BIOLOGICAL SCIENCES

A new virus type found from a boreal lake links ssDNA and dsDNA viruses

Authors: Elina Laanto[1,4], Sari Mäntynen[1,4], Luigi De Colibus[2,4], Jenni Marjakangas[1], Ashley Gillum[2], David I. Stuart[2], Janne J. Ravantti[1,3], Juha T. Huiskonen[2,3 *], and Lotta-Riina Sundberg[1*]

Affiliations: [1] University of Jyväskylä, Centre of Excellence in Biological Interactions, Department of Biological and Environmental Science and Nanoscience Center, P.O. Box 35, 40014 University of Jyväskylä, Finland

[2] Division of Structural Biology, Welcome Trust Centre for Human Genetics, Roosevelt Drive, OX3 7BN, University of Oxford, Oxford, United Kingdom

[3] Institute of Biotechnology and Department of Biosciences, P.O. Box 56, 00014 University of Helsinki, Helsinki, Finland

[4] These authors contributed equally to the study.

[*]Corresponding authors:

Juha Huiskonen: juha@strubi.ox.ac.uk, +441865287844

Lotta-Riina Sundberg: lotta-riina.sundberg@jyu.fi, +358408053931

KEY WORDS: Cryo-electron microscopy, Flavobacterium, lipids, genome, virus structure

Summary

Viruses have impacted the biosphere in numerous ways since the dawn of life. However, the evolution, genetic, structural and taxonomic diversity of viruses remain poorly understood. This is partially due to sparse sampling of the virosphere, which has mostly concentrated on exploring the abundance and diversity of dsDNA viruses. Furthermore, viral genomes are highly diverse, and using only the current sequence-based methods for classifying viruses and studying their phylogeny is complicated. We describe the first virus, FLiP (*Flavobacterium* infecting, lipid-containing phage), with a circular ssDNA genome and an internal lipid membrane enclosed in the icosahedral capsid. The 9,174-nt long genome showed limited sequence similarity to other known viruses. The genetic data implies that this virus might use similar replication mechanisms to those found in other ssDNA replicons. However, the structure of the viral major capsid protein, elucidated at near-atomic resolution using electron cryo-microscopy, is strikingly similar to that observed in dsDNA viruses of the PRD1–adenovirus lineage, characterised by a major capsid protein bearing two beta-barrels. The strong similarity between FLiP and another member of the structural lineage, bacteriophage PM2, extends to the capsid organisation (pseudo *T*=21 *dextro*) in spite of the difference in the genetic material packaged and the lack of significant sequence similarity.

Significance

We describe phage FLiP, the first ssDNA virus with an icosahedral capsid and an internal lipid membrane. FLiP genome shows limited similarity to known sequences, although a ssDNA replication mechanism was implied by genome analysis. However, as the capsid protein fold indicates relatedness with the dsDNA viruses of the PRD1-adenovirus lineage, FLiP exhibits a unique combination of structural and replication modules. It is suggested, that the capsid protein structure could be used to complement the sequence data when classifying viruses, as well as detecting their deep evolutionary relationships, especially in absence of sequence similarities. Furthermore, these findings demonstrate the value of characterizing unknown viruses from diverse environmental sources to understand the diversity of the microbial world.

\body

Introduction

Although conservative calculations estimate the number of viruses in the biome to exceed 10^{31} virions (1) this enormous group of biological entities is largely unexplored, as only a fraction of viruses have been studied in detail. In recent years, the enormous diversity of the viral world (2) has been revealed by high-throughput sequencing of viral genomes. Whereas the traditionally used virus detection methods (epifluorescence microscopy and pulse-field gel electrophoresis and early metagenomic studies) have been selective for dsDNA viruses (3-5) recent metagenomic studies have revealed that the number of ssDNA viruses in nature has been grossly underestimated. In fact, ssDNA viruses are widespread and may predominate in certain habitats, including the oceans and arctic freshwaters (6-10). However, only a few ssDNA viruses have been cultivated in laboratory conditions and therefore detailed structural and biochemical analysis of distinct ssDNA viruses is mostly lacking. Detailed characterization of novel virus isolates is essential to expand our understanding of evolutionary relationships between viruses and their role in microbial communities (11).

Using solely sequence-based methods for virus classification and phylogeny is complicated due to the enormous genetic diversity among viruses (11, 12). However, essential viral structures and functions are often strongly conserved (13, 14). There seems to be only a limited number of folds a functional protein may adopt and stringent structural constraints further decrease the number of folds suitable for forming a capsid protein. This makes the serendipitous invention of new architectures very unlikely, so that only a very limited number of lineages of virion architecture have been observed (15, 16)·(17). Therefore, it has been suggested, that the major capsid protein (MCP) fold and overall virion structure could be used to classify viruses and track down their deeper evolutionary relationships (18).

We describe here the first icosahedral virus with an ssDNA genome and an internal lipid membrane designated as “FLiP” (*Flavobacterium* infecting, lipid-containing phage). The 9,174-nt long genome of FLiP has limited sequence similarities with previously identified viruses. Based on cryo-electron microscopy (at 4 Å resolution), the virion structure of FLiP displays pseudo T=21 *dextro* organisation, previously shown only for marine dsDNA phage PM2. The major capsid protein (MCP) consists of two beta-barrels with jelly-roll topology and resembles closely dsDNA viruses of the PRD1-adenovirus –lineage. Thus, these structural data indicate evolutionary relatedness between some ssDNA and dsDNA viruses, and suggest that it would be beneficial to complement traditional sequence-based systems by structure-based approaches.

