Primordial capsid and spooled ssDNA genome structures penetrate ancestral events of eukaryotic viruses

Anna Munke[1*#], Kei Kimura[2], Yuji Tomaru[3], Han Wang[1], Kazuhiro Yoshida[4], Seiya Mito[5], Yuki Hongo[6], and Kenta Okamoto[1*]

1. The Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden
2. Department of Biological Resource Science, Faculty of Agriculture, Saga University, Saga, Japan
3. Fisheries Technology Institute, Japan Fisheries Research and Education Agency, Hatsukaichi, Hiroshima, Japan
4. Graduate School of Agriculture, Saga University, Saga, Japan
5. Department of Biological Resource Science, Faculty of Agriculture, Saga University, Saga, Japan
6. Bioinformatics and Biosciences Division, Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Fukuura, Kanazawa, Yokohama, Kanagawa, Japan

*Corresponding authors

Corresponding author 1: anna.munke@desy.de

Corresponding author 2: kenta.okamoto@icm.uu.se

[#]Present address: Center for Free Electron Laser Science, Deutsches Elektronen-Synchrotron DESY, Hamburg 22607, Germany

Abstract

Marine algae viruses are important for controlling microorganism communities in the marine ecosystem, and played a fundamental role during the early events of viral evolution. Here, we have focused on one major group of marine algae viruses, the ssDNA viruses from the *Bacilladnaviridae* family. We present the capsid structure of the bacilladnavirus, *Chaetoceros tenuissimus* DNA virus type II (CtenDNAV-II), determined at 2.3 Å resolution. Structural comparison to other viruses supports the previous theory where bacilladnaviruses were proposed to have acquired their capsid protein via horizontal gene transfer from a ssRNA virus. The capsid protein contains the widespread virus jelly-roll fold, but has additional unique features; a third β-sheet and a long C-terminal tail, which are located on the capsid surface and could be important for virus transmission. Further, low-resolution reconstructions of the CtenDNAV-II genome reveal a partially spooled structure, an arrangement previously only described for dsRNA and dsDNA viruses.

Key words: algae virus, asymmetric reconstruction, CRESS-DNA, *Cressdnaviricota,* cryo-EM, *Bacilladnaviridae*, bacilladnavirus, diatom virus, evolution, genome structure, spooled genome

Introduction

Marine algae viruses prevail massively in the oceans and greatly affect the global ecosystem by causing mortality and lysis of microbial communities, releasing organic carbon and other nutrients back into the environment (the 'viral shunt'), thereby affecting the global oxygen level, and the marine nutrient and energy cycling (Fuhrman, 1999; Suttle, 2007; Wilhelm and Suttle, 1999). Algae blooms are global phenomena of dynamic appearance and disappearance of certain algae species. The recurring patterns of algae blooms are associated with the presence of marine diatom/algae viruses (Bratbak et al., 1993; Johannessen et al., 2017; Tarutani et al., 2000). The viruses typically have a very narrow host range, thus causing host-specific mortality and control of algae host populations (Brussaard, 2004), and are therefore attractive for being used as a microbial agent against harmful algae blooms (Nagasaki et al., 1999).

Another incentive for studying marine algal viruses is for penetrating early evolutionary scenarios of primordial eukaryotic viruses. Since unicellular marine organisms were the earliest eukaryotes on earth, they were presumably host of the most ancient viruses (Dolja and Koonin, 2018; Koonin et al., 2015). The extant unicellular organisms, such as unicellular algae, therefore likely retain genetic and structural characteristics from their ascendant (Koonin and Dolja, 2014; Munke et al., 2020). Structural comparison between so-called primordial viruses, such as diatom viruses, and the phylogenetically closely related viruses in the same evolutionary lineages can reveal acquired functional structures in viruses that infect higher eukaryotes such as human, animal and crop pathogens (Munke et al., 2020; Okamoto et al., 2020, 2016).

The *Bacilladnaviridae* viruses in this study are greatly involved in the discussions of both the global ecosystem model and early virus evolution. The *Bacilladnaviridae* viruses infect a diverse range of marine bloom-forming algae species (Kimura and Tomaru, 2015; Tomaru et al., 2013, 2011). These *Bacilladnaviridae* viruses are only infectious to their specific algae host. They lyse the host algae cells and reduce the number significantly during their propagation (Kimura and Tomaru, 2015; Tomaru et al., 2011, 2013). Understanding their host-specific transmission mechanism is the key to clarify how these ssDNA viruses affect the dynamics of bloom-forming algae species. Bacilladnaviruses carry a circular ssDNA genome of ~6 kb, which is partially double-stranded (~700-800 bp) and encodes three proteins; one coat protein, one replication protein (Rep), and a third protein with unknown function (Kimura and Tomaru, 2015). Until recently, bacilladnaviruses have been included in the informal group CRESS DNA viruses (for *circular Rep-encoding ssDNA viruses*), but this group has now formed the phylum *Cressdnaviricota* (Krupovic et al., 2020).

Viruses with different nucleic acid types have traditionally been organised into different classes, however an increasing number of evidence has emerged during the last decade suggesting that RNA-DNA recombination is much more prevalent than previously recognized (Diemer and Stedman, 2012; Kazlauskas et al., 2017; Krupovic, 2013; Tisza et al., 2020). In terms of bacilladnaviruses, Kazlauskas *et al* suggested based on sequence homology that their capsid proteins likely have been acquired through horizontal gene transfer (HGT) from an ancestral noda-like virus (Kazlauskas et al., 2017). This is not an unreasonable scenario considering the prevalence of noda-like viruses in the aquatic environment (Wolf et al., 2020). The icosahedral capsid of the *Bacilladnaviridae* viruses has adapted to pack a ssDNA genome instead of a ssRNA genome, however the lack of available capsid and genome structures of icosahedral ssDNA viruses have prevented addressing this theory.

Here, we present 3D reconstructions that reveal both the capsid and genome organization of the bacilladnavirus *Chaetoceros tenuissimus* DNA virus type II (CtenDNAV-II). An atomic model of the capsid protein could be constructed from the 2.3 Å resolution capsid structure. The capsid protein contains a jelly-roll fold with two additional extensions; a third seven-stranded β-sheet that is intertwined within the jelly-roll and C-terminal tails in two of the three proteins that form the

86 icosahedral asymmetric unit. Both extensions, which are unique to the bacilladnaviruses, are located
87 on the capsid surface and display a higher intraspecies sequence variability compared to other parts of
88 the capsid protein, suggesting a functional role during virus transmission. The virus transmission is
89 further discussed with support from a hemagglutination inhibition assay, adsorption test and sequence
90 data. The DALI program (Holm, 2020a) was used to demonstrate that the capsid of the
91 bacilladnaviruses indeed are structurally more similar to capsids of RNA viruses than to those of other
92 ssDNA viruses, corroborating the HGT theory in early virus evolution. In addition, the first genome
93 structure of a ssDNA virus presented here suggests a spooled genome packaging mechanism, which
94 has previously only been described for dsRNA and dsDNA viruses.

Results

Summary of structure determination

The structure of the CtenDNAV-II viron was determined using cryo-electron microscopy (cryo-EM). A Titan Krios microscope (Thermo Fisher Scientific) equipped with a K2 Summit direct electron detector (Gatan) (see Materials and Methods) was used to record micrographs (Supplementary Fig. 1A) and confirmed mature virons containing a DNA genome. The particles had a diameter of approximately 35 nm and appeared to be morphologically reproducible. The three-dimensional (3D) reconstruction of CtenDNAV-II was performed with RELION (Scheres, 2012). The capsid was reconstructed by imposing icosahedral symmetry (I4) and using 33,507 particles, whereas the genome structure was reconstructed without imposing any symmetry (C1) and using 21,559 particles. The overall resolution of the final maps were estimated using the "gold standard" Fourier shell correlation (FSC) 0.143 criterion (Henderson et al., 2012; Scheres and Chen, 2012) to 2.3 Å for the capsid (Supplementary Fig. 1C) and 13 and 23 Å for the two genome reconstructions (Supplementary Fig. 2). The local resolution of the capsid reconstruction was distributed between 2.3 and 9.1 Å and estimated using ResMap (Kucukelbir et al., 2014) (Supplementary Fig. 1D). An atomic model was built, refined, and validated according to our cryo-EM map. Data acquisition and processing, refinement, and validation statistics are summarized in Table 1.

Table 1. Data collection, processing and refinement statistics

Data collection and processing	**Capsid EMD-12554**	**Genome – outer layer EMD-12555**	**Genome – core EMD-12556**
Magnification	140,000		
Voltage (kV)	300		
Defocus range (μm)	-1.00 to -3.00		
Microscope	Titan Krios		
Camera	K2		
Total electron dose ($e^-/Å^2$)	37		
Pixel size (Å)	1.06		
Final particle number	33,507	21,559	
Symmetry imposed	I4	C1	
Map resolution (Å) (FSC=0.143)	2.3	13	23
Map resolution range (Å)	2.3-9.1	N/A	
Map sharpening B-factor ($Å^2$)	-70.5027		
Model building and refinement			
PDB	XXXX		
Model composition			
Chains	3		
Atoms	13933		
Protein residues	906		
B-factors ($Å^2$) (min/max/mean)	4.67/33.32/12.03		
Bonds (RMSD)			
Length (Å)	0.005		
Angle (°)	0.724		
Validation	1.79		
MolProbity score	1.79		
Clash score	5.60		
Ramachandran plot (%)			
Outliers	0.00		
Allowed	4.01		
Favoured	95.99		
Rotamer outliers (%)	1.90		
Cβ outliers (%)	0.00		
CaBLAM outliers (%)	1.46		
CC (overall)	0.84		
EMRinger score	6.16		
d99 masked (full/half1/half2)	2.2/3.3/3.3		
d99 unmasked (full/half1/half2)	2.1/3.1/3.1		

The capsid protein displays different conformations within the asymmetric unit

The CtenDNAV-II capsid displays T = 3 symmetry, i.e. 180 capsid protein protomers assemble such that the asymmetric unit comprises 3 capsid subunits in 3 quasi-equivalent positions termed A, B, and C (Fig. 1A). For subunit A residues 64-72 and 77-371 were modelled, for subunit B residues 64-371 were modelled, and for subunit C residues 64-357 were modelled. The C-terminus of the A and B subunits form long tails that end on the capsid surface around the 3- and 5-fold axes respectively. Here, the last 19 residues could not be modelled (Fig. 2C). Additional density was however visible in the map when the contour level was decreased (Supplementary Fig. 2), indicating that the C-terminus forms flexible protrusions on the capsid surface. Another 14 residues were unmodelled in the C-terminus of the C subunit (red rectangle in Fig. 2C), nevertheless it is clear that, in contrast to the A- and B-subunits, the C-terminus is directed towards the capsid interior (Fig. 1C). Unmodelled density is observed on the inside of the capsid below the C subunits (Supplementary Fig. 3), which presumably originates from the C-terminus of the C subunit, since corresponding unmodelled density was not visible below the A and B subunits (Supplementary Fig. 3A). Different conformations and/or flexibility were also observed between the three subunits in the N-termini (Fig. 1C), which are all located on the inside of the capsid. The first 63 residues of the N-terminus could not be modelled in any of the three subunits (Fig. 2C). The modelled termini and corresponding cryo-EM map are displayed in Fig. 1D. Apart from the termini, the three subunits have very similar structures (Fig. 1C). In addition, unmodelled density, presumably originating from an ion or small molecule, was visible in the interface of the three subunits that constitute one protomer (Supplementary Fig. 4). The density is surrounded by six arginine residues (R86 and R272), which is an unusual finding among viruses. Similar arginine rings have however been described for the HIV virus, in which the arginines interact with polyanions such as nucleoside triphosphates (dNTPs) and myo-inositol hexaphosphate (IP6) (Mallery et al., 2018).

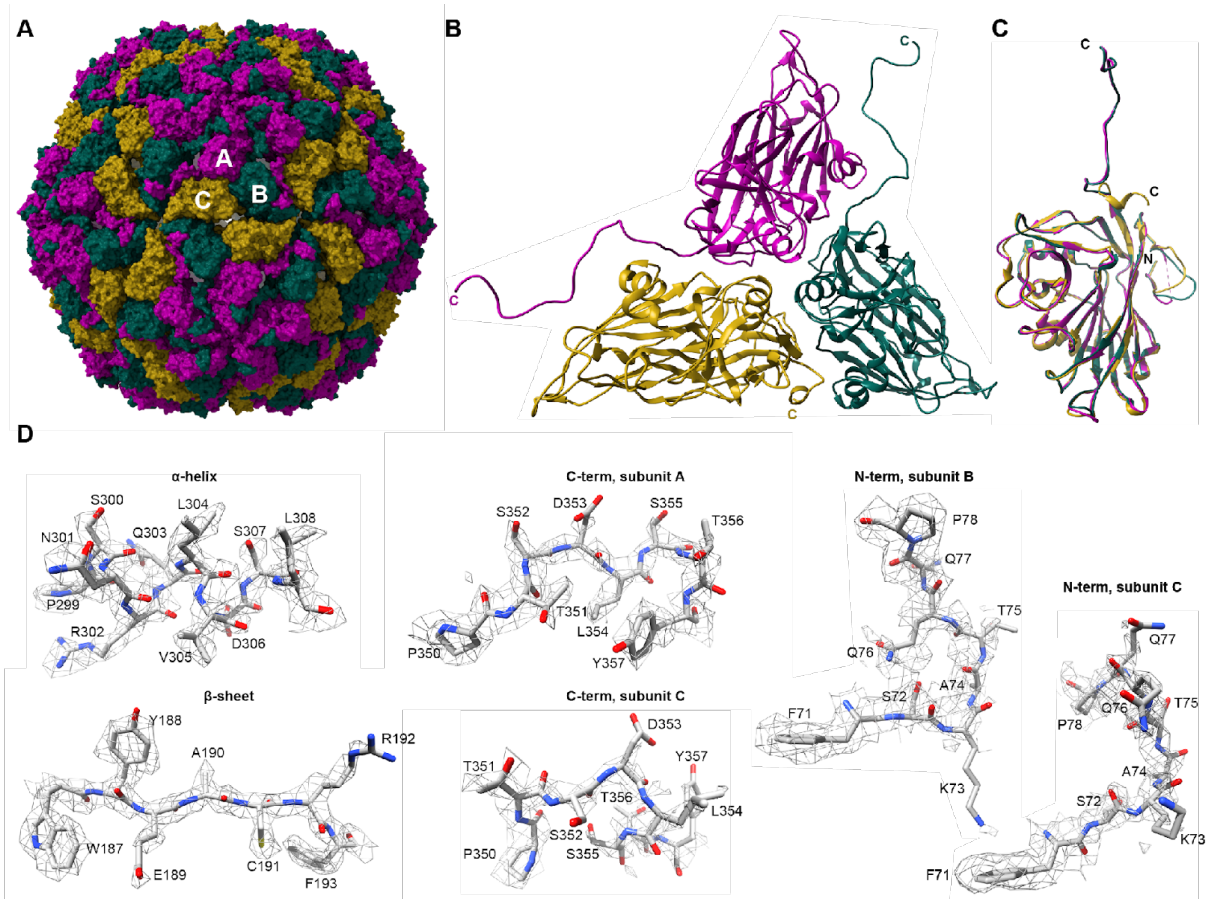


Figure 1. Atomic model of the CtenDNAV-II capsid. The three subunits, A, B and C, are coloured purple, green and yellow, respectively, in panel A-C. (**A**) The entire capsid rendered with a surface representation viewed down an icosahedral two-fold axis. (**B**) The secondary structure of one single asymmetric unit viewed from the outside. (**C**) Superimposition of the three subunits. (**D**) Refined side chains of representatives of secondary structural elements and areas that differ between the three subunits.

