Primordial Capsid and Spooled ssDNA Genome Structures Unravel Ancestral Events of Eukaryotic Viruses

Anna Munke,[a]* Kei Kimura,[b] Yuji Tomaru,[c] Han Wang,[a] Kazuhiro Yoshida,[d] Seiya Mito,[b] Yuki Hongo,[e] Kenta Okamoto[a]

[a]The Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden
[b]Department of Biological Resource Science, Faculty of Agriculture, Saga University, Saga, Japan
[c]Fisheries Technology Institute, Japan Fisheries Research and Education Agency, Hatsukaichi, Hiroshima, Japan
[d]Graduate School of Agriculture, Saga University, Saga, Japan
[e]Bioinformatics and Biosciences Division, Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Fukuura, Kanazawa, Yokohama, Kanagawa, Japan

ABSTRACT Marine algae viruses are important for controlling microorganism communities in the marine ecosystem and played fundamental roles during the early events of viral evolution. Here, we have focused on one major group of marine algae viruses, the single-stranded DNA (ssDNA) viruses from the *Bacilladnaviridae* family. We present the capsid structure of the bacilladnavirus *Chaetoceros tenuissimus* DNA virus type II (CtenDNAV-II), determined at 2.4-Å resolution. A structure-based phylogenetic analysis supported the previous theory that bacilladnaviruses 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. Furthermore, a low-resolution reconstruction of the CtenDNAV-II genome revealed a partially spooled structure, an arrangement previously only described for dsRNA and dsDNA viruses. Together, these results exemplify the importance of genetic recombination for the emergence and evolution of ssDNA viruses and provide important insights into the underlying mechanisms that dictate genome organization.

IMPORTANCE Single-stranded DNA (ssDNA) viruses are an extremely widespread group of viruses that infect diverse hosts from all three domains of life, consequently having great economic, medical, and ecological importance. In particular, bacilladnaviruses are highly abundant in marine sediments and greatly influence the dynamic appearance and disappearance of certain algae species. Despite the importance of ssDNA viruses and the last couple of years' advancements in cryo-electron microscopy, structural information on the genomes of ssDNA viruses remains limited. This paper describes two important achievements: (i) the first atomic structure of a bacilladnavirus capsid, which revealed that the capsid protein gene presumably was acquired from a ssRNA virus in early evolutionary events; and (ii) the structural organization of a ssDNA genome, which retains a spooled arrangement that previously only been observed for double-stranded viruses.

KEYWORDS algae virus, bacilladnavirus, capsid protein evolution, diatom virus, genome structure, horizontal gene transfer, nodavirus, ssDNA virus

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 dynamics of algal blooms, the global oxygen level, and the marine nutrient and energy cycling (1–3). The viruses typically have a very narrow host range, thus causing host-specific mortality and control of algae host populations (4).

Virus capsid evolution is an increasingly growing research field that relies both on

sequence and structural analysis, as well as on computational approaches (5–12). These studies range from narrow comparisons between specific virus groups (6, 8, 10, 11) to large-scale examinations across the entire virosphere (7) and have contributed to the identification of unique structural traits among certain viruses (8, 10–12) and the revelation of evolutionary relationships between seemingly unrelated viruses (5, 6, 9, 13). Moreover, a defined number of viral lineages have been identified based on capsid protein folds (5, 9), which were originally acquired from cells on multiple independent occasions (7). Since unicellular marine organisms were the earliest eukaryotes on earth, they were presumably host of the most ancient viruses (14, 15). Present-day viruses infecting unicellular organisms, such as unicellular algae, therefore likely retain genetic and structural characteristics from their ascendants (8, 16) and are consequently an essential group of viruses for interrogating viral capsid evolution.