Results

FLiP is a novel phage infecting *Flavobacterium*

The virus, designated as "FLiP" (*Flavobacterium* infecting, lipid-containing phage), was isolated together with the host bacterium from a boreal freshwater habitat in Central Finland in September 2010. The host bacterium was identified as a member of genus *Flavobacterium* (Bacteroidetes) by 16S rRNA sequencing. *Flavobacterium* species are important members of freshwater bacterial communities in the boreal regions (19, 20). The greatest similarity to the obtained 1422 bp long 16S rRNA-sequence was to several *Flavobacterium* sp. strains (99 %). Accordingly, the host bacterium strain of FLiP was designated as *Flavobacterium* sp. B330.

FLiP virions were collected for purification from plate lysate. Phage particles were concentrated from the filtered lysate with PEG-NaCl –precipitation and then purified by rate zonal and equilibrium centrifugation to near homogeneity. The rate zonal centrifugation was optimized using different gradient materials (iodixanol, glycerol and sucrose). In sucrose gradient the phage yield was significantly higher, therefore this gradient material was routinely used for purification in both rate zonal and equilibrium centrifugation. This method yielded highly purified particles with specific infectivity of 1.28×10^{12} PFU/mg of protein. The typical recovery and specific infectivity for each purification step is represented in Supplementary Table S1.

FLiP genome is a circular ssDNA molecule with 16 putative ORFs

Genomic nucleic acid of FLiP was extracted from purified virions. Nuclease treatments (Supplementary Fig. S1) revealed that the genome is an ssDNA molecule. Genome sequencing resulted in a 9,174 nt long molecule, with an overall GC content of 34 % (Fig. 1a). A total of 16 putative open reading frames (ORFs) were identified from the genome using programs Glimmer (21) and Genemark (22). All the ORFs were oriented in the same direction. Most of the ORFs showed limited similarity to other sequences in the public databases (Supplementary Table s2), which is common for environmental phage isolates.

Identification of FLiP structural proteins

Purified virions were subjected to sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis. The most abundant protein species detected were approximately 35 kDa in size (Fig. 1b). Resolved structural proteins were identified by mass-spectrometry and N-terminal sequencing was performed for the most abundant protein species (Supplementary Table S3).

According to these analyses five ORFs (7, 8, 9, 11 and 14) were designated as genes as their protein products were identified (Fig. 1b). Due to the high expression level, gene 8 was predicted to encode the major capsid protein. Translated gene 14 showed a high sequence similarity to several lytic transglycosylases and contained the conserved transglycosylase domain (E-value 4.97×10^{-4}). This indicates that a protein with lytic activity is present in the virion structure.

FLiP obtains lipids selectively from the host

Interestingly, FLiP is the first described ssDNA virus with an inner membrane. Low buoyant density (1,21 g/cm^3 and 1,18 g/cm^3 in $CsCl_2$ and sucrose gradients, respectively) chloroform sensitivity and positive Sudan Black staining (data not shown) indicated the presence of a lipid membrane in the virion. The lipid class and molecular species composition of the virion and the host bacterium inner membrane were analysed from the extracted lipids by mass spectrometry (triple quadrupole ESI-MS/MS; Supplementary Fig. 2). FLiP contains significantly more ceramide (60 % of the total lipid composition) than the host cytoplasmic membrane. Ornithine lipids are enriched in the host membrane but present only in a negligible amount in the phage (approximately 7 %). The differing lipid compositions imply that the phage obtains the lipids from the host selectively.

Cryo-EM reconstruction of FLiP virion and MCP structure

We solved the structure of the FLiP virion using cryo-EM (Fig. 2a). Micrographs of purified virions revealed spherical particles that were ~60 nm in diameter (Supplementary Fig. S3a,b). Some of the particles appeared hollow, suggesting they were empty, lacking the DNA genome and possibly also the internal lipids (Supplementary Fig. S3c). Single particle analysis of 2,203 'full' particles and 'gold-standard' three-dimensional structure refinement (23) yielded an icosahedrally symmetric reconstruction of the FLiP virion, solved from a final set of 934 particles (Fig. 2a; Supplementary Table S3). The resolution of the viral protein capsid was on average 4.0 Å, as estimated by Fourier shell correlation at a 0.143-threshold (Supplementary Fig. S3d). Some local areas of the viral surface were less resolved while resolution in the capsid interior was approaching 3.9 Å resolution (Supplementary Fig. S3e,f). Reaching such resolution from a relatively small number of particles has been reported before, for example in the case of deformed wing virus (EMD-3570) (24). In the FLiP reconstruction, pentameric spikes (12 nm tall) protrude from the twelve icosahedral vertices (Fig. 2a,b). The major capsid proteins forming the outer protein shell follow a 'pseudo *T*=21 *dextro*' icosahedral capsid organisation (Fig. 2a), observed also in the dsDNA phage PM2 (25). The shape

of the shell is highly faceted, with facet-to-facet and edge-to-edge distance of 53 nm, and 55 nm, respectively, compared to a vertex-to-vertex distance of 59 nm as measured from the base of the spike. The outer protein shell covers a 5-nm thick lipid bilayer membrane (Fig. 2b). Minor structural components, located between the outer protein shell and the membrane, bridge the two together (Fig.2b).