The capsid protein jelly-roll fold has unique features

The canonical viral jelly-roll consists of eight anti-parallel β-strands that are named from B to I and arranged in two four-stranded sheets (BIDG and CHEF). The loops connecting each strand are named BC, CD, etc. (Harrison et al., 1978; Rossmann et al., 1985). For CtenDNAV-II, the two sheets are formed by strands BIDD'G and C''C'CHEF, respectively (Fig. 2A-C), thus containing three additional strands (D', C' and C'') compared to the standard viral jelly-roll fold. In addition, a third antiparallel β-sheet with seven strands is intertwined within the jelly-roll, i.e. strands from the third sheet are formed by extensions of loops CD, EF and GH of the jelly-roll. The third sheet, which is located on the capsid surface, is thus composed of two C-strands (C''' and C''''), two E-strands (E' and E''), and three G-strands (G', G'' and G''') (Fig. 2B). In conclusion, the jelly-roll fold of CtenDNAV-II has three unique features: three additional strands, an extra surface-exposed β-sheet, and C-terminal tails in subunit A and B.

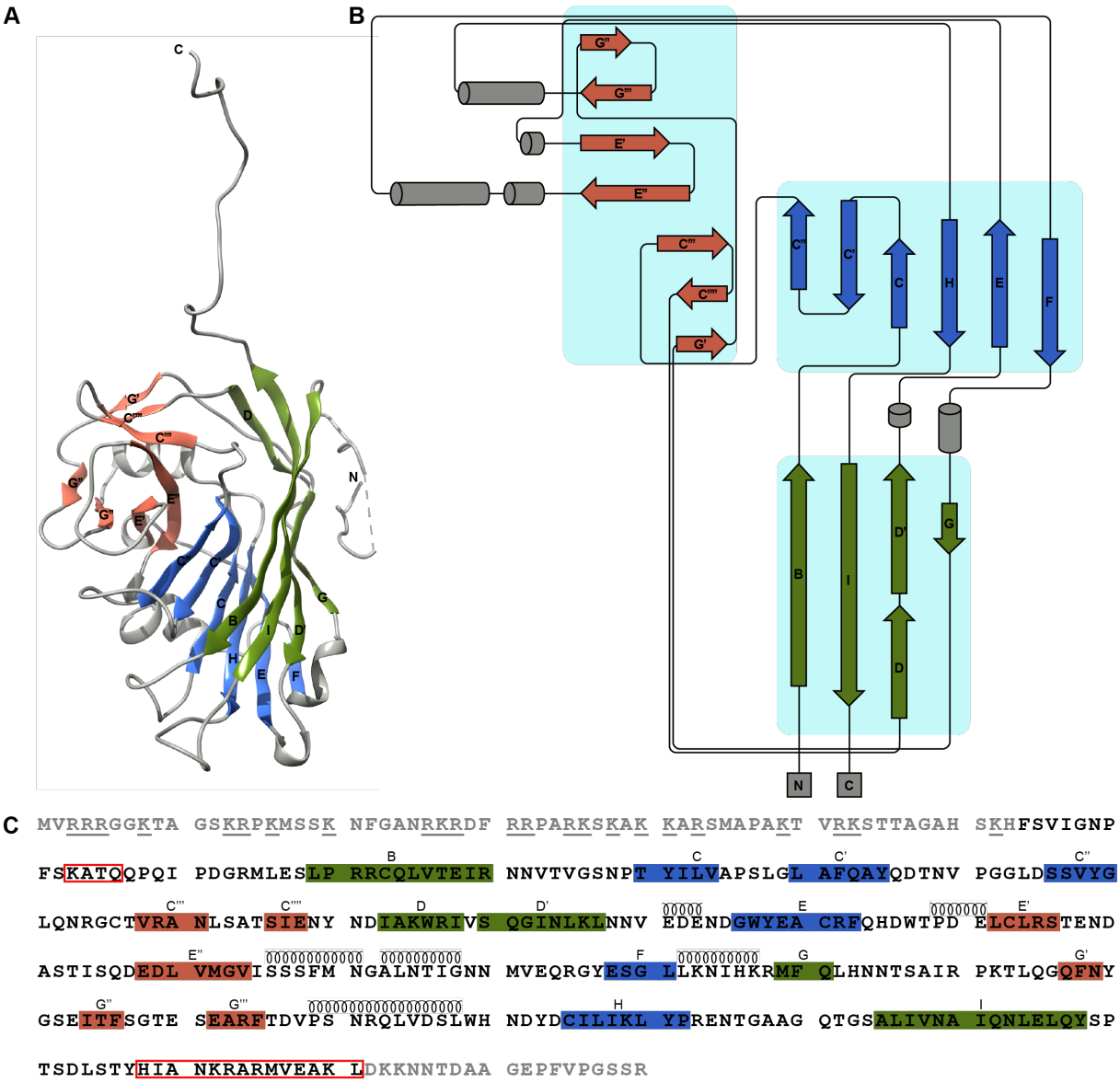


Figure 2. Capsid protein topology and structural organisation. The β-strands are coloured according which β-sheet they belong to. The β-strands within the green and blue β-sheets are named alphabetically according to the conventional jelly-roll fold nomenclature (B to I) (**A**) The secondary structure of subunit A. (**B**) Schematic diagram of the secondary structure. (**C**) The amino acid sequence of the CtenDNAV-II capsid protein, starting from residue 1. Each line has 70 residues and is further subdivided into blocks of 10 residues by spaces within the sequence. The residue numbering is the same as in PDB entry 7NS0. Modelled and unmodelled residues are coloured black and grey, respectively. Residues highlighted with red rectangles were unmodelled in one of the subunits: F73-P76 in subunit A and H358-L371 in subunit C. The assigned secondary structure is shown schematically above the sequence. The underlined residues in the unmodelled N-terminus indicate positively charged residues.

Capsid proteins from CtenDNAV-II and ssRNA viruses are similar

Previous structures of so-called CRESS DNA viruses (phylum *Cressdnaviricota*) (Krupovic et al., 2020) include viruses from families *Ciroviridae* (eg. 3R0R and 5ZJU) and *Geminiviridae* (eg.6F2S and 6EK5). The capsid proteins in viruses from these two families also contain a jelly-roll domain, but lack the third surface-exposed β-sheet and N-terminal tail found in CtenDNAV-II. A search with the

DALI program (Holm, 2020a) revealed that the CtenDNAV-II capsid protein is more similar to capsid proteins of ssRNA viruses than to ssDNA viruses (Fig. 3A), corroborating the previous HGT theory by Kazlauskas *et al* (Kazlauskas et al., 2017). The closest ssDNA virus is that of Beak and feather disease virus (*Circoviridae*), which ends up on 11th place (z-score 8.4) behind ten RNA viruses. Highest similarity is found between ssRNA viruses from families *Carmotetraviridae*, *Alphatetraviridae* and *Nodaviridae* that resulted in z-scores of 14.4-15.1 (Fig. 3A). For interpretation of the results from DALI see Holm (2020) (Holm, 2020b). Superimpositions between capsid proteins from CtenDNAV-II and representatives from each family show that those RNA viruses with highest z-score all have surface projections in addition to the jelly-roll, albeit with different folds, that occupy similar positions as the third β-sheet of CtenDNAV-II (Fig. 3B). However, in none of the other viruses the projection is as intertwined with the jelly-roll as for CtenDNAV-II, and instead the protrusion often forms a separate domain with a linker region, such as the Immunoglobulin (Ig)-like domain of Providence virus (*Carmotetraviridae*) (left most panel in Figure 3B) (Speir et al., 2010). The C-terminal surface protruding tail in subunit A and B of CtenDNAV-II is a unique feature that is not observed in any of the other virus families (Fig. 3B).

In addition to the C-terminal tail and β-sheet surface projection, superimpositions performed by the DALI server also revealed non-conserved loops (HI and C'C'') from the jelly-roll (Supplementary Fig. 5A, B). These areas are all located on the capsid surface clustered together around the five- and three-fold axes (Supplementary Fig. 5C).

A

PDB	Name	Family	Genome	z-score	rmsd	Residues
2qqp	Providence virus*	*Carmotetraviridae*	ssRNA	15.1	2.9	209
3s6p	Helicoverpa Armigera Stunt virus*	*Alphatetraviridae*	ssRNA	14.9	2.9	208
4ftb	Flock House virus*	*Nodaviridae*	ssRNA	14.9	3.1	201
2bbv	Black Beetle virus	*Nodaviridae*	ssRNA	14.7	2.9	199
1ohf	Nudaurelia capensis omega virus	*Alphatetraviridae*	ssRNA	14.5	3.0	207
1f8v	Pariacoto virus	*Nodaviridae*	ssRNA	14.5	3.2	204
1nov	Nodamura virus	*Nodaviridae*	ssRNA	14.4	3.1	202
3ide	Infectious pancreatic necrosis virus*	*Birnaviridae*	dsRNA	12.1	3.0	192
2df7	Infectous brusal disease virus	*Birnaviridae*	dsRNA	12.1	3.3	198
2izw	Ryegrass mottle virus*	*Solemoviridae*	ssRNA	8.8	3.3	162
5zju	Beak and feather disease virus*	*Circoviridae*	ssDNA	8.4	4.1	155

B

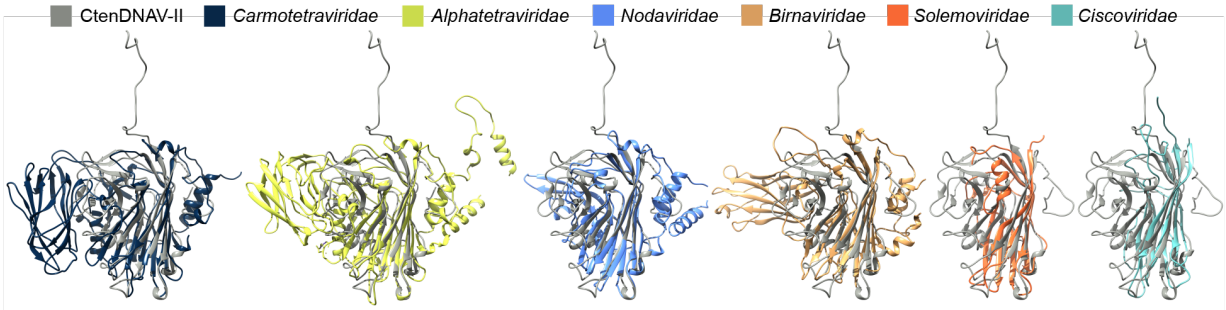
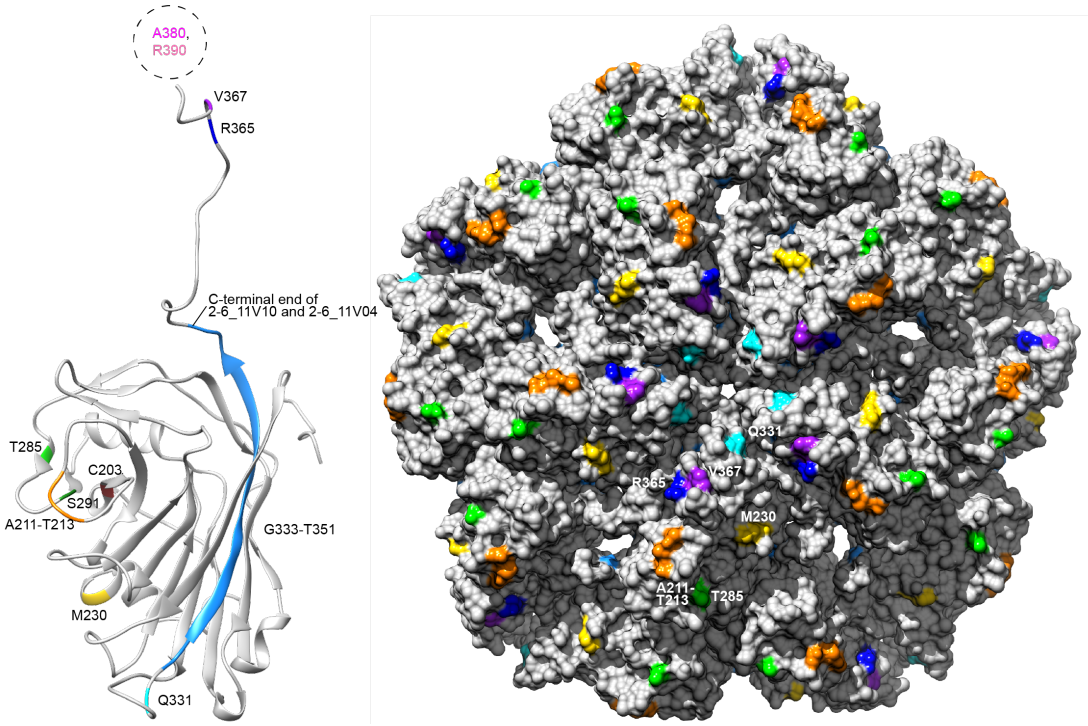


Figure 3. Structural comparison of the CtenDNAV-II capsid protein with corresponding proteins from other viruses. (**A**) DALI-based structural comparison where the top 11 unique viruses (i.e., not unique PDB entries) with highest z-score are listed. Viruses indicated with * were used for the superimpositions in (B). For each comparison the z-score, rmsd and the number of residues that were used for the alignment are listed. (**B**) Superimpositions of the CtenDNAV-II capsid protein (gray) with representatives from each family listed in (A).