Bacilladnaviruses is one of the major group of viruses infecting eukaryotic algae. They carry a circular ssDNA genome of ~6 kb, which is partially double stranded (~700 to 800 bp) and encodes three proteins; one coat protein, one replication-associated protein (Rep), and a third protein with unknown function (17). 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* based on their relatively conserved Rep protein (18). The Rep protein, suggested to be evolved from bacterial plasmids (19), has two functional domains, a His-hydrophobic-His endonuclease and a superfamily 3 helicase, and is involved in the virus genome replication, which is carried out by a rolling-circle mechanism (20). Contrary, the capsid proteins of CRESS DNA viruses are very diverse and have presumably been acquired from RNA viruses on multiple independent occasions (6, 13, 21–23). Specifically, bacilladnaviruses were suggested to have acquired their capsid proteins through horizontal gene transfer (HGT) from an ancestral noda-like virus (6). This is not an unreasonable scenario considering the prevalence of noda-like viruses in the aquatic environment (24). However, structural evidence has until now been pending.

Here, we present 3D reconstructions that reveal both the capsid and genome organizations of the bacilladnavirus *Chaetoceros tenuissimus* DNA virus type II (CtenDNAV-II) (17). An atomic model of the capsid protein could be constructed from the 2.4-Å resolution capsid structure. Structure-based phylogeny was used to demonstrate that the capsid of bacilladnaviruses indeed is structurally more similar to capsids of RNA viruses than to those of other ssDNA viruses, corroborating the HGT theory in early virus evolution. In addition, a low-resolution density map of the ssDNA genome suggests a partially spooled genome packaging mechanism, which has previously only been described for dsRNA and dsDNA viruses.

RESULTS AND DISCUSSION

Summary of structure determination. The structure of the CtenDNAV-II viron was determined using cryo-electron microscopy (cryo-EM). The capsid was reconstructed by imposing icosahedral symmetry (I4) and using 33,507 particles to an overall resolution of 2.4 Å using the "gold standard" Fourier shell correlation (FSC) 0.143 criterion (25, 26) (Fig. S1C). The local resolution of the capsid reconstruction was distributed between 2.3 and 9.1 Å and estimated using ResMap (27) (Fig. S1D). An atomic model of the capsid was built, refined, and validated according to the cryo-EM map. During 3D classification, it became apparent that possibly also the capsid interior possessed higher order structure (Fig. S2A), yet no density was observed in the final icosahedrally averaged high-resolution reconstruction. However, when employing a previously described method of subtracting the contribution of the capsid (28), followed by a number of 3D classification steps, it became apparent that parts of the genome formed an outer layer. Using a subset of 21,559 particles and without imposing any symmetry (C1), the outer genome layer could be reconstructed to 13 Å (FSC = 0.143 criterion). Figure S2B shows the FSC curves of the C1 reconstruction, where also the resolution criterion FSC = 0.5 (20 Å) is indicated as a comparison. An additional subtraction of the
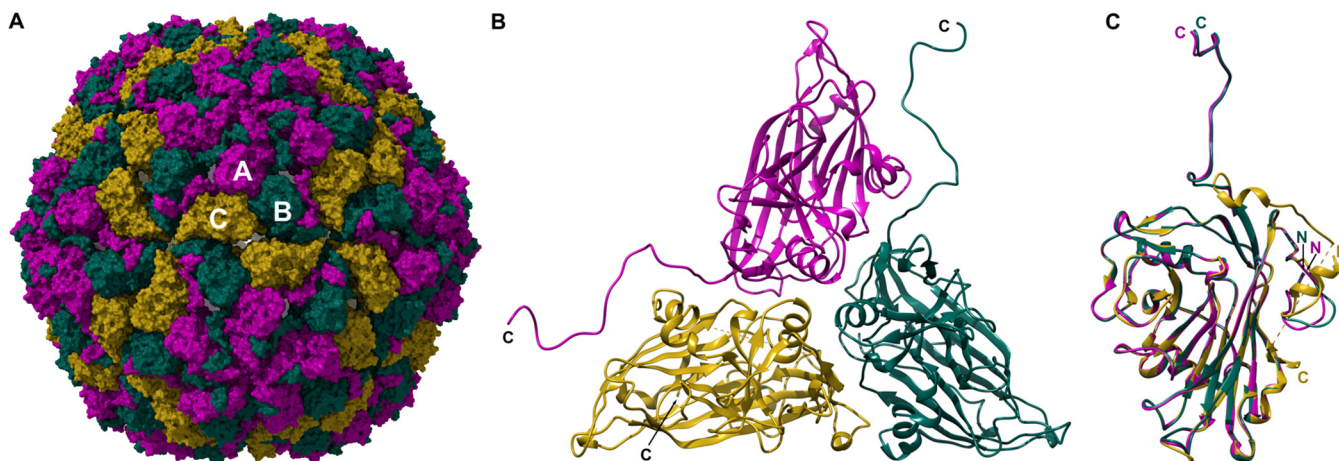
FIG 1 Atomic model of the CtenDNAV-II capsid. The three subunits A, B, and C are colored purple, green, and yellow, respectively, in panels A to C. (A) The entire capsid rendered with a surface representation viewed down an icosahedral 2-fold axis. (B) The secondary structure of one single asymmetric unit viewed from the outside. (C) Superimposition of the three subunits.