The near-atomic resolution of the FLiP cryo-EM reconstruction facilitated building an atomic model of the major capsid protein (MCP) *de novo* (Supplementary Fig. S4; Supplementary Table S5). Minor structural proteins that were resolved in the reconstructions, such as the pentameric spikes, were not built in the absence of sequence assignments. The MCP consists of two beta-barrels/sandwiches, each of which is composed of two anti-parallel beta-sheets (Fig. 3a). Both of the barrels have a jelly-roll topology (Fig. 3b). Strands C, H, E and F contribute to the sheets facing the surface of the trimer whereas strands B, I, D and G contribute to the sheets facing towards the centre of the trimer (Fig. 3a, b). Strand B is absent in the N-terminal beta-barrel. The two beta-barrels are decorated by loops, in addition to alpha-helices embedded between strands F and G. A long C-terminal alpha-helix resides horizontally at the base of the MCP. Trimers of the double-beta-barrel MCP, are formed with the six beta-barrels of the MCP subunit arranged to form a pseudo-hexameric molecule (Fig. 2).

FLiP MCP fold suggests the virus belongs to PRD1-Adenovirus lineage

The overall virion architecture and details of the MCP fold (including the topology and disposition of the beta-barrels in addition to the position of the embedded alpha-helices) resemble closely those observed in members of the PRD1–adenovirus lineage (14) (Fig. 3c, Fig. 4). This lineage of icosahedral viruses has until now comprised solely dsDNA viruses, albeit infecting organisms ranging from bacteria (e.g. PRD1 (26) and PM2 (27)) through archaea (*Sulfolobus* turreted icosahedral virus (28)) to green algae (*Paramecium bursaria Chlorella* virus 1 (29)) and higher eukaryotes (e.g. adenovirus (30)). We calculated a structure-based phylogenetic tree by aligning MCPs of different members of this lineage against FLiP using Homologous Structure Finder program (31) (Fig. 3c). FliP branches from the bacterial/archaeal arm of the tree, being in no sense an outlier, and is most closely structurally related to bacteriophage PM2. For these two viruses, the structural homology extends beyond the conserved double-beta-barrel MCP fold. Notably, both FLiP and PM2 share the same pseudo *T*=21 *dextro* capsid organization, although PM2 is slightly larger in size than FLiP (vertex-to-vertex distance 63 nm vs. 59 nm; Fig. 4) (25). The two viruses differ also in how the capsid is bridged to the internal lipid bilayer. In both viruses, minor

membrane proteins interact with the MCP inner surface via alpha-helices that are parallel to the capsid (Fig. 4d). In FLiP, these parallel helices correspond to the C-terminus of the MCP (Fig. 2e; Fig. 3a; α'', Ser285–Lys360) whereas in PM2, the helices belong to another minor protein P3 (27).

Discussion

Here, we describe a novel virus, FLiP, which infects a *Flavobacterium* host. The FLiP virion consists of an icosahedrally symmetric protein capsid enclosing a circular ssDNA genome of 9,174 nucleotides in length (Fig. 1). The inner surface of the capsid is covered by a lipid membrane, which makes FLiP unique among previously described ssDNA phages. Even outside of those ssDNA genomes, few bacterial viruses with internal lipid membranes have been characterized (32). The presence of the membrane evidently facilitates virus-host interactions (32), and similarly to FLiP, some viruses have been shown to derive lipids selectively from the host (33, 34). FLiP contained significantly more ceramide than the host cytoplasmic membrane. This may relate to the cone-shaped structure of ceramide, which could favour the formation of the highly curved viral membrane.

The FLiP genome is considerably larger than that of the previously isolated of non-tailed ssDNA phages except for the *Cellulophaga* phage phi48:2 (35). FLiP shows limited sequence similarity to any known sequences in the database (Supplementary Table S2), which is typical for environmental phage isolates and probably reflects the limited number of cultured virus isolates. Similarities were detected to *Flavobacterium* and *Zunongwangia profunda* as well as ssDNA phages infecting *Cellulophaga* (35, 36). *Zunongwangia profunda* and *Cellulophaga* belong to the Bacteroidetes phylum as the FLiP host *Flavobacterium*. Moreover, ORF4 contains conserved DNA-binding helix-turn-helix domain of Xre family. Interestingly, ORF15 has resemblance to rolling circle replication initiator proteins from *Staphylococcus aureus* and *Geobacillus stearothermophilus*, which suggests that FLiP might use replication mechanisms similar to those found in other ssDNA replicons.

However, despite the unique combination of ssDNA genome and lipid membrane and the limited sequence similarities, the further structural characterization of FLiP implies intriguing connections to previously identified dsDNA viruses in the PRD1-adenovirus lineage. Although the number of available high resolution structures of viral MCPs is increasing, most viruses seem to fit into just a few structure-based lineages, to the extent that certain DNA viruses (e.g. ssDNA phage phiX174 as well as papilloma and polyomaviruses with dsDNA genomes) have been assigned into the same

lineage together with some positive sense ssRNA viruses (15). The overall virion structure of FLiP suggests a relationship with the members of the PRD1-adenovirus lineage and PM2 in particular. In addition, the MCP of PM2 is the closest homolog of the FLiP MCP, strongly suggesting that FLiP is a member of the PRD1-adenovirus lineage, despite the fact that the genomes of other members of that lineage are dsDNA, as opposed to circular ssDNA of FLiP (Fig. 3c). This contrasts with the traditional classification of viruses, accepted by the International Committee on the Taxonomy of Viruses (ICTV), where genome type is used as the top level of classification, namely into ssDNA, dsDNA, ssRNA and dsRNA virus groups. The data we present here, together with previous observations (15, 37) raise the question, of whether it would be appropriate to place more emphasis on the virion architecture and the structure of the capsid proteins, when classifying viruses and assigning distant viruses to higher taxonomic levels. To gain a more comprehensive view on the evolution of the virus world, it would seem beneficial to complement the genomic analyses with structural data.