206 **Intraspecies sequence variability in surface-exposed regions**

207 Since the presented structure is the first of a virus from the *Bacilladnaviridae* family and its most
208 structurally close-related viruses are ssRNA viruses (Fig. 3), we sequenced (see Supplementary
209 methods) 15 CtenDNAV-II strains with the anticipation of improving our understanding of the
210 structure. The alignment (Supplementary Fig. 6) revealed 11 areas with mutations, of which 9 are
211 surface-exposed primarily on the C-terminal tail (R365, V367, A380 and R390) or beta-sheet
212 projection (A211-T213, T285 and S291) (Fig. 4). Thus, the mutated areas correlate with the unique
213 structural features for the CtenDNAV-II capsid protein that was identified from the superimpositions
214 with the structurally closely related viruses (Fig. 3B and Supplementary Fig. 5A). Interestingly, two
215 strains, 2-6_11V10 and 2-6_11V04, shared the same C-terminus sequence, which were completely
216 different from the other 13 strains (Fig. 4 and Supplementary Fig. 6).
217



218
219 **Figure 4. Location of mutations identified among CtenDNAV-II strains.** Subunit A (left), and a
220 pentamer viewed from the outside looking down the five-fold axis (right) with mutations having the
221 same colour code as in Supplementary Figure 6. Two mutations (A380 and R390), in dashed circle,
222 are located in the unmodelled region of the C-terminus.
223

224 ***Chaetoceros tenuissimus* viruses employ different transmission strategies**

225 Sialic acids, acidic sugars that terminate glycan chains on the cell surface, are one of the common viral
226 receptors, and well-described for vertebrate viruses such as influenza, corona and polyoma. The
227 existence of sialic acids in algae and the discussions on their importance for algae virus interactions is
228 however more recent (Fulton et al., 2014; Munke et al., 2020; Wagstaff et al., 2018). Wagstaff and co-
229 workers could demonstrate that *Prymnesium parvum* is capable of *de novo* synthesis of the deaminated
230 sialic acid Kdn, and through bioinformatics analysis, that sialic acid biosynthesis is widespread among
231 microalgae. However, we failed to detect homologous sequences in *C. tenuissimus* (Accession
232 numbers BLLK01000001-BLLK01000085) for any of the synthesases identified in Wagstaff *et al*. To
233 further corroborate our findings, we performed a classical hemagglutination inhibition (HI) assay (see
234 Supplementary methods) to investigate the sialic acid-binding capability of CtenDNAV-II. An RNA
235 virus (CtenRNAV-II) (family *Marnaviridae*) that infects the same diatom host was also tested as a

representative of another major group of diatom viruses. The HI test demonstrated that CtenRNAV-II shows partial HI activity, especially at high concentrations (no and 2 times dilution), while CtenDNAV-II did not show any HI activity (Supplementary Fig. 6). In conclusion, both the synthesase sequence analysis and HI assay indicate that the transmission of CtenDNAV-II is not sialic acid-mediated.

Chaetoceros tenuissimus viruses display different host cell affinities

The adsorption kinetics of CtenDNAV-II and CtenRNAV-II were examined with two diatom species including three culture strains (see Supplementary methods). The adsorption of CtenDNAV-II was not detected for *Chaetoceros socialis* f. *radians* NIES-3713 (Supplementary Fig. 8). This result is reasonable, since the DNA virus does not lyse *C. socialis* (Kimura and Tomaru 2015). Although CtenDNAV-II lyses both *C. tenuissimus* strains NIES-3714 and NIES -3715 (Kimura and Tomaru 2015), significant adsorption of the virus was only observed with the former strain (Supplementary Fig. 8). As for CtenRNAV-II, significant adsorption was detected for *C. tenuissimus* strain NIES-3715, but not to other diatom species and strains (Supplementary Fig. 8). A previous study reported that CtenRNAV-II lyses *C. tenuissimus* strain NIES-3715, but not NIES-3714, and has a low lytic activity on *C. socialis* f. *radians* (Kimura and Tomaru 2015). Therefore, the present result would be consistent with the phenotypic relationship and suggest that the specific adsorption of CtenRNAV-II is determined at host strain level. In conclusion, the degree of the virus and diatom host cell affinities would be determined by combinations of the virus species and host strain level.

CtenDNAV-II genome structure is partially spooled

The genome organization of CtenDNAV-II is shown in Figure 5. The CtenDNAV-II genome consists of an ordered outer layer (Fig. 5B and E) that is partially spooled around a disordered core (Fig. 5C). The outer layer (EMDB-12555) was reconstructed at 13 Å by masking out the capsid, while the core (EMDB-12556) was reconstructed at 23 Å by using a spherical mask (see Materials and Methods for details). The reconstruction of the outer genome layer displays a coil of three turns that are positioned between the icosahedral 5-fold axes (Fig. 5E). On each side of the three turns are additional DNA fragments that do not follow the same spooling arrangement (Fig. 5B). The spooled genome packaging has previously only been described in viruses with double stranded genomes (Ilca et al., 2019; Liu et al., 2019), whereas ssRNA viruses instead have a branched network formed by the genome secondary structure (Dai et al., 2017; Gorzelnik et al., 2016; Koning et al., 2016). The distance between the parallel turns are about 28 Å (Fig. 5E), which is about the same distance as what has been described for dsRNA viruses (Ilca et al., 2019). Below each icosahedral 5-fold axis of the capsid the reconstruction of the outer genome layer displays protrusions towards the capsid (Fig. 5B and E). However, these protrusions are not necessarily true at every position and could partially be artefacts from the reconstruction, which is supported by the fact that some of the protrusions loose connections to the remaining outer layer when the contour level is decreased while other protrusions remain intact. Connections between the outer genome layer and the core seem to be confined to two specific areas on opposite sides of the outer layer (Supplementary Fig. 9). Supplementary Movie 1 shows the asymmetric reconstruction in relation to the capsid.

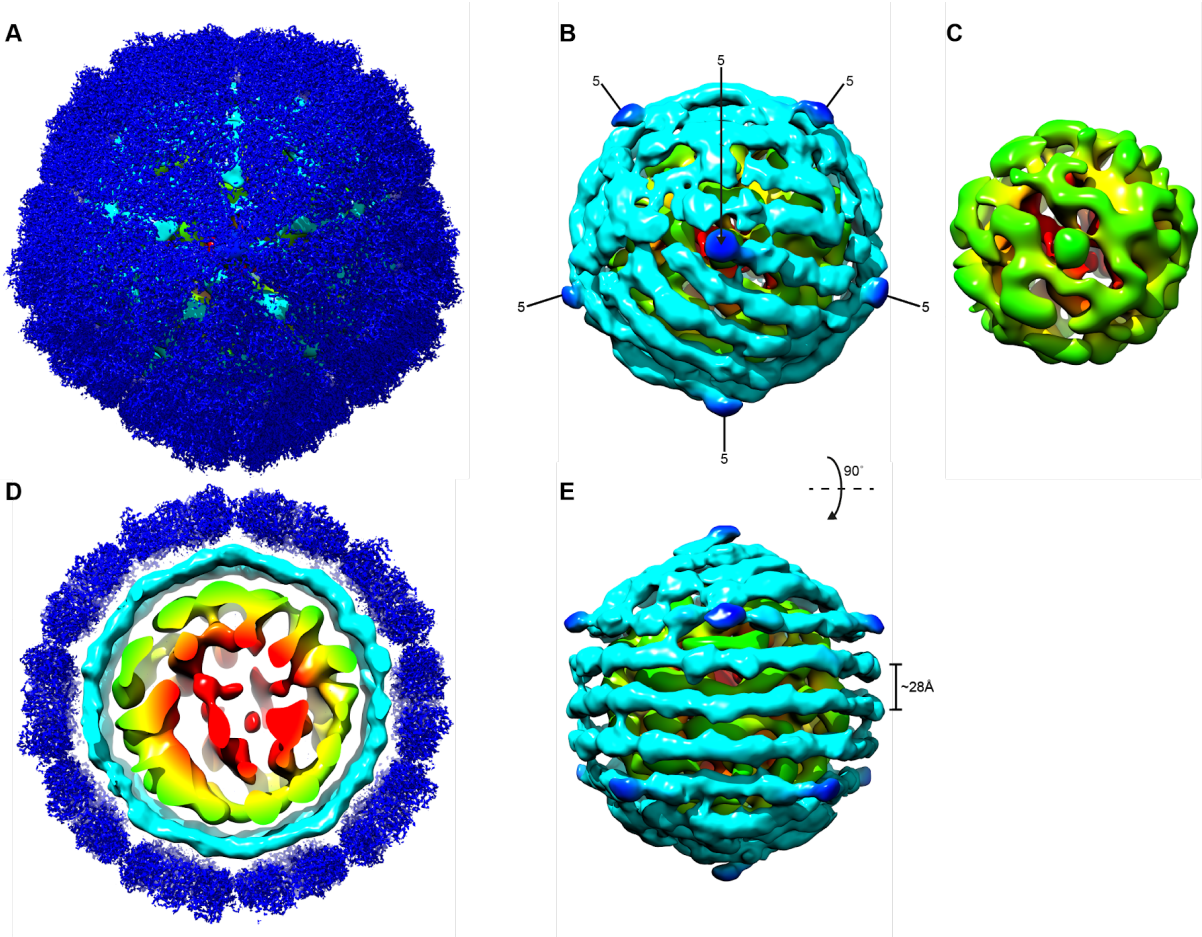


Figure 5. Asymmetric reconstructions of CtenDNAV-II radially coloured from red to blue. (**A**) The capsid (blue) reconstructed with icosahedral symmetry and the asymmetric reconstructions sighted below viewed down the icosahedral five-fold axis. (**B**) Same view as (A) but with the icosahedral reconstruction removed. A protrusion (blue) from the outer genome layer (cyan) is visualised at each five-fold axis (labelled as 5). (**C**) Same view as in (A) and (B) but with the icosahedral and outer layer reconstructions removed. (**D**) A thin slice of the reconstructions shown in (A). (**E**) The picture in (B) has been rotated 90°. The resolution (FSC=0.143) of the outer layer and core are 13 Å and 23 Å, respectively. The FSC curves are shown in Supplementary Figure 2.

Discussion

Transmission mechanisms of *Chaetoceros tenuissimus* viruses

This paper describes the first structure (Figs. 1-2) of a *Bacilladnaviridae* virus, which is one of the major groups of marine bloom-forming algae viruses. The CtenDNAV-II capsid protein contains the widespread virus fold, the jelly-roll, but has in addition a third β-sheet and a C-terminal extension, both of which form surface domains in close proximity to each other (Supplementary Fig.5) on the capsid surface and that to our knowledge are unprecedented in other jelly-roll fold viruses determined to date (Fig. 3, Supplementary Figs. S2 and S5). We have previously determined the structure of a *C. tenuissimus* infecting RNA virus (family *Marnaviridae*), which has a primordial surface loop and lacks the receptor-binding motif found in other structurally and phylogenetically related animal viruses (Munke et al., 2020). This implies that these host-specific algae viruses have unique receptor-binding mechanisms. The molecular details on the infection mechanisms of diatom viruses are unknown. However, a putative transmission route could exist through interactions with host sialic acids, since sialic acid biosynthesis has been shown to be widespread among microalgae (Wagstaff et al., 2018). In addition, many structurally related jelly-fold viruses interact with sialic acids for cell entry, such as polyomaviruses via their conserved surface loops (Ströh et al., 2020). Likewise, the C-terminal P-domain of the betanodavirus capsid protein interacts with sialic acids, and its deletion affects virulence and host specificity (Liu et al., 2005; Moreno et al., 2019). The combined results of no homologous sequences in *C. tenuissimus* for any of the synthesases identified in Wagstaff *et al* (Wagstaff et al., 2018), with the lack of HI activity (Supplementary Fig. 6), strongly refute a sialic acid interaction between CtenDNAV-II and its host. This implies that the CtenDNAV-II virus employs an alternative receptor and transmission strategy. The RNA virus infecting the same diatom host (CtenRNAV-II) did however interact weakly with the mammalian sialic acid, which could indicate that microalgae might have sialic acids or structurally related acidic sugars that have not yet been described in the literature. The CtenRNAV-II virus is phylogenetically closely related to *Picornaviridae* viruses, some of which use sialic acids as their entry receptor (Zocher et al., 2014). The different receptor-binding strategies employed by the DNA and RNA virus could explain their different host cell affinities (Supplementary Fig. 8).

Nevertheless, the CtenDNAV-II capsid protein has two unique surface features, the C-terminal tail and β-sheet projection, compared to the most structurally close-related viruses (Fig. 3). These extensions from the jelly-roll fold were in addition found to have an increasing number of mutations between different virus species compared to other parts of the structure (Fig. 4 and Supplementary Fig. 6). Host specificity of algae viruses have previously been correlated with surface-exposed amino acid substitutions (Nagasaki et al., 2005). Future research on CtenDNAV-II strains and their infection patterns might reveal similar relationships. The locations of the mutations do however make the β-sheet and C-terminal extensions putative interaction candidates.

Structural insights on eukaryotic jelly-roll fold viruses in evolution

About 30% of all viruses adopt the single jelly-roll fold (Krupovic and Koonin, 2017), and the results presented herein provide structural insights on the evolution of this fold. Our findings corroborate the theory that the jelly-roll capsid proteins of ssDNA, ssRNA and dsRNA viruses originate from the same virus with a primordial jelly-roll fold (Fig. 3) (Kazlauskas et al., 2017; Krupovic and Koonin, 2017). According to the HGT theory, the capsid protein of CRESS-DNA viruses could have been acquired from the ssRNA protist viruses in early evolutionary events (Kazlauskas et al., 2019, 2017). Indeed, the capsid protein of CtenDNAV-II is structurally more similar to ssRNA and dsRNA viruses than to other CRESS-DNA viruses (Fig. 3). Because the rolling-circle replication proteins of CRESS-DNA viruses are polyphyletic, the capsid protein genes have probably been acquired from ssRNA

viruses on several independent occasions and at different time-points in evolution (Krupovic, 2013). Similarly, icosahedral dsRNA viruses are thought to have originated from ancestral ssRNA nodaviruses and tetraviruses (Coulibaly et al., 2005). Yet another example includes the large CRESS-DNA virus FLiP (Flavobacterium-infecting, lipid-containing phage) that shows a dsDNA adenovirus-like capsid structure (Laanto et al., 2017), which potentially links CRESS-DNA and dsDNA viruses. To put our new structural insights into a broader context, we have updated the putative emergence of the ssDNA, ssRNA and dsRNA viruses (Fig. 6).