outer genome layer was attempted to reconstruct the core; however, the even lower resolution of the resulting reconstruction made it uninterpretable and unreliable. Data acquisition and processing, refinement, and validation statistics are summarized in Table S1.

Different conformations of capsid proteins 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 quasiequivalent positions termed A, B, and C (Fig. 1A and B). For subunit A and B, residues 64 to 371 were modeled, and for subunit C, residues 60 to 365 and 378 tp 384 were modeled (Fig. 2C). The A and B subunits were close to identical (Fig. 1C): 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 modeled (Fig. 2C); however, additional density was visible in the map when the contour level was decreased (Fig. S3), indicating that the C terminus forms flexible protrusions on the capsid surface. The termini of subunit C differs from the A and B subunits (Fig. 1C). Contrary to the long tail toward the capsid surface, the C terminus of the C subunit is directed toward the capsid interior (Fig. 1C and 6). The N termini of all subunits are presumably located on the capsid interior; however, the first 59 to 63 residues could not be modeled (Fig. 2C). Numerous positively charged residues can be found in the N terminus, thus possibly interacting with the genome. Residues 366 to 385 were only partially modeled, and the last residues (386 to 390) were unmodeled in all subunits (Fig. 2C), suggesting different conformations and/or flexibility. The N terminus of the C subunit has a small α-helix on the capsid interior that was not present in the other two subunits (Fig. 1C). The modeling of the C subunit was aided by a structure prediction (Fig. S4) from the AlphaFold package (29). The termini of the predicted model were