Pervasive shuffling of genes and gene modules is undoubtedly a key feature in the virus evolution. Viral genomes commonly encompass structural and replication modules, which may have different evolutionary provenances (38, 39). The recombination of these modules may provide adaptive advantage for viruses and give rise to novel virus types (38). The genetic resemblance of FLiP ORF15 to rolling circle replication initiation proteins implies, that this virus may display a fascinating combination of a replication mechanism typical of ssDNA viruses and structural characteristics of dsDNA viruses. However, the replication mechanism of FLiP is yet to be confirmed. Interestingly, PM2-like prophages with differing replication machineries seem to be common in aquatic bacteria, although their structural proteins have remain conserved (40).

It is evident that bacterial viruses represent an enormous reservoir of genetic diversity. However, only a small fraction of them have been structurally and biochemically characterized. The finding of FLiP exemplifies the importance of detailed characterization of novel viruses from diverse environments in obtaining deeper understanding of the microbial world. This is especially true for ssDNA viruses whose diversity and ecological impact have remained poorly understood, despite their abundance.

Methods

Isolation of the bacteriophage and host bacterial strains

Bacteriophage FLiP (*Flavobacterium* infecting, lipid-containing phage) and its host bacterium *Flavobacterium* sp. B330 were isolated from a freshwater sample, taken from lake Jyväsjärvi in Central Finland (N62°13.840' E 025°44.510) as described elsewhere (41). Afterwards B330 was routinely grown in Shieh medium (42) at RT, using a constant agitation of 110 rpm. FLiP virions were isolated from a filtered (pore size 0.45 µm, Micropore, MA, USA) water sample by enriching host bacterium in Luria–Bertani (LB) liquid medium diluted to 1/5 with the filtrated water sample at RT for 2 days (110 rpm). Enrichment culture was mixed with soft agar (0.7 % [w/v]) in 1/5 LB and plated on solid agar. The plate was incubated at RT for 2 days, after which single plaques were picked, suspended in 500 µl of 1/5 LB medium and stored at 4°C. Three rounds of plaque purification were performed. Bacteriophage was propagated using the standard double agar layer technique (43). In brief, 100 µl of the bacteriophage suspension in different dilutions was added to 3 ml of molten soft agar (0.7 % [w/v]) together with 200 µl of the host cell culture. The mixture was poured onto the surface of Shieh agar plates (1 % [w/v]) and incubated overnight at RT.

16S rRNA sequencing of the host bacterium

Genomic DNA of B330 was extracted with GeneJET Genomic DNA Purification kit (Fermentas) according to manufacturer´s instructions. The gene encoding the 16S rRNA gene was amplified by PCR using primers fD1 and rD1 (44). The PCR products were purified with QIAquick PCR purification kit (Thermo Scientific) and sequenced using BigDye Terminator v3.1 kit and ABI Prism Genetic Analyzer 3100 (Life Technologies). The sequence data was analysed in June 2015 using BlastN (45).

Amplification, purification and stability of FLiP-virions

For the preparation of high titer phage lysates, plates showing confluent lysis were used. 5 ml of Shieh was added onto each plate and the suspension was incubated in cold (approximately 6°C) for 3 hours with agitation. Medium was collected from the plates and the resulting supernatant was filtrated through a 0.45-µm-pore-size membrane filter (PALL Life Sciences). Viral titer of the filtrated supernatant was determined by the plaque assay and the phage stock was stored at 4°C and for longer periods at –80°C with 20 % glycerol.

For phage purification, the lysate was collected from several hundreds of confluent plates. The lysate was filtered and virus particles were precipitated with 10 % (w/v) polyethylene glycol 6000

and 0.5 M NaCl. Phage precipitate was collected by centrifugation (Sorvall SLA3000, 11,000 × g, 30 min, 4°C) and suspended in 20 mM PPB (pH 7.2). Three different gradient materials (all in 20 mM PPB) were tested for optimal purification: glycerol (5–30 %), OptiPrep (5–20 %) and sucrose (5–20 %), the latter of which was selected. The suspension was layered on top of a 5–20 % (w/v) sucrose gradient (in 20 mM PPB pH 7.2) for subsequent rate zonal centrifugation (Beckman Optima L-K90 ultracentrifuge, SW28 rotor, 104,000 × g, 1 h, 15°C). The light-scattering zone containing the virus was collected and further purified by equilibrium centrifugation in 20–70 % (w/v) sucrose gradient (in 20 mM PPB pH 7.2; Beckman Optima L-K90, SW41 rotor, 175,000 × g, 19 h, 5°C). Light-scattering zones were collected, after which the phage particles were pelleted by centrifugation (Beckman Optima L-K90, 70Ti rotor, 112,000 × g, 3 h, 5°C), suspended in 20 mM PPB (pH 7.2) and stored at –20°C or –80°C.

Analysis of phage proteins

The protein concentrations of purified phage solutions were measured by the Bradford assay (46), using bovine serum albumin as a standard. Structural proteins of the purified virions were resolved by 16 % SDS-PAGE. The gel was stained with Coomassie brilliant blue to detect the proteins. For structural protein identification, protein bands were cut from the gel.