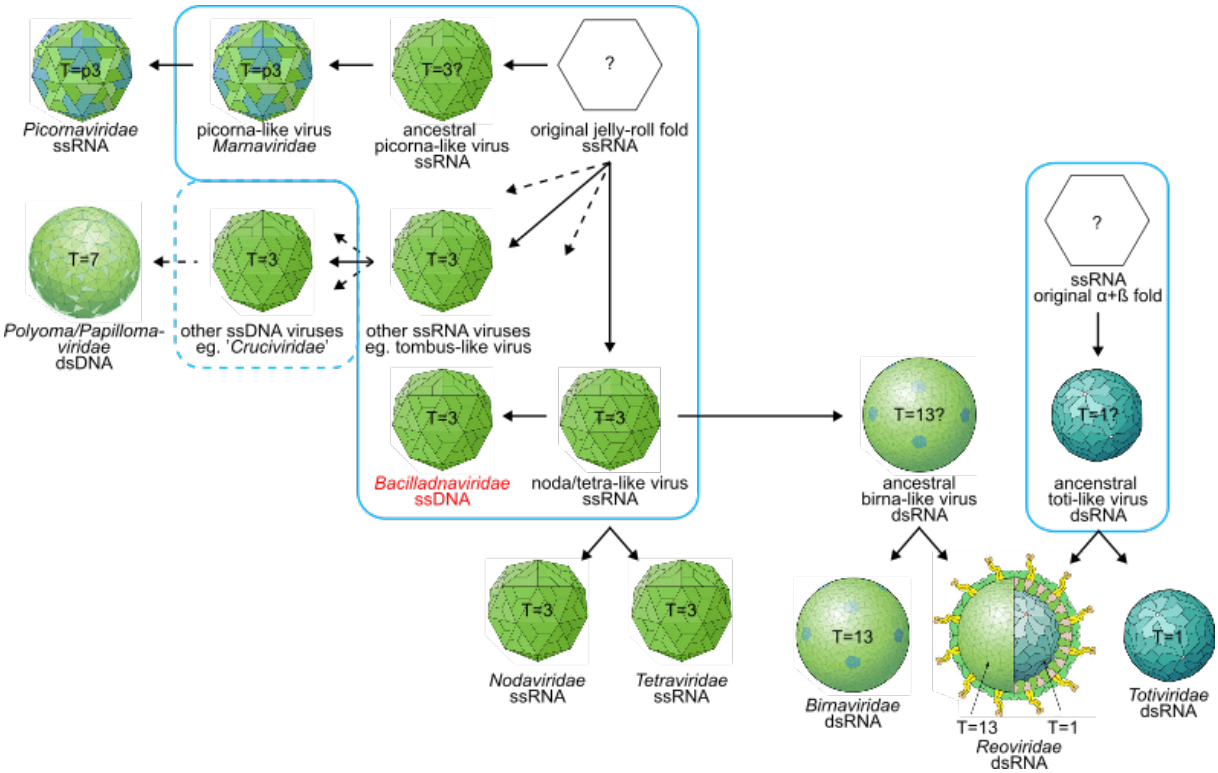


Figure 6. A scheme depicting the putative emergence of virus capsids carrying different nucleic acid genomes. The figure illustrates based on current knowledge the most relevant virus groups and the evolutionary relationship between their virus capsids. Events that likely took place at an early stage of evolution in ancient algal pools are circled in blue. The dashed blue line represents that ssRNA and ssDNA virus recombination has occurred at several independent occasions, at different time points in evolution. The figure is based on results described in this paper as well as previous results and discussions (Chiba et al., 2020; Coulibaly et al., 2005; Diemer and Stedman, 2012; Kazlauskas et al., 2019; Krupovic, 2013; Munke et al., 2020; Okamoto et al., 2020, 2016). The viron pictures were derived from ViralZone, SIB Swiss Institute of Bioinformatics (Hulo et al., 2011) (https://viralzone.expasy.org/) licensed under a Creative Commons Attribution 4.0 International Licens.

Animal and plant ssRNA, dsRNA and CRESS-DNA viruses with icosahedral capsids that contain the jelly-roll fold should have acquired additional structural motifs for infecting their hosts. The primordial jelly-roll capsid protein structure of the CtenDNAV-II highlights the acquired features of these viruses that infect multicellular eukaryotes. Examples of such acquisitions are the Ig-like domain of tetraviruses and the P-domain of nodaviruses and birnaviruses (Fig. 3). These structures have been described as putative receptor-binding domains of those viruses; however, an exhaustive characterization still remains. The external Ig-like domain in tetraviruses (family *Carmotetraviridae*

359 and *Alphatetraviridae*) is believed to function in receptor binding to host cells (Helgstrand et al., 2004;
360 Penkler et al., 2016). Some *Nodaviridae* viruses, so-called gammanodaviruses, have an external P-
361 domain that is important for virus attachment and entry (Ho et al., 2018), whereas the capsid proteins
362 of other nodaviruses retain primordial features that are structurally much more similar to the
363 CtenDNAV-II capsid (Johnson et al., 2001) (Fig. 3). The dsRNA *Birnaviridae* viruses display an
364 external highly antigenic P-domain that is likely involved in the initial attachment to the host cells
365 (Coulibaly et al., 2005; Garriga et al., 2006). The CRESS-DNA polyomaviruses possess a conserved
366 jelly-roll fold VP1 protein, and has acquired extra minor proteins VP2 and VP3 that has an ER-to-
367 cytosol penetration function during cell entry (Inoue and Tsai, 2011). Several studies report on the
368 presence of a large number of diverse unclassified animal CRESS-DNA viruses, including putative
369 human pathogenic viruses (Ng et al., 2015; Phan et al., 2015; Tisza et al., 2020). Structural
370 comparison of their capsid proteins with that of CtenDNAV-II could reveal the acquired features of
371 these animal CRESS-DNA viruses.

372 **A spooled genome requires structural adaptations of the capsid**

373 The technological and computational improvements in cryo-EM structure determination have led to an
374 increasing number of virus genome structures during the last decade. However, determining the
375 genome structure within the symmetric icosahedral capsid still remains challenging, especially for
376 single-stranded genomes (Koning et al., 2016). By employing a previously described method of
377 subtracting the contribution of the capsid (Ilca et al., 2019), we were here able to demonstrate the first
378 entire genome structure of a ssDNA virus (Fig. 5 and Supplementary Movie 1).
379 Previous cryo-EM studies have revealed three types of genome organization for non-enveloped
380 icosahedral viruses. The dsDNA and some dsRNA viruses such as Bacteriophage Φ6 forms spooled
381 genome structures with certain interfilament spacing (Ilca et al., 2019; Liu et al., 2019; Wang et al.,
382 2019). Segmented dsRNA reoviruses instead form non-spooled or partially-spooled genomes with
383 pseudo-D3 symmetry that interact with the RNA-dependent RNA polymerase (RdRp) (Ding et al.,
384 2019; Liu and Cheng, 2015; Wang et al., 2018; Zhang et al., 2015). The third variant is found in
385 *Leviviridae* viruses where the secondary structure of their ssRNA genomes form a branched network
386 of stem-loops (Dai et al., 2017; Gorzelnik et al., 2016; Koning et al., 2016), however it remains
387 unknown if a similar organisation is found in other ssRNA viruses.
388 The structure presented here is the first entire genome structure for a ssDNA virus. The genomes of
389 ssDNA viruses have the capability to form biologically functional secondary structures similar to
390 ssRNA viruses (Muhire et al., 2014) and since the capsid gene of bacilladnaviruses has been
391 horizontally transferred from ssRNA viruses, a similar branched structural network is imaginable also
392 for ssDNA viruses. However, the structure of the CtenDNAV-II genome is much more similar to those
393 of the spooled dsDNA or the partially spooled dsRNA genomes (Fig. 5 and Supplementary Movie 1).
394 Two major questions arise for the genome packaging of bacilladnaviruses. The first question is how to
395 stabilize the spooled ssDNA genome without a secondary structure organization? Potentially, a partial
396 secondary structure might exist, and could together with the dsDNA segment that exist in
397 bacilladnaviruses (Kimura and Tomaru, 2015; Tomaru et al., 2011, 2013), be sufficient for stabilizing
398 the ssDNA genome inside the capsid. Another possibility is that the dsDNA gene of bacilladnaviruses
399 help stabilizing the ssDNA without having any secondary structure. The second question is how the
400 ssRNA virus-like capsid protein was adapted to packing the dsDNA/dsRNA-like spooled genome of
401 the CtenDNAV-II? The spooled genome structures of the dsDNA and dsRNA viruses are packaged by
402 flexible interactions between the capsid protein and the genome. The interactions are mediated by
403 small contacts with hydrophobic and/or positively-charged amino acid residues of their capsid proteins
404 (Ding et al., 2019; Ilca et al., 2019; Liu and Cheng, 2015, 2015; Wang et al., 2019, 2018; Zhang et al.,
405 2015). In contrast, the secondary structure of the (+)ssRNA viral genome binds specifically to the

406 capsid (Dai et al., 2017; Gorzelnik et al., 2016; Hesketh et al., 2018; Koning et al., 2016). Our
407 hypothesis is that the unmodelled N-terminus, and possibly the C-terminus of the C-subunit, which
408 both possess numerous positively charged amino acids, interact with the ssDNA genome, in a flexible
409 and non-specific manner similar to dsDNA and dsRNA viruses. In contrary, the structurally related
410 tetra- and nodaviruses seem to interact more specifically with their ssRNA genome, since small RNA
411 segments (Speir et al., 2010) or parts of the genome (Johnson et al., 2001) have been revealed even
412 when the icosahedral symmetry has not been broken during the structure determination. This could
413 explain why the comparison between CtenDNAV-II and the RNA viruses revealed different structural
414 features of their capsids' interior, where the CtenDNAV-II was unmodelled and the ssRNA viruses
415 had additional α-helices (Fig. 3). A 7 nucleotide-stem loop gene fragment of the ssDNA geminivirus
416 has been reported to interact with its capsid (Hesketh et al., 2018), which could indicate that not all
417 ssDNA genomes are spooled.
418 The spooling genome structure is efficient for packaging long genomes of dsRNA and dsDNA
419 viruses (Ilca et al., 2019; Liu and Cheng, 2015). The acquisition of the T=3 capsid gene from a ssRNA
420 virus could facilitate the packaging of a larger CRESS-DNA genome that cannot be accommodated by
421 capsids of small T=1 CRESS-DNA viruses (Kazlauskas et al., 2017). Many other ssDNA and dsDNA
422 viruses such as animal tumor polyomaviruses and papillomaviruses, and plant geminiviruses are
423 thought to be originated from primordial CRESS-DNA viruses (Kazlauskas et al., 2017; Koonin et al.,
424 2015; Koonin and Dolja, 2014). Considering the polyphyletic origin of ssDNA viruses and their
425 acquisitions of different RNA capsid proteins during the course of evolution, it will be interesting
426 when future ssDNA virus structures are unravelled.

Materials and methods

Virus production and purification

CtenDNAV-II was produced as previously described (Kimura and Tomaru, 2015). The crude virus suspension was loaded onto 15 to 50% (w/v) sucrose density gradients and centrifuged at 24,000 × *rpm* (102,170 × *g*) for 18 h at 4°C (Sw40Ti rotor, Beckman Coulter). The fractions of the sucrose gradient were applied to SDS-PAGE. The VP2 capsid protein fractions were pooled and subjected to centrifugation at 28,000 rpm (139,065 × *g*) for 3 h at 4°C (Sw40Ti rotor, Beckman Coulter). The pellet was resuspended in 50 mM Tris (pH 7.4), 100 mM NaCl, and 0.1 mM EDTA.

Cryo-EM and data collection

An aliquot (3 µl) of purified CtenDNAV-II virons (10 mg ml^{-1}) was deposited onto freshly glow-discharged holey carbon-coated copper grids (Quantifoil R 2/2, 300 mesh, copper) followed by 2 s of blotting in 100% relative humidity for plunge-freezing (Vitrobot Mark IV) in liquid ethane. Images were acquired using a Titan Krios microscope (Thermo Fisher Scientific) operated at 300 kV and equipped with a K2 Summit direct electron detector (Gatan) and an energy filter.

Image processing and 3D reconstruction

The micrographs were corrected for beam-induced drift using MotionCor2 1.2.6 (Zheng et al., 2017), and contrast transfer function (CTF) parameters were estimated using Gctf 1.06 (Zhang, 2016). The RELION 3.1 package (Zivanov et al., 2018) was used for particle picking, 2D and 3D classifications, *de novo* 3D model generation and refinement. A reconstruction of the capsid was generated in I4 symmetry using 33,507 particles, which were obtained by performing 9 consecutive 2D classification steps. The two genome reconstructions were generated in C1 symmetry using 21,559 particles, which were generated by performing 6 consecutive 3D classifications of the 33,507 particles that were obtained from the 2D classification step. Resolutions were estimated using the gold standard Fourier shell correlation (threshold, 0.143 criterion) (Henderson et al., 2012; Scheres and Chen, 2012). The data set and image processing are summarized in Table 1.

To reconstruct the genome without icosahedral symmetry a similar procedure to what has been described by Ilca et al (Ilca et al., 2019) was carried out. The contribution of the capsid was first subtracted from the map created during the final iteration of the I4 refinement job using the Particle subtraction function in Relion. To create a mask for the particle subtraction, the capsid model was first transformed to a density map using the molmap command in Chimera and then an inverted soft edged mask was created from the density map using relion_mask_create with the --invert option. The subtraction was followed by 3D classification (C1 symmetry), which generated the subset of 21,559 particles that was used for the final C1 refinement. The 3D classifications revealed clear density of an outer layer, and an additional subtraction was therefore performed using a spherical mask of 100 Å before the final refinement. The spherical mask was created using relion_mask_create with the --denovo and --outer_radius options. The two maps (the capsid density created by molmap and the circular map) were combined using Chimeras vop command before creating a new inverted mask using relion_mask_create, which was used for subtraction before the final 3D refinement. To reconstruct the core alone the spherical mask was used for the subtraction procedure.

Model building and refinement

The atomic model of CtenDNAV-II capsid protein was manually built into the density map using Coot (Emsley and Cowtan, 2004). The model was further improved through cycles of real-space refinement in PHENIX (Adams et al., 2010) with geometric and secondary structure restraints, and subsequent manual corrections by Coot were carried out iteratively. Refinement statistics are summarized in Table 1. Figures were generated using the programs UCSF Chimera (Pettersen et al., 2004) and UCSF

Chimera X (Goddard et al., 2018). Structural comparison of the CtenDNAV-II capsid protein was carried out by the DALI server as a heuristic search against all structures (as of 2020-06-03) in the PDB (Holm, 2020a).

Data availability

The atomic coordinates of CtenDNAV-II have been submitted to the Protein Data Bank under accession no. 7NS0. The cryo-EM density maps of the CtenDNAV-II capsid and genome has been deposited at the Electron Microscopy Data Bank under no. EMD-12554, 12555 and 12556, respectively.

Acknowledgements

The data were collected at the Cryo-EM Swedish National Facility funded by the Knut and Alice Wallenberg, Erling Persson Family, and Kempe Foundations, SciLifeLab, Stockholm University and Umeå University. We thank Julian Conrad and Dustin Morado for help with data collection. We also want to thank Afonso Vieira for valuable discussions.

Funding was provided by the following agencies: Vetenskapsrådet (VR)/The Swedish Research Council (to K.O., grant no. 2018-03387), the Swedish Foundation for International Cooperation in Research and Higher Education (STINT) (to Janos Hajdu and K.O., grant no. JA2014-5721), FORMAS research grant from the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning (to K.O., grant no. 2018-00421), the Royal Swedish Academy of Sciences (to K.O., grant no. BS2018-0053), and the Japan Society for the Promotion of Science KAKENHI (to Keizo Nagasaki, K.K., and Y.T., grant no. 16H06429, 16K21723, 16H06437, and 19H00956).