more similar to the C subunit than to the A and B subunits. The N terminus of the AlphaFold model was confidently predicted (pLDDT >90) from H60, which was also the first residue of the C subunit that could be confidently modeled based on the cryo-EM map (Fig. 2C). As expected, based on the experimental data, the first 59 residues of the N terminus of the predicted structure are disordered. The C subunit was confidently modeled based on the cryo-EM map until R365 (Fig. 2C). With guidance from the AlphaFold model, another seven residues (D378 to F384) could be modeled. Residues 366 to 377 could not be confidently modeled based on the experimental data and are thus not included in the final model (PDB: 7NS0); however, based on the appearance of the cryo-EM density and the predicted model that has a pLDDT score ranging between 77 and 93 in this part, an α-helix is likely located there (Fig. S4). The modeled termini and corresponding cryo-EM map are displayed in Fig. S1E.
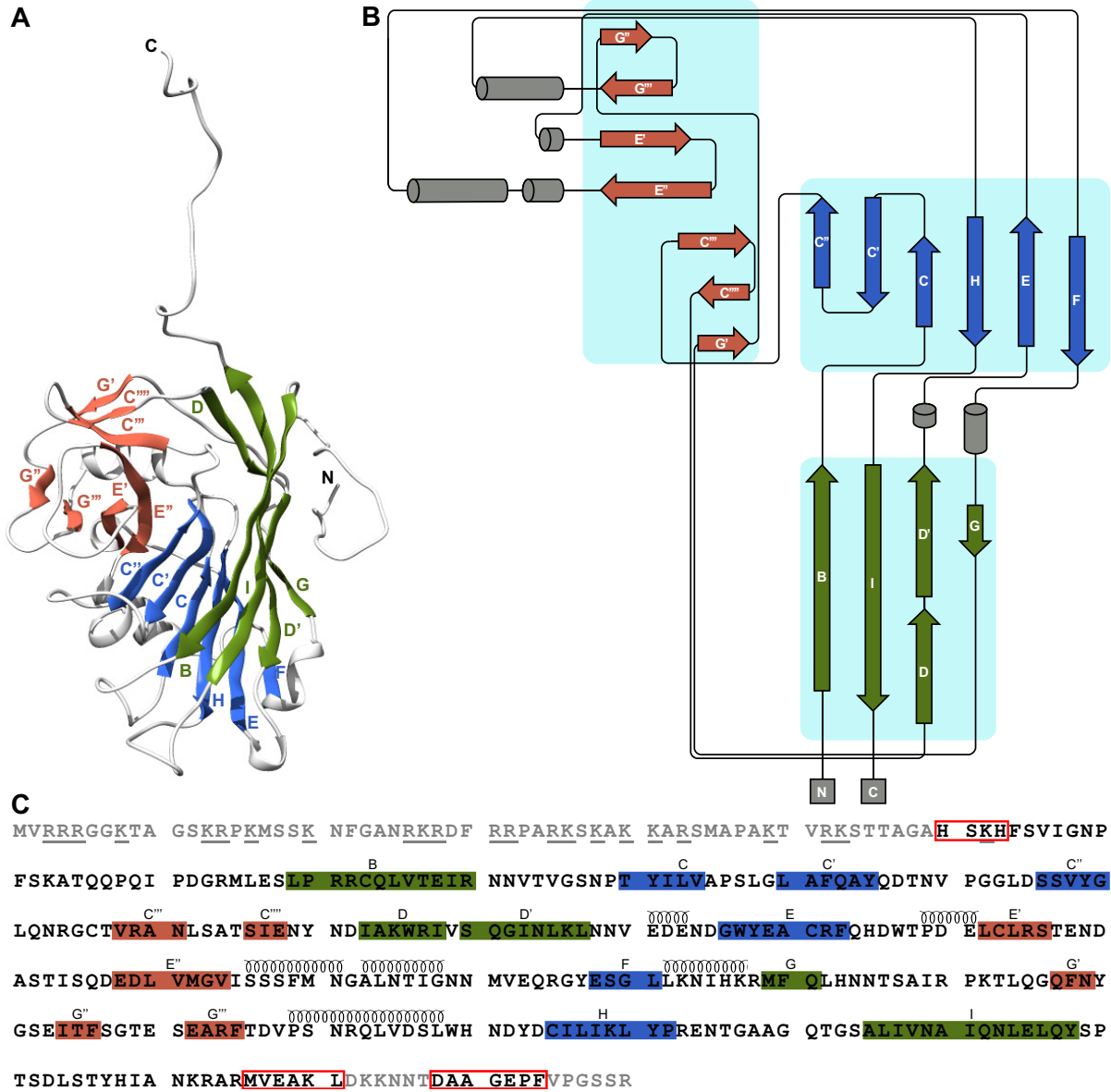
FIG 2 Capsid protein topology and structural organization. The β-strands are colored according to which β-sheet they belong to. The β-strands within the green and blue β-sheets are named alphabetically according to the conventional jelly-roll fold nomenclature (B to I). (A) The secondary structure of 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. Modeled and unmodeled residues are colored black and gray, respectively. Residues highlighted with red rectangles were partly unmodeled: H60 to H63 in subunits A and B, M366 to L377 in subunit C, and D378 to F384 in subunits A and B. The assigned secondary structure is shown schematically above the sequence. The underlined residues in the unmodeled N terminus indicate the numerous positively charged residues.