Structural proteins were identified by nano liquid chromatography-electrospray ionization-tandem mass spectrometry (nanoHPLC-ESI-MS/MS) in Proteome Factory (Proteome Factory AG, Berlin, Germany). The LCMS system included an Agilent 1100 nanoHPLC system (Agilent, Waldbronn, Germany), PicoTip electrospray emitter (New Objective, Woburn, MA) and a LTQ-FT or an Orbitrap XL Ultra mass spectrometer (ThermoFisher, Bremen, Germany). For LC-MS/MS analysis SDS-PAGE-separated proteins were excised from the gels and in-gel digested by trypsin. Peptides were first trapped and then desalted on the enrichment column (Zorbax SB C18, 0.3 × 5 mm, Agilent) for 5 minutes (solvent: 2.5 % acetonitrile/0.5 % formic acid), after which they were separated on a Zorbax 300SB-C18, 75 µm × 150 mm column (Agilent) using a linear gradient from 10 % to 32 % B (solvent A: 5 % acetonitrile in water, solvent B: acetonitrile, both with 0.1% formic acid). Ions of interest were data-dependently subjected to MS/MS according to the expected charge state distribution of peptide ions. Proteins were identified by database search against a database containing the RefSeq entries of *Flavobacterium johnsoniae* UW101 (National Center for Biotechnology Information, Bethesda, USA) and FLiP sequences using MS/MS ion search of the Mascot search engine (Matrix Science, London, England). Only peptides matches with a score of 20 or above were accepted.

306 **Gel electrophoresis and Edman degradation**

307 Gel electrophoresis and Edman degradation were conducted in Proteome Factory (Proteome
308 Factory AG, Berlin, Germany). 5–23 µg of sample were loaded onto 15 % SDS–PAGE under
309 reducing conditions. After electrophoresis, proteins were transferred onto an Immobilon-P
310 membrane (PVDF, pore size 0.45 mm; Millipore, Bedford, MA) using a Trans-Blot SD Semi-Dry
311 Transfer Cell (Biorad, München, Germany; 5 V; 16 h, 4 °C; blot-buffer: 40 mM borat [pH 9], 20 %
312 methanol) and visualized by Coomassie blue staining. Afterwards an ABI Procise 491 Protein
313 Sequencer (Applied Biosystems) was used for automated N-terminal sequencing according to
314 manufacturer's protocol.

315 **Isolation, sequencing and bioinformatics of phage genome**

316 To isolate phage DNA, purified phage particles were disrupted by treatment with 2 % SDS and 1.2
317 µg/ml proteinase K (37 °C, 45 min), followed by phenol-ether extraction and precipitation with
318 sodium acetate and ethanol. The purified phage genome was digested with different nucleases
319 [DNase I (Fermentas), RNAse A (Sigma-Aldrich), RNase I (Thermo Scientific), S1 nuclease
320 (Thermo Scientific), Mung bean nuclease (Promega) and *Eco*RI (Fermentas)], according to the
321 manufacturer´s instructions. The same treatments were performed for the single-stranded DNA
322 genome of φX174 (Thermo Scientific) as control. Digestion products were analysed by agarose gel
323 electrophoresis (1 % agarose in 1× Tris-acetate-ethylene diamine tetra-acetic acid buffer).

324 Complementary strands were synthesized for the single-stranded DNA genome of FLiP using
325 random hexamer primers and Klenow Fragment. Subsequently, the double-stranded complementary
326 DNA was amplified by Illustra GenomiPhi V2 DNA Amplification Kit, and sequenced using Roche
327 454 GS FLX+ -sequencer at the Institute of Biotechnology, University of Helsinki, Finland. This
328 initial sequence data was utilized to design specific oligonucleotides for primer walking to fill in the
329 sequence gaps between six contigs. Klenow Fragment was again utilized in these complementary
330 PCR reactions. The resulting PCR products were first amplified by the Illustra GenomiPhi V2 DNA
331 Amplification Kit and then sequenced using conventional Sanger sequencing with BigDye
332 Terminator v3.1 Cycle Sequencing kit and ABI Prism Genetic Analyzer 3100 (Life Technologies)
333 to obtain the complete genome sequence of FLiP. Finally, the whole genome was covered by primer
334 walking as described above.

335 GC content of the genome was calculated using Infoseq-program from EMBOSS-package. Putative
336 protein-encoding open reading frames (ORFs) were predicted from the genome using the

algorithms of Genemark (22) and Glimmer (21). Proteins homologous to translated FLiP ORFs were found using NCBI programs BlastP and PSI-Blast (47) against the nonredundant GenBank protein database. NCBI Conserved Domain Search was used to search for protein motifs. Translated ORFs were characterized by molecular mass and isoelectric point using programs included in the ExPASy Proteomics tools. Transmembrane domains were predicted using TMHMM-program (48).

Extraction of cytoplasmic membrane (CM) and lipid analysis

Extraction of cytoplasmic membrane (CM) from the cells of *Flavobacterium* sp strain B330 was done according to Laurinavičius et al (33) using a floatation gradient for separation of membranes. Lipid extraction and analysis of CM and purified phage was done commercially by the Lipidomics Unit at the University of Helsinki, using mass spectrometry (triple quadrupole ESI-MS/MS).

Electron cryo-microscopy

An aliquot (3 µl) of purified particles was applied on a glow-discharged electron microscopy grid (C-flat; Protochips, Raleigh, NC) and plunge frozen into liquid ethane using a vitrification apparatus (Vitrobot mark IV, FEI, Hillsboro, OR), operated at room temperature and above 80% relative humidity. Data were acquired using a 300-kV transmission electron microscope (Tecnai F30 'Polara'; FEI), operated at liquid nitrogen temperature and equipped with an energy filter (GIF Quantum LS; Gatan, Pleasanton, CA; zero-loss mode with 20-eV slit width) and direct electron detector (K2 Summit; Gatan). Movies (22 frames, total electron dose 22 $e^{-}/Å^2$) were collected in electron counting mode using a dose rate of 8 e^{-}/pix/s, frames were aligned in MotionCor2 to compensate for specimen drift and electron beam induced damage, and averaged together (49). Particles were picked from averaged images automatically using ETHAN (50). The picked particles were manually divided into two datasets, consisting of full and empty particles. CTF parameters were estimated locally using GCTF (51). The 3D structure full particles was determined in Relion using established protocols for image classification and gold-standard structure refinement (23). As a starting model, the previously published structure of PM2 (EMDB:1082) was used and it was filtered to low resolution (40 Å) to avoid bias. The resolution of the reconstruction was estimated by Fourier shell correlation using threshold of 0.143. Map was sharpened by applying an inverse B-factor of –100 $Å^2$ and local resolution was estimated in Relion.