A.M., K.K., Y.T., and K.O. prepared the cryo-EM samples. A.M. and K.O. designed the experiments. A.M. and K.O. collected the cryo-EM data and analysed the data. H.W performed the HI assay. Y.T. performed the adsorption test. Y.H., K.Y., S.M. and K.K. performed the sequence analyses. A.M. and K.O. wrote the manuscript. All of the authors discussed the results and proofread the manuscript.

We declare no competing financial interests.

References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**:213–221. doi:10.1107/S0907444909052925

Bratbak G, Egge JK, Heldal M. 1993. Viral mortality of the marine alga Emiliania huxleyi (Haptophyceae) and termination of algal blooms. *Mar Ecol Prog Ser* **93**:39–48.

Brussaard CPD. 2004. Viral Control of Phytoplankton Populations-a Review1. *J Eukaryot Microbiol* **51**:125–138. doi:10.1111/j.1550-7408.2004.tb00537.x

Chiba Y, Tomaru Y, Shimabukuro H, Kimura K, Hirai M, Takaki Y, Hagiwara D, Nunoura T, Urayama S. 2020. Viral RNA Genomes Identified from Marine Macroalgae and a Diatom. *Microbes Environ* **35**:n/a. doi:10.1264/jsme2.ME20016

Coulibaly F, Chevalier C, Gutsche I, Pous J, Navaza J, Bressanelli S, Delmas B, Rey FA. 2005. The Birnavirus Crystal Structure Reveals Structural Relationships among Icosahedral Viruses. *Cell* **120**:761–772. doi:10.1016/j.cell.2005.01.009

Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, Sun R. 2017. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **541**:112–116. doi:10.1038/nature20589

Diemer GS, Stedman KM. 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol Direct* **7**:13. doi:10.1186/1745-6150-7-13

Ding K, Celma CC, Zhang X, Chang T, Shen W, Atanasov I, Roy P, Zhou ZH. 2019. In situ structures of rotavirus polymerase in action and mechanism of mRNA transcription and release. *Nat Commun* **10**:2216. doi:10.1038/s41467-019-10236-7

Dolja VV, Koonin EV. 2018. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res* **244**:36–52. doi:10.1016/j.virusres.2017.10.020

Emsley P, Cowtan K. 2004. *Coot* : model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**:2126–2132. doi:10.1107/S0907444904019158

Fuhrman JA. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**:541–548. doi:10.1038/21119

Fulton JM, Fredricks HF, Bidle KD, Vardi A, Kendrick BJ, DiTullio GR, Van Mooy BAS. 2014. Novel molecular determinants of viral susceptibility and resistance in the lipidome of *E miliania huxleyi*: Novel lipids in *Emiliania huxleyi*. *Environ Microbiol* **16**:1137–1149. doi:10.1111/1462-2920.12358

Garriga D, Querol-Audí J, Abaitua F, Saugar I, Pous J, Verdaguer N, Castón JR, Rodriguez JF. 2006. The 2.6-Angstrom Structure of Infectious Bursal Disease Virus-Derived T=1 Particles Reveals New Stabilizing Elements of the Virus Capsid. *J Virol* **80**:6895–6905. doi:10.1128/JVI.00368-06

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis: UCSF ChimeraX Visualization System. *Protein Sci* **27**:14–25. doi:10.1002/pro.3235

Gorzelnik KV, Cui Z, Reed CA, Jakana J, Young R, Zhang J. 2016. Asymmetric cryo-EM structure of the canonical *Allolevivirus* Qβ reveals a single maturation protein and the genomic ssRNA in situ. *Proc Natl Acad Sci* **113**:11519–11524. doi:10.1073/pnas.1609482113

Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. 1978. Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**:368–373. doi:10.1038/276368a0

Helgstrand C, Munshi S, Johnson JE, Liljas L. 2004. The refined structure of Nudaurelia capensis ω Virus reveals control elements for a T = 4 capsid maturation. *Virology* **318**:192–203. doi:10.1016/j.virol.2003.08.045

Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, Egelman EH, Feng Z, Frank J, Grigorieff N, Jiang W, Ludtke SJ, Medalia O, Penczek PA, Rosenthal PB, Rossmann MG, Schmid MF, Schröder GF, Steven AC, Stokes DL, Westbrook JD, Wriggers W, Yang H, Young J, Berman HM, Chiu W, Kleywegt GJ, Lawson CL. 2012. Outcome of the First Electron Microscopy Validation Task Force Meeting. *Structure* **20**:205–214. doi:10.1016/j.str.2011.12.014

554 Hesketh EL, Saunders K, Fisher C, Potze J, Stanley J, Lomonossoff GP, Ranson NA. 2018. The 3.3 Å
555 structure of a plant geminivirus using cryo-EM. *Nat Commun* **9**:2369. doi:10.1038/s41467-018-04793-
556 6
557 Ho KL, Gabrielsen M, Beh PL, Kueh CL, Thong QX, Streetley J, Tan WS, Bhella D. 2018. Structure
558 of the Macrobrachium rosenbergii nodavirus: A new genus within the Nodaviridae? *PLOS Biol*
559 **16**:e3000038. doi:10.1371/journal.pbio.3000038
560 Holm L. 2020a. DALI and the persistence of protein shape. *Protein Sci* **29**:128–140.
561 doi:10.1002/pro.3749
562 Holm L. 2020b. Using Dali for Protein Structure Comparison In: Gáspári Z, editor. Structural
563 Bioinformatics, Methods in Molecular Biology. New York, NY: Springer US. pp. 29–42.
564 doi:10.1007/978-1-0716-0270-6_3
565 Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, Le Mercier P. 2011.
566 ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res* **39**:D576–D582.
567 doi:10.1093/nar/gkq901
568 Ilca SL, Sun X, El Omari K, Kotecha A, de Haas F, DiMaio F, Grimes JM, Stuart DI, Poranen MM,
569 Huiskonen JT. 2019. Multiple liquid crystalline geometries of highly compacted nucleic acid in a
570 dsRNA virus. *Nature* **570**:252–256. doi:10.1038/s41586-019-1229-9
571 Inoue T, Tsai B. 2011. A Large and Intact Viral Particle Penetrates the Endoplasmic Reticulum
572 Membrane to Reach the Cytosol. *PLoS Pathog* **7**:e1002037. doi:10.1371/journal.ppat.1002037
573 Johannessen T, Larsen A, Bratbak G, Pagarete A, Edvardsen B, Egge E, Sandaa R-A. 2017. Seasonal
574 Dynamics of Haptophytes and dsDNA Algal Viruses Suggest Complex Virus-Host Relationship.
575 *Viruses* **9**:84. doi:10.3390/v9040084
576 Johnson JE, Tang L, Johnson KN, Ball LA, Lin T, Yeager M. 2001. The structure of Pariacoto virus
577 reveals a dodecahedral cage of duplex RNA. *Nat Struct Biol* **8**:77–83. doi:10.1038/83089
578 Kazlauskas D, Dayaram A, Kraberger S, Goldstien S, Varsani A, Krupovic M. 2017. Evolutionary
579 history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA
580 nodaviruses. *Virology* **504**:114–121. doi:10.1016/j.virol.2017.02.001
581 Kazlauskas D, Varsani A, Koonin EV, Krupovic M. 2019. Multiple origins of prokaryotic and
582 eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat Commun* **10**:3425.
583 doi:10.1038/s41467-019-11433-0
584 Kimura K, Tomaru Y. 2015. Discovery of Two Novel Viruses Expands the Diversity of Single-
585 Stranded DNA and Single-Stranded RNA Viruses Infecting a Cosmopolitan Marine Diatom. *Appl*
586 *Environ Microbiol* **81**:1120–1131. doi:10.1128/AEM.02380-14
587 Koning RI, Gomez-Blanco J, Akopjana I, Vargas J, Kazaks A, Tars K, Carazo JM, Koster AJ. 2016.
588 Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nat Commun*
589 **7**:12524. doi:10.1038/ncomms12524
590 Koonin EV, Dolja VV. 2014. Virus World as an Evolutionary Network of Viruses and Capsidless
591 Selfish Elements. *Microbiol Mol Biol Rev* **78**:278–303. doi:10.1128/MMBR.00049-13
592 Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The
593 ultimate modularity. *Virology* **479–480**:2–25. doi:10.1016/j.virol.2015.02.039
594 Krupovic M. 2013. Networks of evolutionary interactions underlying the polyphyletic origin of
595 ssDNA viruses. *Curr Opin Virol* **3**:578–586. doi:10.1016/j.coviro.2013.06.010
596 Krupovic M, Koonin EV. 2017. Multiple origins of viral capsid proteins from cellular ancestors. *Proc*
597 *Natl Acad Sci U S A* **114**:E2401–E2410. doi:10.1073/pnas.1621061114
598 Krupovic M, Varsani A, Kazlauskas D, Breitbart M, Delwart E, Rosario K, Yutin N, Wolf YI,
599 Harrach B, Zerbini FM, Dolja VV, Kuhn JH, Koonin EV. 2020. *Cressdnaviricota* : a Virus Phylum
600 Unifying Seven Families of Rep-Encoding Viruses with Single-Stranded, Circular DNA Genomes. *J*
601 *Virol* **94**:e00582-20, /jvi/94/12/JVI.00582-20.atom. doi:10.1128/JVI.00582-20
602 Kucukelbir A, Sigworth FJ, Tagare HD. 2014. Quantifying the local resolution of cryo-EM density
603 maps. *Nat Methods* **11**:63–65. doi:10.1038/nmeth.2727
604 Laanto E, Mäntynen S, De Colibus L, Marjakangas J, Gillum A, Stuart DI, Ravantti JJ, Huiskonen JT,
605 Sundberg L-R. 2017. Virus found in a boreal lake links ssDNA and dsDNA viruses. *Proc Natl Acad*
606 *Sci* **114**:8378–8383. doi:10.1073/pnas.1703834114
607 Liu H, Cheng L. 2015. Cryo-EM shows the polymerase structures and a nonspooled genome within a
608 dsRNA virus. *Science* **349**:1347–1350. doi:10.1126/science.aaa4938

Liu W, Hsu C-H, Hong Y-R, Wu S-C, Wang C-H, Wu Y-M, Chao C-B, Lin C-S. 2005. Early endocytosis pathways in SSN-1 cells infected by dragon grouper nervous necrosis virus. *J Gen Virol* **86**:2553–2561. doi:10.1099/vir.0.81021-0

Liu Y-T, Jih J, Dai X, Bi G-Q, Zhou ZH. 2019. Cryo-EM structures of herpes simplex virus type 1 portal vertex and packaged genome. *Nature* **570**:257–261. doi:10.1038/s41586-019-1248-6

Mallery DL, Márquez CL, McEwan WA, Dickson CF, Jacques DA, Anandapadamanaban M, Bichel K, Towers GJ, Saiardi A, Böcking T, James LC. 2018. IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife* **7**. doi:10.7554/eLife.35335

Moreno P, Souto S, Leiva-Rebollo R, Borrego JJ, Bandín I, Alonso MC. 2019. Capsid amino acids at positions 247 and 270 are involved in the virulence of betanodaviruses to European sea bass. *Sci Rep* **9**:14068. doi:10.1038/s41598-019-50622-1

Muhire BM, Golden M, Murrell B, Lefeuvre P, Lett J-M, Gray A, Poon AYF, Ngandu NK, Semegni Y, Tanov EP, Monjane AL, Harkins GW, Varsani A, Shepherd DN, Martin DP. 2014. Evidence of Pervasive Biologically Functional Secondary Structures within the Genomes of Eukaryotic Single-Stranded DNA Viruses. *J Virol* **88**:1972–1989. doi:10.1128/JVI.03031-13

Munke A, Kimura K, Tomaru Y, Okamoto K. 2020. Capsid Structure of a Marine Algal Virus of the Order *Picornavirales*. *J Virol* **94**:e01855-19, /jvi/94/9/JVI.01855-19.atom. doi:10.1128/JVI.01855-19

Nagasaki K, Shirai Y, Takao Y, Mizumoto H, Nishida K, Tomaru Y. 2005. Comparison of genome sequences of single-stranded RNA viruses infecting the bivalve-killing dinoflagellate Heterocapsa circularisquama. *Appl Environ Microbiol* **71**:8888–8894. doi:10.1128/AEM.71.12.8888-8894.2005

Nagasaki K, Tarutani K, Yamaguchi M. 1999. Growth Characteristics of Heterosigma akashiwo Virus and Its Possible Use as a Microbiological Agent for Red Tide Control. *Appl Environ Microbiol* **65**:898.

Ng TFF, Zhang W, Sachsenröder J, Kondov NO, da Costa AC, Vega E, Holtz LR, Wu G, Wang D, Stine CO, Antonio M, Mulvaney US, Muench MO, Deng X, Ambert-Balay K, Pothier P, Vinjé J, Delwart E. 2015. A diverse group of small circular ssDNA viral genomes in human and non-human primate stools. *Virus Evol* **1**:vev017. doi:10.1093/ve/vev017

Nishihara T, Kurano N, Shinoda S. 1986. Calculation of Most Probable Number for Enumeration of Bacteria on a Micro-Computer. *Eisei Kagaku* **32**:226–228. doi:10.1248/jhs1956.32.226

Okamoto K, Ferreira RJ, Larsson DSD, Maia FRNC, Isawa H, Sawabe K, Murata K, Hajdu J, Iwasaki K, Kasson PM, Miyazaki N. 2020. Acquired Functional Capsid Structures in Metazoan Totivirus-like dsRNA Virus. *Structure* **28**:888-896.e3. doi:10.1016/j.str.2020.04.016

Okamoto K, Miyazaki N, Larsson DSD, Kobayashi D, Svenda M, Mühlig K, Maia FRNC, Gunn LH, Isawa H, Kobayashi M, Sawabe K, Murata K, Hajdu J. 2016. The infectious particle of insect-borne totivirus-like Omono River virus has raised ridges and lacks fibre complexes. *Sci Rep* **6**:33170. doi:10.1038/srep33170

Penkler DL, Jiwaji M, Domitrovic T, Short JR, Johnson JE, Dorrington RA. 2016. Binding and entry of a non-enveloped T =4 insect RNA virus is triggered by alkaline pH. *Virology* **498**:277–287. doi:10.1016/j.virol.2016.08.028

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera?A visualization system for exploratory research and analysis. *J Comput Chem* **25**:1605–1612. doi:10.1002/jcc.20084

Phan TG, Mori D, Deng X, Rajindrajith S, Ranawaka U, Fan Ng TF, Bucardo-Rivera F, Orlandi P, Ahmed K, Delwart E. 2015. Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. *Virology* **482**:98–104. doi:10.1016/j.virol.2015.03.011

Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht H-J, Johnson JE, Kamer G, Luo M, Mosser AG, Rueckert RR, Sherry B, Vriend G. 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**:145–153. doi:10.1038/317145a0

Scheres SHW. 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**:519–530. doi:10.1016/j.jsb.2012.09.006

Scheres SHW, Chen S. 2012. Prevention of overfitting in cryo-EM structure determination. *Nat Methods* **9**:853–854. doi:10.1038/nmeth.2115

Speir JA, Taylor DJ, Natarajan P, Pringle FM, Ball LA, Johnson JE. 2010. Evolution in Action: N and C Termini of Subunits in Related T = 4 Viruses Exchange Roles as Molecular Switches. *Structure* **18**:700–709. doi:10.1016/j.str.2010.03.010

Ströh LJ, Rustmeier NH, Blaum BS, Botsch J, Rößler P, Wedekink F, Lipkin WI, Mishra N, Stehle T. 2020. Structural Basis and Evolution of Glycan Receptor Specificities within the Polyomavirus Family. *mBio* **11**:e00745-20, /mbio/11/4/mBio.00745-20.atom. doi:10.1128/mBio.00745-20

Suttle CA. 2007. Marine viruses — major players in the global ecosystem. *Nat Rev Microbiol* **5**:801–812. doi:10.1038/nrmicro1750

Suttle CA. 1993. Enumeration and isolation of virusesHandbook of Methods in Aquatic Microbial Ecology. Boca Raton: Lewis Publishers.

