DNA based on phage precipitation by zinc chloride. *Nucleic Acids Res.* **19**, 5442 (1991).
51. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679 (2007).
52. Besemer, J., Lomsadze, A. & Borodovsky, M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**, 2607–2618 (2001).
53. Johnson, M. et al. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–9 (2008).
54. Zimmermann, L. et al. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred Server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
55. Chin, C. S. et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
56. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
57. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
58. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
59. de la Rosa-Trevin, J. M. et al. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
60. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
61. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

62. Rohou, A. & Grigorieff, N. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
63. Kivioja, T., Ravantti, J., Verkhovskiy, A., Ukkonen, E. & Bamford, D. Local average intensity-based method for identifying spherical particles in electron micrographs. *J. Struct. Biol.* **131**, 126–134 (2000).
64. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
65. Abrishami, V. et al. Localized reconstruction in Scipion expedites the analysis of symmetry mismatches in cryo-EM data. *Prog. Biophys. Mol. Biol.* **160**, 43–52 (2021).
66. Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**, e11182 (2015).
67. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
68. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–394 (2015).
69. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
70. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
71. Laskowski, R. A., Jablonska, J., Pravda, L., Varekova, R. S. & Thornton, J. M. PDBsum: structural summaries of PDB entries. *Protein Sci.* **27**, 129–134 (2018).
72. Pei, J. & Grishin, N. V. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700–712 (2001).

## Acknowledgements

We thank Mark McBride for bacterial strains and comments on the manuscript as well as Denise Tremblay, Benita Löflund, Kati Mäkelä, Ville Hoikkala and Pasi Laurinmäki for technical assistance. This study was supported by the Academy of Finland (342988 to I.R., 321985 to E.L. and 314939 and 346772 to L.-R.S.) and Emil Aaltonen Foundation (to L.-R.S.). The facilities and expertise of the HiLIFE cryo-EM unit at the University of Helsinki, a member of Instruct-ERIC Centre Finland, FINStruct, and Bio-center Finland are gratefully acknowledged. The authors wish to acknowledge CSC—IT Center for Science, Finland, for computational resources. S.M. holds the Canada Research Chair in Bacteriophages.

## Author contributions

Conceptualization, E.L., S.M., L.-R.S., J.T.H.; Formal analysis, N.K.; Investigation, N.K., E.L., I.R., M.S., J.R., J.T.H.; Writing—original draft, J.T.H.; Writing—review & Editing, all authors; Visualization, E.L., I.R., J.R., L.-R.S., J.T.H.; Resources, S.M., L.-R.S., J.T.H.; Software, V.A.; Supervision, L.-R.S., J.T.H.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-35123-6>.

**Correspondence** and requests for materials should be addressed to Lotta-Riina Sundberg or Juha T. Huiskonen.

**Peer review information** *Nature Communications* thanks Mart Krupovic and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022