Model building and refinement

PM2 major capsid protein (pdb:2VVF) was used to generate the FLiP asymmetric unit; 10 protein

chains were fitted in the FLiP cryo-EM density using Chimera (52). Subsequently the density corresponding to the asymmetric unit was extracted with Phenix suite (53) phenix.map_box and CCP4 suite (54). To facilitate the sequence assignment the map features were improved by B-factor sharpening in REFMAC (55) and density modification in PHENIX (53). The atomic model of a single MCP protomer was manually traced in the density map filtered to 3.9 Å using COOT (56). Initial model was re-built and refined in Rosetta release version 2016.32.58837 using protocols optimized for cryo-EM maps (57). The best-scoring model as estimated by density fit and geometry was selected and used in COOT to guide further model building and optimization. Final model was refined with Phenix suite phenix.real_space_refine using icosahedral constrains. Model was validated by calculating model to map real-space cross-correlation and Fourier shell correlation with and without masking the map around the model. Side-chains were validated by EMRinger (58) and model geometry by Molprobity (59). Further experimental details on model building and refinement will be published elsewhere. Structural alignment of FLiP MCP and other viral MCPs with a double-beta-barrel fold and a phylogenetic tree were calculated using HSF (Homologous Structure Finder) software (31). Figures were generated with Chimera (52).

References

1. Suttle CA (2007) Marine viruses — major players in the global ecosystem. *Nat Rev Micro* 5(10):801–812.

2. Mokili JL, Rohwer F, Dutilh BE (2012) Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* 2(1):63–77.

3. Tomaru Y, Nagasaki K (2007) Flow Cytometric Detection and Enumeration of DNA and RNA Viruses Infecting Marine Eukaryotic Microalgae. *J Oceanography* 63:215–221.

4. Steward GF (2001) Fingerprinting viral assemblages by Pulsed Field Gel Electrophoresis (PFGE). *Marine Microbiology*, Methods in Microbiology. (Elsevier), pp 85–103.

5. Breitbart M, et al. (2002) Genomic analysis of uncultured marine viral communities. *P Natl Acad Sci USA* 99:14250–14255.

6. Angly FE, et al. (2006) The marine viromes of four oceanic regions. *PLoS Biol* 4(11):e368.

7. Wegley L, Edwards R, Rodriguez-Brito B, Liu H, Rohwer F (2007) Metagenomic analysis of the microbial community associated with the coral Porites astreoides. *Environ Microbiol* 9(11):2707–2719.

8. Desnues C, et al. (2008) Biodiversity and biogeography of phages in modern stromatolites

401 and thrombolites. *Nature* 452(7185):340–343.

402 9. Tucker KP, Parsons R, Symonds EM, Breitbart M (2010) Diversity and distribution of
403 single-stranded DNA phages in the North Atlantic Ocean. *The ISME Journal* 5(5):822–830.

404 10. Aguirre de Carcer D, Lopez-Bueno A, Pearce DA, Alcami A (2015) Biodiversity and
405 distribution of polar freshwater DNA viruses. *Science Advances* 1(5):e1400127–e1400127.

406 11. Krupovic M (2013) Networks of evolutionary interactions underlying the polyphyletic origin
407 of ssDNA viruses. *Curr Opin Virol* 3(5):578–586.

408 12. Hatfull GF (2008) Bacteriophage genomics. *Curr Opin Microbiol* 11(5):447–453.

409 13. Bamford DH, Burnett RM, Stuart DI (2002) Evolution of Viral Structure. *Theor Popul Biol*
410 61(4):461–470.

411 14. Bamford DH (2003) Do viruses form lineages across different domains of life? *Res*
412 *Microbiol* 154(4):231–236.

413 15. Abrescia NGA, Bamford DH, Grimes JM, Stuart DI (2012) Structure Unifies the Viral
414 Universe. *Annu Rev Biochem* 81(1):795–822.

415 16. Oksanen HM, et al. (2012) Virus Universe: Can It Be Constructed from a Limited Number of
416 Viral Architectures. *Viruses: Essential Agents of Life* (Springer Netherlands, Dordrecht), pp
417 83–105.

418 17. Abrescia NGA, Bamford DH, Grimes JM, Stuart DI (2012) Structure unifies the viral
419 universe. *Annu Rev Biochem* 81:795–822.

420 18. Krupovic M, Bamford DH (2010) Order to the viral universe. *J Virol* 84(24):12476–12479.

421 19. Eiler A, Bertilsson S (2004) Composition of freshwater bacterial communities associated
422 with cyanobacterial blooms in four Swedish lakes - Eiler - 2004 - Environmental
423 Microbiology - Wiley Online Library. *Environ Microbiol*.

424 20. Eiler A, Bertilsson S (2007) Flavobacteria Blooms in Four Eutrophic Lakes: Linking
425 Population Dynamics of Freshwater Bacterioplankton to Resource Availability. *Appl Environ*
426 *Microbiol* 73(11):3511–3518.

427 21. Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and
428 endosymbiont DNA with Glimmer. *Bioinformatics* 23(6):673–679.

429 22. Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: a self-training method for
430 prediction of gene starts in microbial genomes. Implications for finding sequence motifs in
431 regulatory regions. *Nucleic Acids Res* 29(12):2607–2618.