Tarutani K, Nagasaki K, Yamaguchi M. 2006. Virus adsorption process determines virus susceptibility in Heterosigma akashiwo (Raphidophyceae). *Aquat Microb Ecol* **42**:209–213. doi:10.3354/ame042209

Tarutani K, Nagasaki K, Yamaguchi M. 2000. Viral Impacts on Total Abundance and Clonal Composition of the Harmful Bloom-Forming PhytoplanktonHeterosigma akashiwo. *Appl Environ Microbiol* **66**:4916–4920. doi:10.1128/AEM.66.11.4916-4920.2000

Tisza MJ, Pastrana DV, Welch NL, Stewart B, Peretti A, Starrett GJ, Pang Y-YS, Krishnamurthy SR, Pesavento PA, McDermott DH, Murphy PM, Whited JL, Miller B, Brenchley J, Rosshart SP, Rehermann B, Doorbar J, Ta'ala BA, Pletnikova O, Troncoso JC, Resnick SM, Bolduc B, Sullivan MB, Varsani A, Segall AM, Buck CB. 2020. Discovery of several thousand highly diverse circular DNA viruses. *eLife* **9**:e51971. doi:10.7554/eLife.51971

Tomaru Y, Kimura K. 2016. Rapid quantification of viable cells of the planktonic diatom *Chaetoceros tenuissimus* and associated RNA viruses in culture. *Plankton Benthos Res* **11**:9–16. doi:10.3800/pbr.11.9

Tomaru Y, Shirai Y, Toyoda K, Nagasaki K. 2011. Isolation and characterisation of a single-stranded DNA virus infecting the marine planktonic diatom Chaetoceros tenuissimus. *Aquat Microb Ecol* **64**:175–184. doi:10.3354/ame01517

Tomaru Y, Toyoda K, Suzuki H, Nagumo T, Kimura K, Takao Y. 2013. New single-stranded DNA virus with a unique genomic structure that infects marine diatom Chaetoceros setoensis. *Sci Rep* **3**:3337. doi:10.1038/srep03337

Wagstaff BA, Rejzek M, Field RA. 2018. Identification of a Kdn biosynthesis pathway in the haptophyte *Prymnesium parvum* suggests widespread sialic acid biosynthesis among microalgae. *J Biol Chem* **293**:16277–16290. doi:10.1074/jbc.RA118.004921

Wang F, Liu Y, Su Z, Osinski T, de Oliveira GAP, Conway JF, Schouten S, Krupovic M, Prangishvili D, Egelman EH. 2019. A packing for A-form DNA in an icosahedral virus. *Proc Natl Acad Sci* **116**:22591–22597. doi:10.1073/pnas.1908242116

Wang X, Zhang F, Su R, Li X, Chen W, Chen Q, Yang T, Wang J, Liu H, Fang Q, Cheng L. 2018. Structure of RNA polymerase complex and genome within a dsRNA virus provides insights into the mechanisms of transcription and assembly. *Proc Natl Acad Sci* **115**:7344–7349. doi:10.1073/pnas.1803885115

Wilhelm SW, Suttle CA. 1999. Viruses and Nutrient Cycles in the Sea. *BioScience* **49**:781–788. doi:10.2307/1313569

Wolf YI, Silas S, Wang Y, Wu S, Bocek M, Kazlauskas D, Krupovic M, Fire A, Dolja VV, Koonin EV. 2020. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol* **5**:1262–1270. doi:10.1038/s41564-020-0755-4

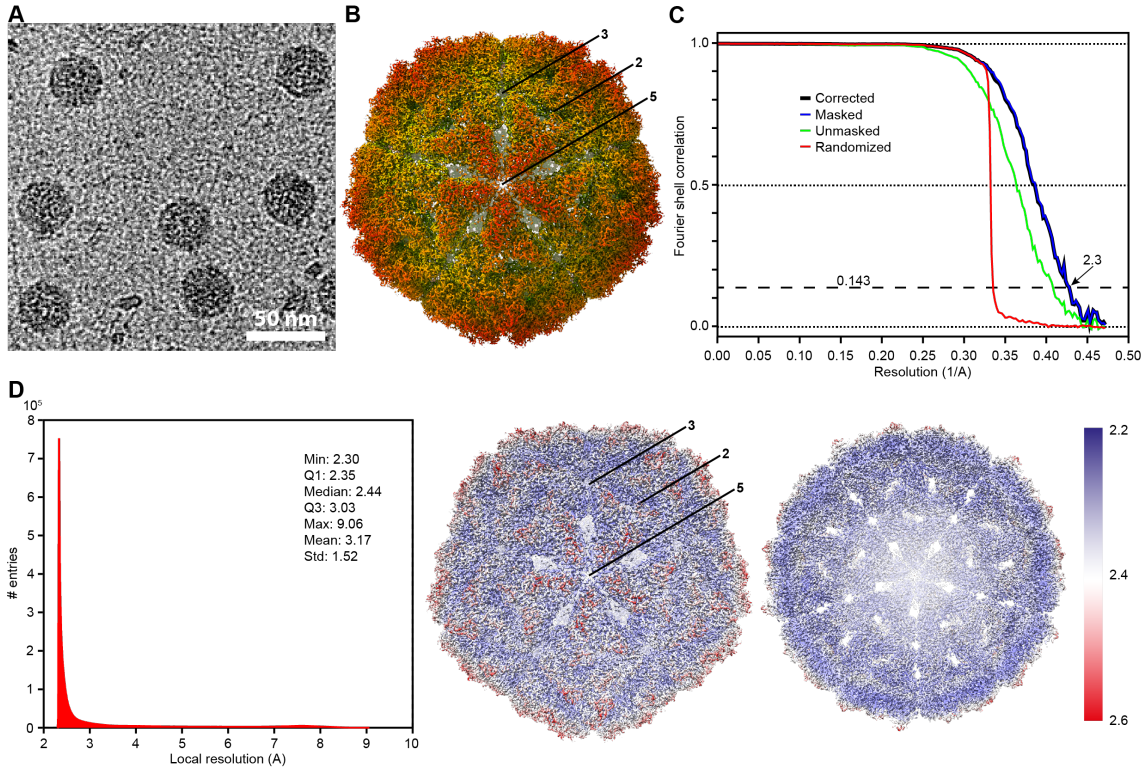
Zhang K. 2016. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**:1–12. doi:10.1016/j.jsb.2015.11.003

Zhang X, Ding K, Yu X, Chang W, Sun J, Hong Zhou Z. 2015. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature* **527**:531–534. doi:10.1038/nature15767

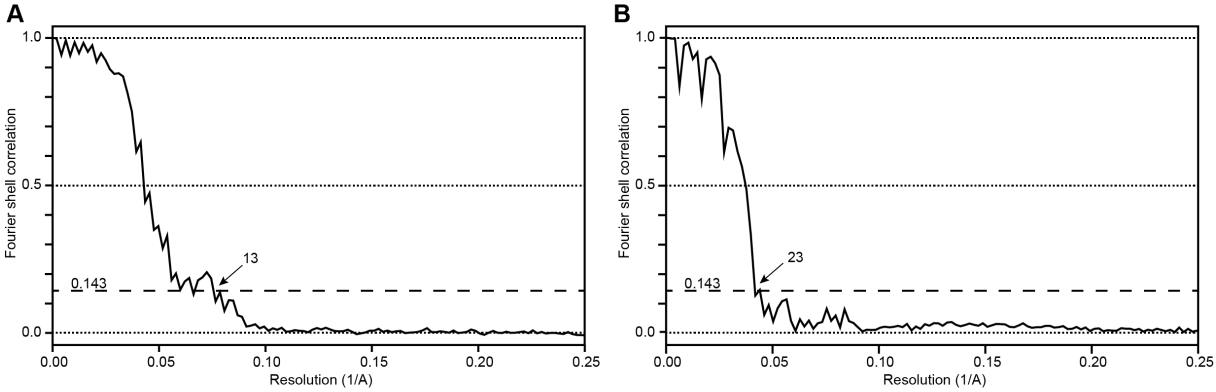
Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**:331–332. doi:10.1038/nmeth.4193

716 Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SH. 2018. New tools
717 for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**:e42166.
718 doi:10.7554/eLife.42166
719 Zocher G, Mistry N, Frank M, Hähnlein-Schick I, Ekström J-O, Arnberg N, Stehle T. 2014. A Sialic
720 Acid Binding Site in a Human Picornavirus. *PLoS Pathog* **10**:e1004401.
721 doi:10.1371/journal.ppat.1004401
722

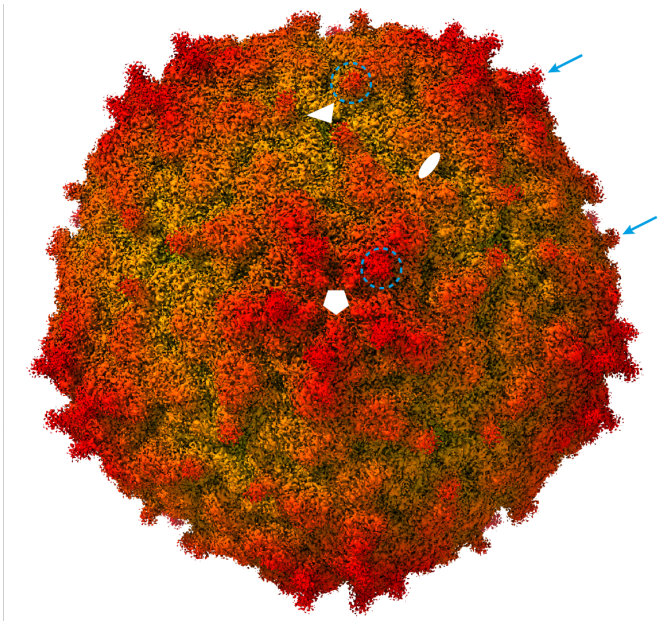
Supplementary figures



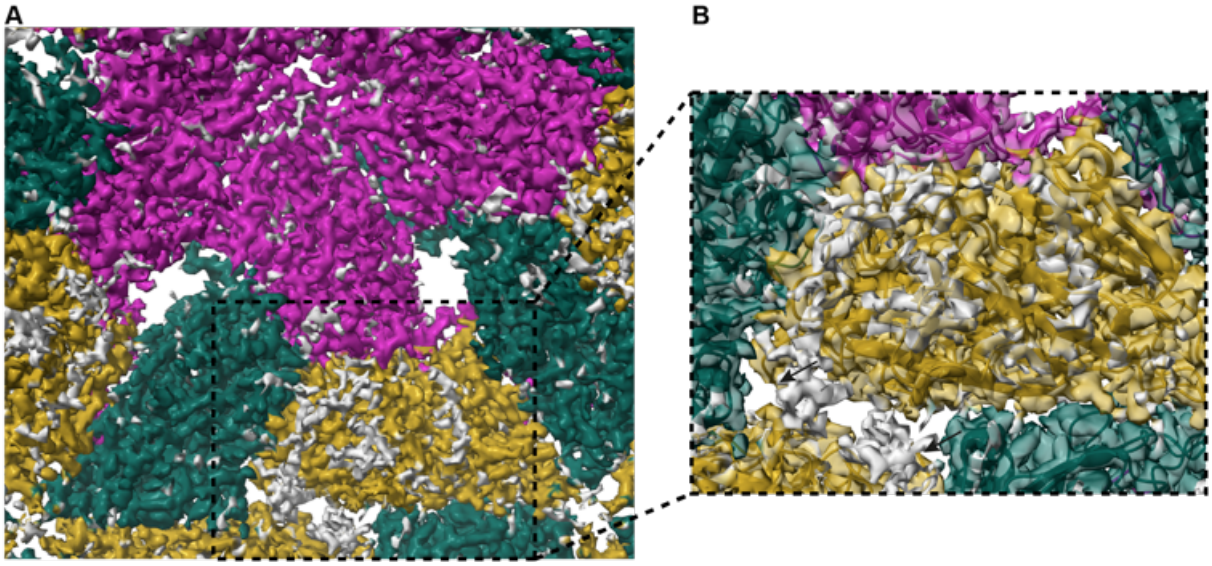
Supplementary Figure 1. Data collection and reconstruction of the capsid. (**A**) Cryo-EM raw image of CtenDNAV-II. (**B**) Cryo-EM 3D reconstruction of the CtenDNAV-II capsid. The capsid is viewed down the five-fold axis and radially coloured from blue to red. The icosahedral five-fold, three-fold, and two-fold axes are labelled as 5, 3, and 2, respectively. (**C**) The gold standard FSC resolution curves of masked (blue) and unmasked (green) reconstructions of the CtenDNAV-II capsid. Possible effects of the masking were compensated for by noise randomization (red), to create the final FSC curve (black). The resolution at which the correlation drops below the FSC = 0.143 (gold standard threshold (Henderson et al., 2012; Scheres and Chen, 2012) is 2.3 Å. (**D**) Local resolution of the final reconstruction determined by Relion. The left panel shows a histogram of the local resolution, and the middle and right panel shows the reconstruction coloured from red to white to blue that corresponds to local resolutions 2.6, 2.4 and 2.2, respectively. The icosahedral five-fold, three-fold, and two-fold axes are labelled as 5, 3, and 2, respectively in the middle panel. The front half of the reconstruction is removed in the right most panel to visualize the inside of the capsid.