432 23. Scheres SHW (2012) RELION: implementation of a Bayesian approach to cryo-EM structure
433 determination. *J Struct Biol* 180(3):519–530.

434 24. Škubník K, et al. (2017) Structure of deformed wing virus, a major honey bee pathogen. *P*
435 *Natl Acad Sci Usa* 114(12):3210–3215.

25. Huiskonen JT, Kivelä HM, Bamford DH, Butcher SJ (2004) The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. *Nat Struct Mol Biol* 11(9):850–856.

26. Benson SD, Bamford JKH, Bamford DH, Burnett RM (1999) Viral Evolution Revealed by Bacteriophage PRD1 and Human Adenovirus Coat Protein Structures. *Cell* 17:825–833.

27. Abrescia NGA, et al. (2008) Insights into Virus Evolution and Membrane Biogenesis from the Structure of the Marine Lipid-Containing Bacteriophage PM2. *Molecular Cell* 31(5):749–761.

28. Khayat R, et al. (2005) Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *P Natl Acad Sci Usa* 102:18944–18949.

29. Nandhagopal N, et al. (2002) The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *P Natl Acad Sci Usa* 99(23):14758–14763.

30. Roberts MM, White JL, Grütter MG, Burnett RM (1986) Three-dimensional structure of the adenovirus major coat protein hexon. *Science* 232(4754):1148–1151.

31. Ravantti J, Bamford D, Stuart DI (2013) Automatic comparison and classification of protein structures. *J Struct Biol* 183(1):47–56.

32. Atanasova NS, et al. (2015) Comparison of Lipid-Containing Bacterial and Archaeal Viruses. Advances in Virus Research. (Elsevier), pp 1–61.

33. Laurinavičius S, Käkelä R, Somerharju P, Bamford DH (2004) Phospholipid molecular species profiles of tectiviruses infecting Gram-negative and Gram-positive hosts. *Virology* 322(2):328–336.

34. Laurinavičius S, Bamford DH, Somerharju P (2007) Transbilayer distribution of phospholipids in bacteriophage membranes. *Biochim Biophys Acta* 1768(10):2568–2577.

35. Holmfeldt K, et al. (2013) Twelve previously unknown phage genera are ubiquitous in global oceans. *P Natl Acad Sci Usa*:1–14.

36. Holmfeldt K, Odić D, Sullivan MB, Middelboe M, Riemann L (2012) Cultivated single-stranded DNA phages that infect marine Bacteroidetes prove difficult to detect with DNA-binding stains. *Appl Environ Microbiol* 78(3):892–894.

37. Pietilä MK, Roine E, Sencilo A, Bamford DH, Oksanen HM (2015) Pleolipoviridae, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. *Arch Virol* 161(1):249–256.

38. Koonin EV, Dolja VV, Krupovic M (2015) Origins and evolution of viruses of eukaryotes_ The ultimate modularity. *Virology* 479-480(C):2–25.

39. Krupovic M, Koonin EV (2017) Multiple origins of viral capsid proteins from cellular ancestors. *P Natl Acad Sci Usa* 114(12):E2401–E2410.

40. Krupovic M, Bamford DH (2007) Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. *BMC Genomics* 8(1):236.

41. Laanto E, Sundberg L-R, Bamford JKH (2011) Phage specificity of the freshwater fish pathogen Flavobacterium columnare. *Appl Environ Microbiol* 77(21):7868–7872.

42. Song YL, Fryer JL, Rohovec JS (1988) Comparison of six media for the cultivation of Flexibacter columnaris. *Fish Pathol* 23:81–94.

43. Adams MH (1959) *Bacteriophages* (John Wiley & Sons, Ltd, New York).

44. Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 173(2):697–703.

45. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A Greedy Algorithm for Aligning DNA Sequences. *J Comput Biol* 7(1-2):203–214.

46. Bradford MM (1976) 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.

47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.

48. Sonnhammer E (1998) Protein family databases for automated protein domain identification. *Database*.

49. Zheng S, Palovcak E, Armache J-P, Cheng Y, Agard D (2016) Anisotropic Correction of Beam-induced Motion for Improved Single-particle Electron Cryo-microscopy. *BiorXiv*:1–30.

50. Kivioja T, Ravantti J, Verkhovsky A, Ukkonen E, Bamford D (2000) Local Average Intensity-Based Method for Identifying Spherical Particles in Electron Micrographs. *J Struct Biol* 131(2):126–134.

51. Zhang K (2016) Gctf: Real-time CTF determination and correction. *J Struct Biol* 193(1):1–12.

52. Pettersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.

53. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):213–221.

54. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):235–242.

55. Nicholls RA, Long F, Murshudov GN, IUCr (2012) Low-resolution refinement tools in REFMAC5. *Acta Crystallogr Sect D Biol Crystallogr* 68(4):404–417.

56. Emsley P, Cowtan K, IUCr (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* 60(12):2126–2132.

57. DiMaio F, et al. (2015) Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. *Nature Methods* 12(4):361–365.

58. Barad BA, Echols N, Wang R, Cheng Y, DiMaio F (2015) EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nature*.

59. Chen VB, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr* 66(1):12–21.

Acknowledgements

This work was supported by the Finnish Centre of Excellence Program of the Academy of Finland; the CoE in Biological Interactions 2012-2017 (#252411), by Academy of Finland grants #251106 (J.K.H.B) and #266879 (L-R.S.), by Jane and Aatos Erkko Foundation, by Medical Research Council #MR/N00065X/1 (D.I.S.) and by European Research Council under the European Union's Horizon 2020 research and innovation programme (649053 to J.T.H). The Oxford Particle Imaging Centre was founded by a Wellcome Trust JIF award (060208/Z/00/Z) and is supported by a WT equipment grant (093305/Z/10/Z). The Wellcome Trust Centre for Human Genetics is supported by Wellcome Trust Centre grant 090532/Z/09/Z. A.G. is funded by a Wellcome Trust 4-year PhD studentship (106274/Z/14/Z). The authors wish to thank Ms Heidi Pirttinen for assistance in the laboratory, Mr Petri Papponen for the help with TEM, and Dr. Alistair Siebert for cryo-TEM support in addition to Dr. Pietro Roversi, Professor Axel Brünger and Professor Frank DiMaio for their helpful discussion on model building and refinement in EM maps. In memory of Professor Jaana Bamford.