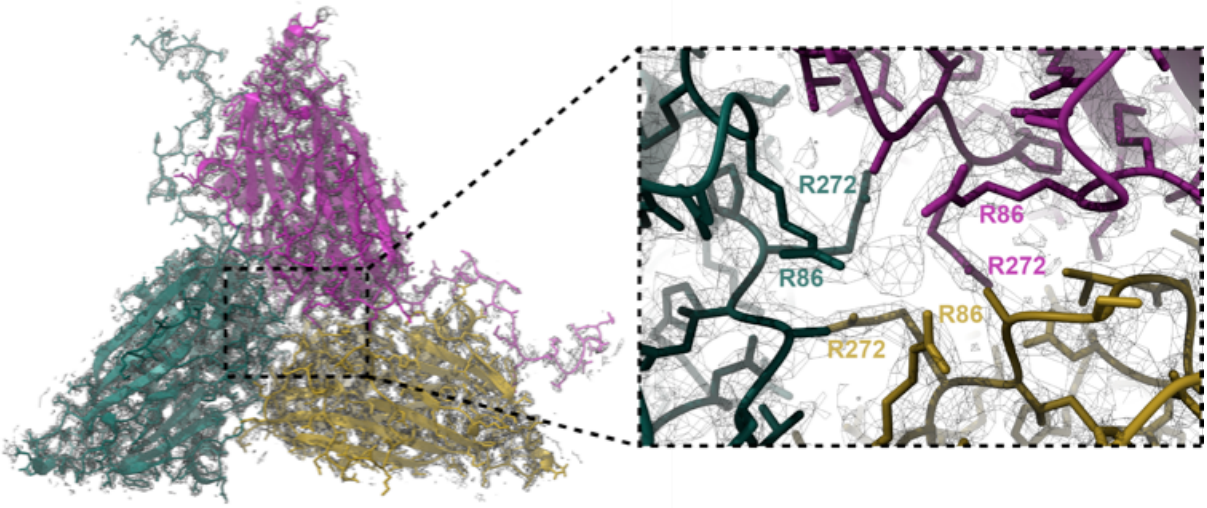
Supplementary Figure 2. The gold standard FSC resolution curves of the asymmetric reconstructions. (**A**) FSC curve of the outer layer. (**B**) FSC curve of the core.



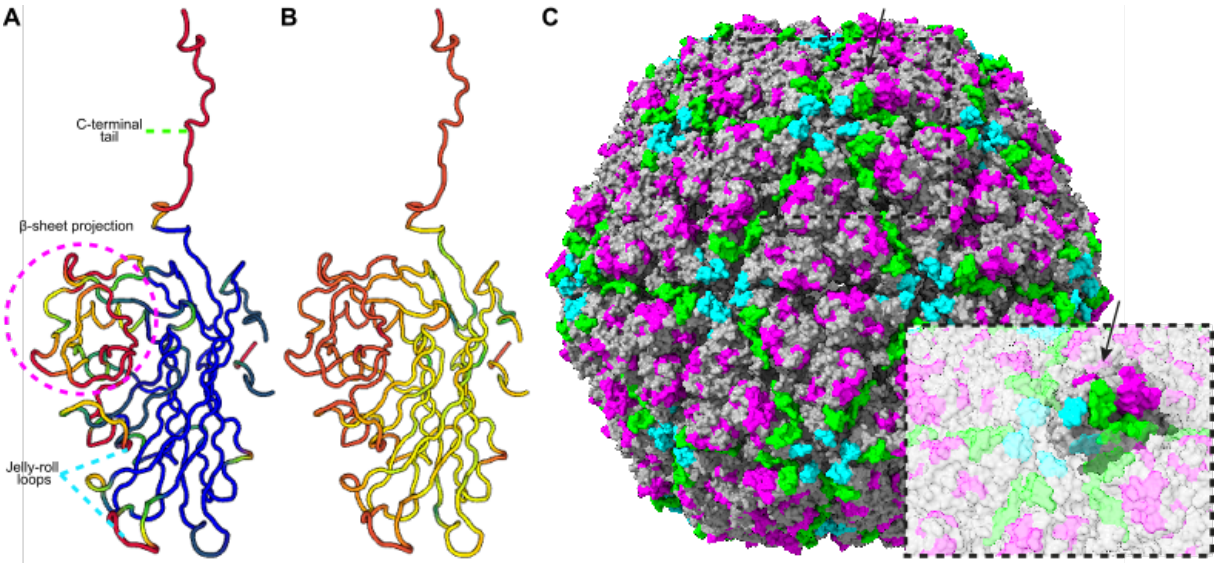
Supplementary Figure 2. Cryo-EM 3D reconstruction of the CtenDNAV-II capsid at low contour level. Unmodelled densities around the three- and five-fold axes are indicated with blue arrows and circles, which correspond to the C-terminal ends of subunit A and B. The capsid is viewed down the five-fold axis and radially coloured from blue to red. The positions of the five-, three-, and two-fold axes are shown by a white pentamer, triangle, and ellipse, respectively.



Supplementary Figure 3. Unmodelled density is visualized on the inside of the capsid below the C subunit. (**A**) Inside view of the cryo-EM reconstruction coloured based on the atomic model of the three subunits. By decreasing the contour level, unmodelled densities are visualized in grey. (**B**) Close-up view of the C subunit, where the reconstruction is transparent to reveal the model in ribbon representation. The last modelled residues in the C-termini of the two 2-fold symmetry related C subunits are indicated with black arrows.



Supplementary Figure 4. Putative ion located in the subunit interface. The model of a single icosahedral protomer viewed from the inside of the capsid and corresponding cryo-EM map is visualized to the left. To the right, is a close-up view of the subunit interface, displaying density in the centre possibly originating from an ion and the surrounding arginine residues.



Supplementary Figure 5. Superimposition performed by the DALI server of the CtenDNAV-II capsid protein against corresponding capsid proteins from viruses in Fig. 3B. The figures show CtenDNAV-II subunit A coloured according to (**A**) structural conservation and (**B**) sequence conservation where conservation mapping goes from blue for the highest values through green to red for the lowest values. (**A**) The three areas where structural conservation has the lowest values; the C-terminal tail, β-sheet projection, and jelly-roll loops C'-C'' and H-I are highlighted in green, magenta and cyan dashed lines. (**C**) The entire capsid rendered with a surface representation viewed down an icosahedral two-fold axis. The amino acids with lowest structural conservation (red in panel A) are coloured green, magenta and cyan that corresponds to the areas N-terminal tail, β-sheet projection and jelly-roll loops labelled in (A). The insert is zooming in on one of the five-folds and highlights one A-subunit and the C-terminus (green) coming from the B-subunit in an adjacent five-fold related protomer.

775
776

Strain	PDGRMLESLP	RRCQLVTEIR	NNVTVGSNPT	YILVAPSLGL	AFQAYQDTNV	PGGLDSSVYG
SS10V-8V	**********	**********	**********	**********	**********	**********
SS10V-35V	**********	**********	**********	**********	**********	**********
2-6_11V10	**********	**********	**********	**********	**********	**********
2-6_11V12	**********	**********	**********	**********	**********	**********
2-6_10V07	**********	**********	**********	**********	**********	**********
2-6_10V05	**********	**********	**********	**********	**********	**********
2-6_10V17	**********	**********	**********	**********	**********	**********
2-6_10V20	**********	**********	**********	**********	**********	**********
SS10V-24V	**********	**********	**********	**********	**********	**********
2-6_10V27	**********	**********	**********	**********	**********	**********
2-6_10V33	**********	**********	**********	**********	**********	**********
2-6_11V04	**********	**********	**********	**********	**********	**********
2-6_11V08	**********	**********	**********	**********	**********	**********
2-6_11V06	**********	**********	**********	**********	**********	**********
2-6_11V17	**********	**********	**********	**********	**********	**********

777

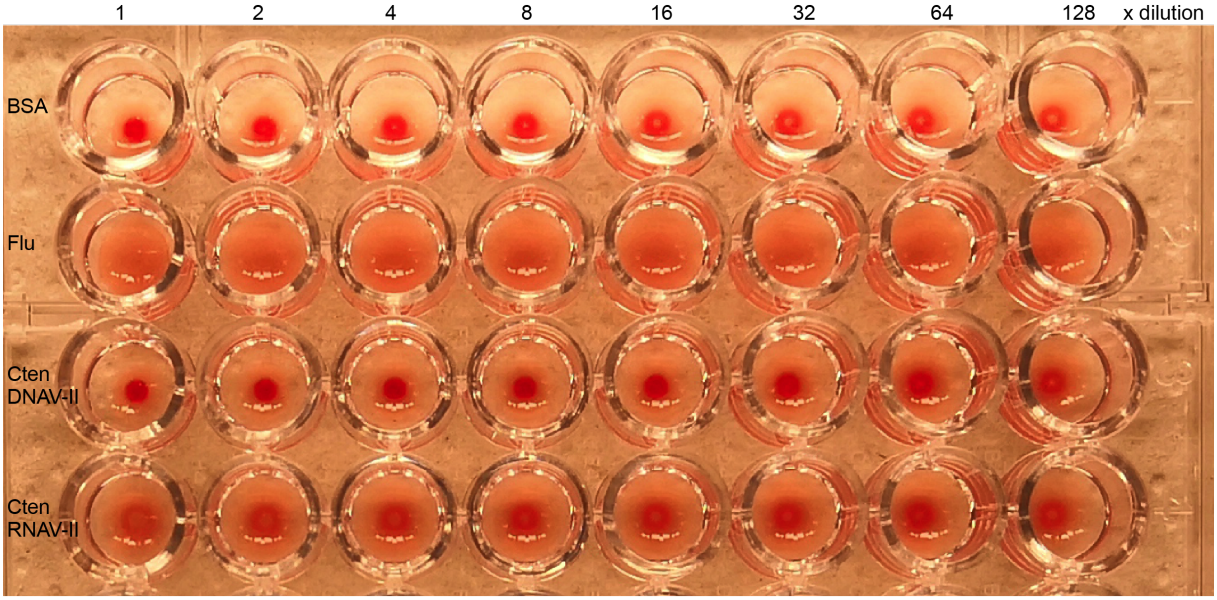
Strain	LQNRGCTVRA	NLSATSIENY	NDIAKWRIVS	QGINLKLNNV	EDENDGWYEA	CRFQHDWTPD	ELCLRSTEND
SS10V-8V	**********	**********	**********	**********	**********	**********	**********
SS10V-35V	**********	**********	**********	**********	**********	**********	**********
2-6_11V10	**********	**********	**********	**********	**********	**********	**********
2-6_11V12	**********	**********	**********	**********	**********	**********	**********
2-6_10V07	**********	**********	**********	**********	**********	**********	**********
2-6_10V05	**********	**********	**********	**********	**********	**********	**********
2-6_10V17	**********	**********	**********	**********	**********	**********	**********
2-6_10V20	**********	**********	**********	**********	**********	**********	**********
SS10V-24V	**********	**********	**********	**********	**********	**********	**********
2-6_10V27	**********	**********	**********	**********	**********	**********	**********
2-6_10V33	**********	**********	**********	**********	**********	**********	**********
2-6_11V04	**********	**********	**********	**********	**********	**********	**********
2-6_11V08	**********	**********	**********	**********	**********	**********	**S*******
2-6_11V06	**********	**********	**********	**********	**********	**********	**********
2-6_11V17	**********	**********	**********	**********	**********	**********	**********

Strain	ASTISQDEDL	VMGVISSSFL	NGALNTIGNN	MVEQRGYESG	LLKNIHKRMF	QLHNNTSAIR	PKTLQGQFNY
SS10V-8V	.....	.....M	.....	.....	.....	.....	.....
SS10V-35V	.....	.....S	.....	.....	.....	.....	.....
2-6_11V10	.....	.....	.....	.....	.....	.....	.....
2-6_11V12	.....	.....	.....	.....	.....	.....	.....
2-6_10V07	.....	.....	.....	.....	.....	.....	.....
2-6_10V05	.....	.....	.....	.....	.....	.....	.....
2-6_10V17	TNN.....	.....	.....	.....	.....	.....	.....
2-6_10V20	.....	.....	.....	.....	.....	.....	.....
SS10V-24V	T.....	.....	.....	.....	.....	.....	.....
2-6_10V27	.....	.....	.....	.....	.....	.....	.....
2-6_10V33	.....	.....	.....	.....	.....	.....	.....
2-6_11V04	.....	.....	.....	.....	.....	.....	.....
2-6_11V08	.NS.....	.....	.....	.....	.....	.....	.....
2-6_11V06	.NS.....	.....	.....	.....	.....	.....	.....
2-6_11V17	.....	.....	.....	.....	.....	.....	.....