Author Contributions

E.L., S.M., and J.M. isolated FLiP and its host and performed the biological and genomic characterisations. A.G. and J.T.H. collected and analysed the cryo-EM images. L.D.C. performed the model building, density modification and structure refinement. D.I.S. and J.T.H. supervised the structural analysis. J.J.R. and L.R.S. designed and supervised the project. E.L., S.M., L.D.C. and J.T.H. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Data accessibility

The genome sequence of FLiP will be submitted to GenBank upon acceptance. The cryo-EM 3D reconstruction map of FLiP will be deposited with the Electron Microscopy Data Bank (EMDB) under accession code EMD-XXX. The modeled structure of the virus MCP shell will be deposited

539 with the Protein Data Bank (PDB) under the accession code XXXX. The authors declare no
540 competing financial interests.

541

Figure 1 | Circular ssDNA genome of FLiP. a, A graphic representation of the genome (9,174 nt, numbering starting from a unique *Eco*RII restriction site) with predicted open reading frames labeled and shown in blue. Those identified by proteomics to encode for structural proteins are indicated in purple. **b,** Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (16 % acrylamide) analysis of the structural proteins. Molecular weight standards (Fermentas #26619) are shown in lane 1. Major protein bands (lane 2), which were subjected to mass spectrometry, are indicated by an asterisk. N-terminal sequences determined by N-terminal protein sequencing are indicated.

Figure 2 | Three dimensional cryo-electron microscopy reconstruction of the FLiP virion. a, Surface rendering of the FLiP virion. The major capsid proteins are colored in yellow, green, blue, and red. Other structural components are colored in grey. Icosahedral 2-fold (ellipse), 3-fold (triangle) and 5-fold (pentagons) axes of symmetry are indicated. The triangulation number T=21 of the icosahedral lattice of the capsid is calculated as $T=h^2 + h\times k + k^2$ where lattice indices are h=4 and k=1 as indicated. The lattice is handed, and is right-handed (*dextro*) as indicated by the right turn (arrow) **b**, A section of virion density is shown from the area indicated in *a*. Spike structure, minor proteins (asterisk), outer (OL) and inner (IL) leaflet of the lipid bilayer and density corresponding to the ssDNA are indicated. **c**, The capsid is composed of 20 faces, each consisting a group of 10 pseudo-hexameric MCP trimers (ribbons). Each face is divided in three asymmetric units (one outlined), each consisting of three MCP trimers (1,2, and 4) and one chain from MCP trimer 3 (in addition to the base domain of the spike). **d–e**, Group-of-ten seen from the side (*d*) and below (*e*). Approximate footprints of the minor capsid proteins indicated in *b* are outlined.

Figure 3 | FLiP major capsid protein fold and structure-based phylogeny. a, The structure of FLiP major capsid protein (MCP) is shown as a ribbon representation, colored from the N-terminus (blue) to the C-terminus (red). The beta-strands in the N-terminal beta-barrel are labeled C–I and the beta-strands in the C-terminal beta-barrel B'–I'. Alpha-helices are labeled α, α' and α''. **b**, A schematic diagram to illustrate the topology of the double-beta barrel jelly-roll fold. The secondary structure elements are labeled and colored as in *a*. The N- and C-termini are labeled with the respective letters. **c**, Structure-based phylogenetic tree derived from alignment of MCPs (31) from FLiP, bacteriophage PM2 (pdb:2WOC), bacteriophage PRD1 (pdb:1HX6), Sulfolobus turreted icosahedral virus (STIV; pdb:2BBD), Paramecium bursaria chlorella virus 1 (PBCV-1; pdb:1M3Y), Vaccinia virus Western Reserve D13 (D13; pdb:2YGB), Sputnik virophage (pdb:3J26), and human adenovirus 5 (h-Ad5; 1P30) in addition to three other adenoviruses (2OBE, 1P2Z, 2INY; not shown). The region of each protein used in the structure-based alignment is colored in green. A generalized common core of the fold is shown as a C-alpha trace in the inset. Branches corresponding to viruses with dsDNA genomes are shaded in light blue. The branch corresponding to FLiP, the only characterized member the lineage with a ssDNA genome, is highlighted in red.

Figure 4 | Comparison of FLiP and PM2 capsid architecture. a, Surface rendered views of FLiP (this study) and PM2 (EMD-1082) virion structures are shown. Both maps have been low-pass filtered to 8-Å resolution. Major capsid protein (MCP) trimers are colored in yellow, blue, green and red, depending on their location in the icosahedral capsid. Other densities, including the spikes in the icosahedral vertices, are colored in gray. **b**, A central section of density is shown for the two virions. Coloring is as in *a*. Capsid-to-membrane contacts are indicated with arrowheads. In FLiP,

these densities have not been assigned. In PM2, they correspond to minor protein P3 (27). **c–d**, a ribbon representation of four different types of MCPs seen from outside (*c*) and inside (*d*) of the virion. In addition to the MCP, alpha-helices of the PM2 minor protein P3 interacting with the MCP (27) are shown in dark blue.