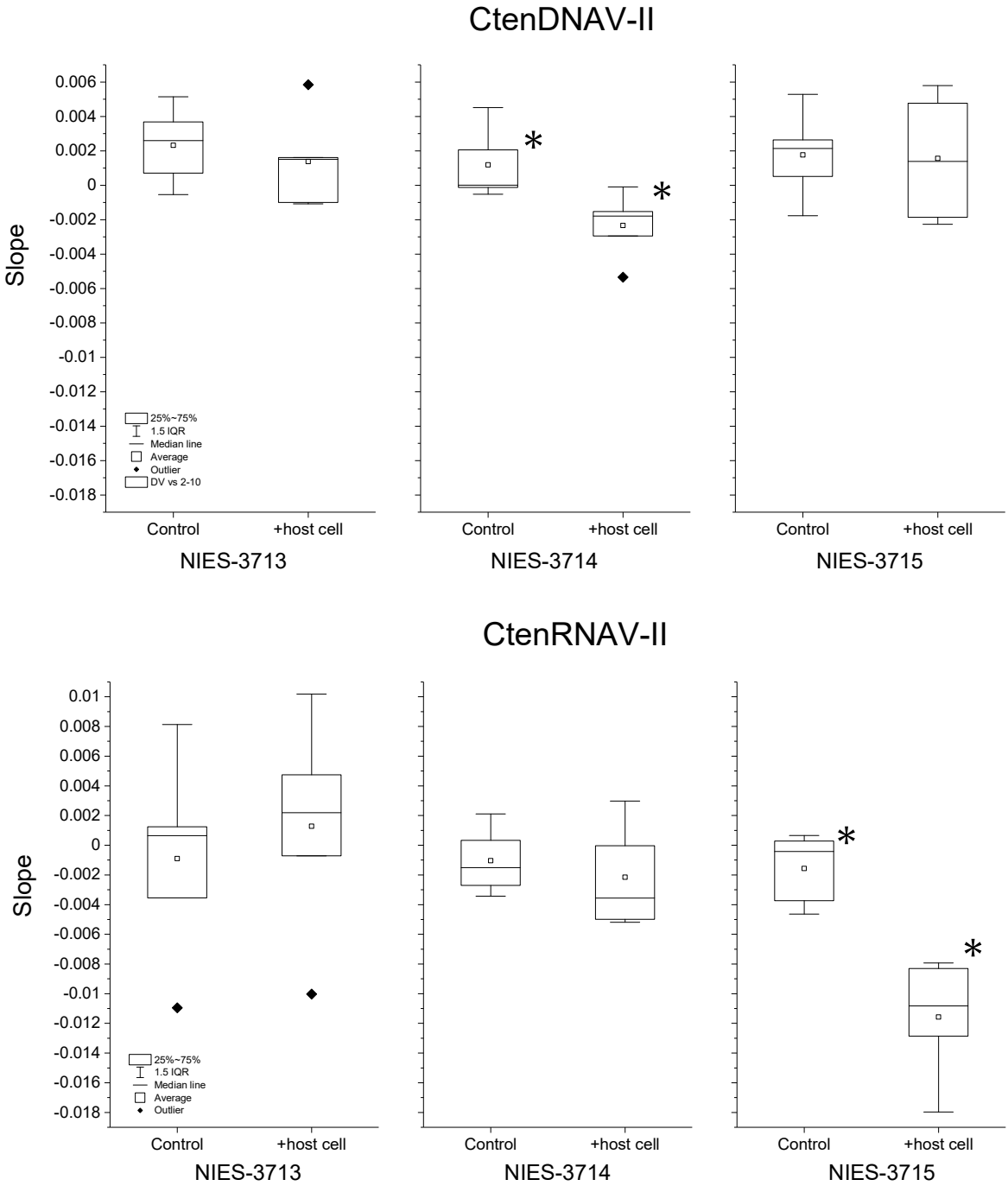
Strain	GSEITFSGTE	SEARFTDVPS	NRQLVDSLWH	NDYDCILIKL	YPRENTGAAG	QTGSALIVNA	IQNLELQYSP
SS10V-8V	.....	.....	.....	.....	.....	.....	.....
SS10V-35V	.....	.....	.....	.....	.....	.....	.....
2-6_11V10	.....	.....	.....	.....	.....	R.AKRPHCKR	YSKFGAAIFS
2-6_11V12	.....	.....	.....	.....	.....	R.....	.....
2-6_10V07	.....	N.....	.....	.....	.....	.....	.....
2-6_10V05	.....	N.....	.....	.....	.....	.....	.....
2-6_10V17	.....	.....	.....	.....	.....	R.....	.....
2-6_10V20	.....	N.....	.....	.....	.....	.....	.....
SS10V-24V	.....	.....	.....	.....	.....	.....	.....
2-6_10V27	.....	.....	.....	.....	.....	.....	.....
2-6_10V33	.....	N.....	.....	.....	.....	.....	.....
2-6_11V04	....I.....	.....	.....	.....	.....	..AKRPHCKR	YSKFGAAIFS
2-6_11V08	....I.....	.....	.....	.....	.....	.....	.....
2-6_11V06	.....	.....	.....	.....	.....	.....	.....
2-6_11V17	.....	.....	.....	.....	.....	R.....	.....

```
Strain      | TSDLSTYHIA NKRARMVEAK LDKKNNTDAA GEPFVPGSSR*
SS10V-8V    | |||||||||| |||||||||| |||||||||| ||||||||||*
SS10V-35V   | |||||||||| |||||||||| |||||||||| ||||||||||*
2-6_11V10   | N*
2-6_11V12   | |||||||||| |||||||||| |||||||||P ||||||||||T*
2-6_10V07   | |||||||||| |||||||||| |||||||||| ||||||||||*
2-6_10V05   | |||||||||| |||||||||| |||||||||| ||||||||||*
2-6_10V17   | |||||||||| ||||K|G||| |||||||||P ||||||||||T*
2-6_10V20   | |||||||||| |||||||||| |||||||||| ||||||||||*
SS10V-24V   | |||||||||| |||||||||| |||||||||| ||||||||||*
2-6_10V27   | |||||||||| ||||||G||| |||||||||| ||||||||||T*
2-6_10V33   | |||||||||| |||||||||| |||||||||| ||||||||||*
2-6_11V04   | N*
2-6_11V08   | |||||||||| |||||||||| |||||||||| ||||||||||*
2-6_11V06   | |||||||||| |||||||||| |||||||||| ||||||||||*
2-6_11V17   | |||||||||| |||||||||| |||||||||P ||||||||||T*
```

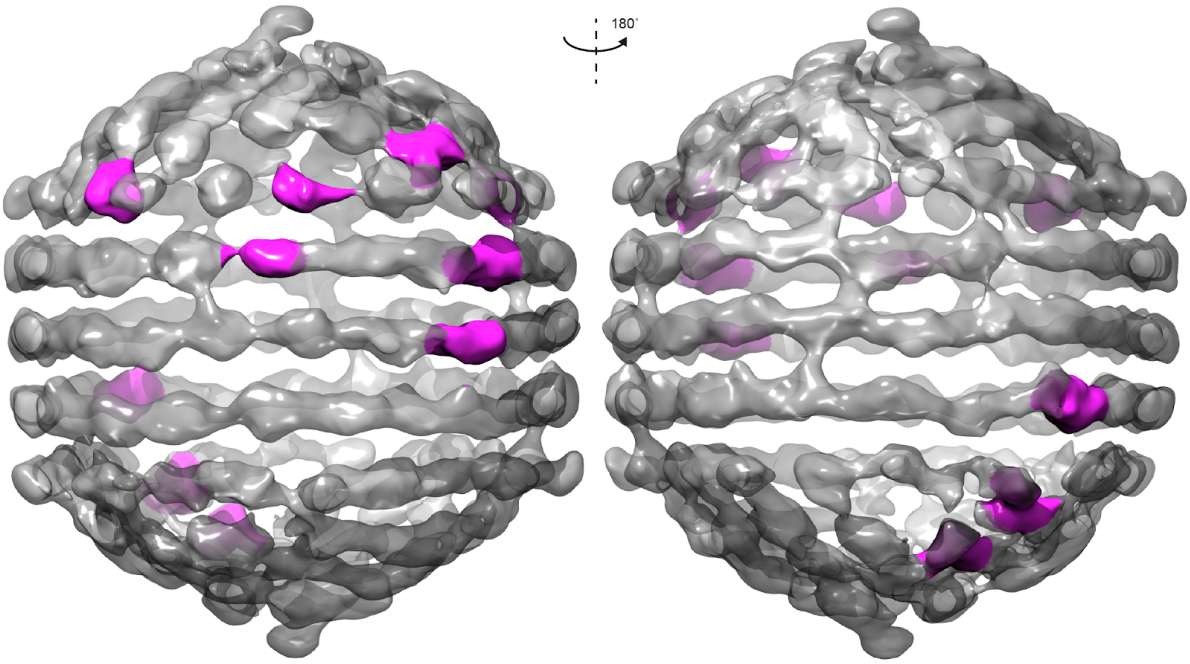
Supplementary Figure 6. Amino acid alignment of capsid proteins from 15 CtenDNAV-II strains. The consensus sequence is displayed in grey and the mutations in colour. The amino acids are organised in the same scheme as in Figure 2 with divisions into 70 residues per line, which are further subdivided into blocks of 10 residues by spaces within the sequence. The N-termini were not successfully sequenced and the first residue is therefore P81. This paper describes the structure of strain SS10V-8V.



Supplementary Figure 7. Hemagglutination inhibition assay. Twofold serially diluted 1 mg/mL BSA, influenza virus (flu), CtenDNAV-II and CtenRNV-II are tested for HI activity. The BSA and flu were used for negative and positive controls for the assay.



Supplementary Fig. 8. Adsorption tests between viruses and diatoms. The results are shown in slope of free virus titer in a culture for 2h (see materials and methods). Significant differences were detected in CtenDNAV-II vs. NIES-3714, and CtenRNAV-II vs. NIES-3715 (t-test, $p<0.05$, $n=5$). Their adsorption coefficients were $1.4\pm0.9\times10^{-9}$ mL min^{-1} and $2.7\pm1.1\times10^{-9}$ mL min^{-1}, respectively. NIES-3713: *Chaetoceros socialis* f. *radians*; NIES-3714: *Chaetoceros tenuissimus*; NIES-3715: *Chaetoceros tenuissimus.*



Supplementary Figure 9. Connections between the outer genome layer and the core. The outer genome layer where the sites that connect to the core are highlighted in magenta. The connections between the outer genome layer and the core seem to be confined to two specific areas on approximately opposite sides. One area is shown in the left panel and the second is shown to the right, which is rotated 180°.

807 # Supplementary methods

808 ## Sequencing of viral capsid protein gene

809 Two hundred µL of *C. tenuissimus* NIES-3715 strain culture was added to each well of a 96-well
810 culture plate, and 10 uL of virus strains (SS10V-8V, SS10V-24V, SS10V-35V, 2-6_10V-05, 2-
811 6_10V-07, 2-6_10V-17, 2-6_10V-20, 2-6_10V-27, 2-6_10V-33, 2-6_11V-04, 2-6_11V-06, 2-6_11V-
812 08, 2-6_11V-10, 2-6_11V-12, 2-6_11V-17) isolated from *C. tenuissimus* were added to each separate
813 well. Diatom cultures with viral solution were grown at 15 °C in modified SWM3 medium (Kimura
814 and Tomaru, 2015) under a 12/12-h light-dark cycle of ca. 110-150 mmol of photons / m2 / s1 by
815 using white fluorescent illumination. After 24 hours of viral inoculation, the cultures were transferred
816 to microtubes and the cells were pelleted by centrifugation. One hundred uL of TE were added to each
817 cell pellet, and they were heated at 100 °C for 10 min. These solutions were used as DNA samples.
818 The PCR reaction was conducted as follows: 1 cycle at 94 °C for 2 min, 25 cycles each at 94 °C for 30
819 s, 60 °C for 30 s, 72 °C for 1 min, and 1 cycle at 72 °C for 10 min. The primer set of middle part of the
820 gene (P31_CtDV-Ari-F2: 5'-AGCAACCACAAATCCCCG-3'; P32_CtDV-Ari-R1: 5'-
821 ATTTTGAATAGCGTTTACAATGAG-3') and the primer set of last part of the gene
822 (CtD2V_VP2inF1: 5'-AATACTAGTGCCATTCGTCC-3', CtD2V_VP2outR2: 5'-
823 ACTACCATCATGATTGAGACTG-3'), they detect the gene for the viral capsid protein were used.
824 Each PCR product was purified with the ISOSPIN PCR Product (Nippon gene, Tokyo) and sequenced
825 on an ABI 3130 DNA Analyzer using the Sanger dideoxy method. Each DNA sequence was
826 combined and converted to amino acid sequences. The partial capsid protein amino acid sequences of
827 each virus were automatically aligned using Clustal W.

828 ## Hemagglutination inhibition (HI) assay

829 The HI test was performed according to classical procedures. The purified BSA, influenza virus (flu),
830 CtenDNAV-II and CtenRNAV-II were adjusted to 1 mg/mL in PBS. Twofold serial dilutions of these
831 samples were prepared for the assay, of which 25 µL of each dilution was incubated with 50 µL of 125
832 times PBS-diluted red blood cells from sheep blood (Håtunalab). The assay was conducted in a V-
833 bottom, 96-well plate. The plate was incubated for 1 hr at RT, and used for observing the
834 hemagglutination of the red blood cells.

835 ## Adsorption test

836 *Host diatom*

837 The virus adsorption kinetics of CtenDNAV-II and CtenRNAV-II was examined using the following
838 diatom strains: *Chaetoceros socialis* f. *radians* NIES3713, *C. tenuissimus* NIES-3714 and *C.*
839 *tenuissimus* NIES-3715. Previously reported compatibilities between host diatoms and viruses based
840 on (Kimura and Tomaru, 2015) are as following:

	CtenDNAV-II	CtenRNAV-II
C. socialis f. *radians* NIES3713 (L-4)	Not lysed	Lysed
C. tenuissimus NIES-3714 (2-6)	Lysed*	Not lysed
C. tenuissimus NIES-3715 (2-10)	Lysed	Lysed*

841 *original host-virus combination

842 *Viral inocula*

843 Exponentially growing cultures of *C. tenuissimus* NIES-3715 were inoculated with CtenDNAV-II and
844 CtenRNAV-II (0.1% v v^{-1}), which were stored at 4°C in the dark, and incubated for 7 d under the
845 growth conditions described above. Lysates were passed through a 0.2-µm polycarbonate membrane

filter (Whatman® Nuclepore Track-Etched Membranes, Merck KGaA, Darmstadt, Germany) to remove cellular debris and stored at 4°C. Filtered lysates were used as experimental inocula.

Virus enumeration

The number of infectious viral units was determined using the extinction dilution method (Suttle, 1993). Briefly, stored filtrates were rapidly thawed at 25°C, and then diluted with SWM3 medium in a series of 10-fold dilution steps. Aliquots (100 µL) of each dilution were added to 8 wells in cell-culture plates with 96 flat-bottom wells and mixed with 150 µL of an exponentially growing host algae culture, *C. tenuissimus* NIES-3715. The cell culture plates were incubated at 20°C under a 12:12 h light:dark cycle of 130–150 µmol photons·$m^{-2}s^{-1}$ (cool white fluorescent illumination) and cultures were monitored using an optical microscope (Nikon Ti, Tokyo, Japan) for up to 14 d for culture lysis. Culture lysis due to viral infection was usually observed as a complete degradation of the host cells in a well. Virus abundance was calculated from the number of wells in which algal lysis occurred using a BASIC programme as previously described (Nishihara et al., 1986).

Adsorption test

The host cell cultures were grown in SWM3 medium under a 12:12 h light:dark cycle of approximately 400 µmol of photons·$m^{-2}s^{-1}$ using white LED illumination at 25°C for *C. tenuissimus* (NIES-3714 and -3715) and 20°C for *C. socialis* f. *radians* (NIES-3713). Cell counts were performed immediately with no fixation using Fuchs-Rosenthal hemocytometer or an image-based cytometer (Tali® Image Cytometer, Thermo Fisher Scientific Ltd., Waltham, MA, USA) (Tomaru and Kimura, 2016). When the cultures reached stationary phase, cell concentration of >10^6 cells mL^{-1}, 20 mL of the culture was poured into a sterilized polystyrene flask (n=5). An aliquot of the culture was passed through a 0.2µm pore sized filter to remove the diatom cells. The resultant filtrate was poured into the polystyrene flask as a control (n=5). The virus suspensions were inoculated into these cell-suspended and filtrate cultures at 0.1%v/v (20µL). The virus inoculations were conducted during 1h and 2h after the light cycle started. The final virus titers in the cultures were set at 10^5~10^6 infectious units mL^{-1} (multiplicity of infections = 10^{-1}~10^0).

Aliquots of the cultures were sampled at just after virus inoculations (0h), 1h and 2h of post inoculation. The samples were used for cell counts and unadsorbed virus enumeration. The samples used for viral titer estimations (infectious units mL^{-1}) were passed through 0.2-µm filters (DISMIC®-25CS020AS, Advantec, Tokyo, Japan) to remove cellular debris. Filtrates were stored at -80°C until analysis.

The adsorption coefficient (Cd; mL min^{-1}) was determined as follows (Tarutani et al., 2006):

$$C_d = \frac{\alpha - \alpha_C}{N}$$

where α (min^{-1}) and α_c (min^{-1}) is the slope determined by linear regression for the natural logarithm of the percentage of free viruses in diatom cell culture and control culture against sampling time, respectively. N (cells mL^{-1}) is the host cell number.