Unique features of the CtenDNAV-II capsid protein jelly-roll fold. 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. (30, 31). For CtenDNAV-II, the two sheets are formed by strands BIDD'G and C"C'CHEF, respectively (Fig. 2A to 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
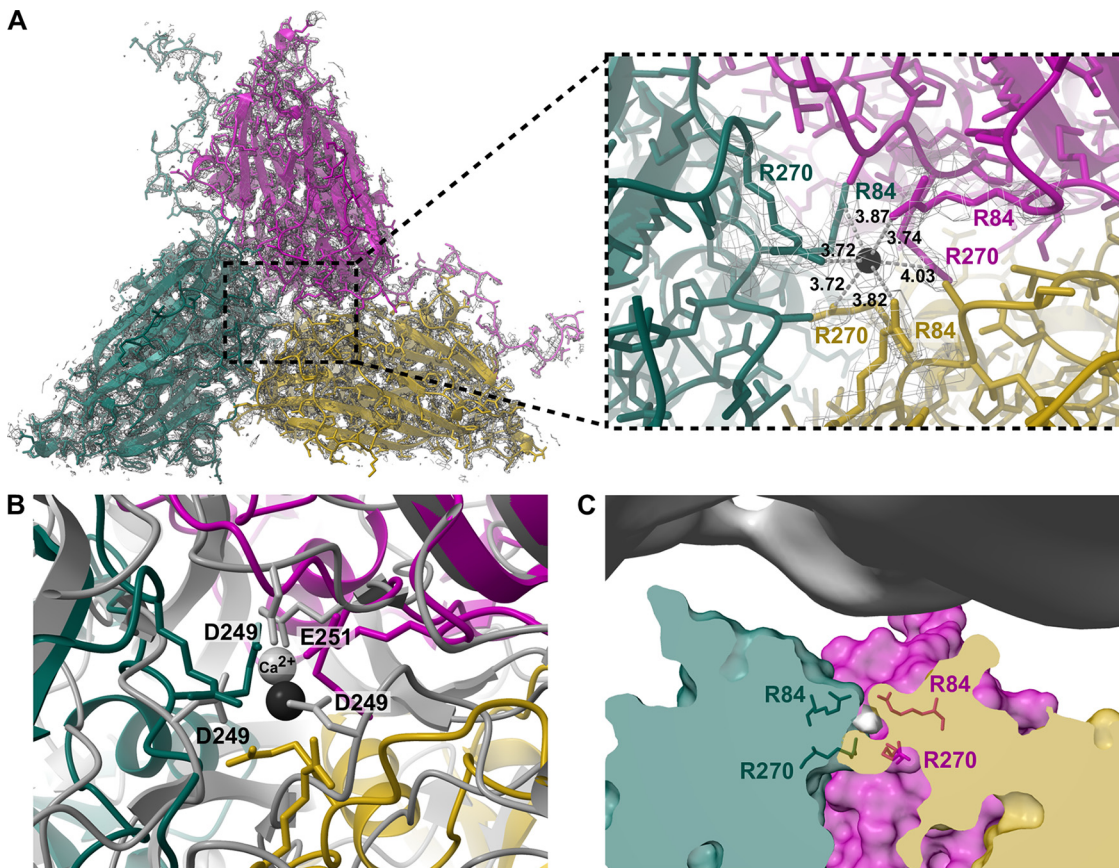
FIG 3 Unassigned density in the subunit interface. (A) The model of a single icosahedral protomer viewed from the inside of the capsid and corresponding cryo-EM map is visualized to the left. To the right is a closeup view of the subunit interface. A marker (black) was placed in the center of the unmodeled density using Chimera X to measure the distances to the surrounding arginine residues. (B) Same view as in panel A, but aligned with Flock House virus (*Nodaviridae*) (PDB: 4FTB). Putative amino acids interacting with Ca^{2+} (gray ball) are labeled and shown as sticks. (C) A clipped side view of an icosahedral protomer with unmodeled density in gray surrounded by arginine residues. The map from the outer genome layer is shown in the top.

two C-strands (C‴ and C⁗), two E-strands (E′ and E″), and three G-strands (G′, G,″ and G‴) (Fig. 2B). Notably, AlphaFold accurately and confidently predicted the jelly-roll fold as well as the fold of the additional β-sheet, which had pLDDT scores of >90 and >70, respectively (Fig. S4). In conclusion, the capsid protein of CtenDNAV-II has three unique features: three additional strands in the jelly-roll fold, an extra surface-exposed β-sheet, and an unusual conformation of the C termini of subunit A and B that extends out from the jelly-roll core as long tails toward the capsid surface.

Unmodeled density blob in interface between capsid protein subunits. An unmodeled density blob was visible in the interface between the three subunits that constitute one protomer (Fig. 3A). Metal ions, commonly Ca^{2+} (32, 33) but also Zn^{2+} (34), are sometimes found in the interface of capsid protein subunits and have shown to be important for viron stability. The unmodeled density in the subunit interface of CtenDNAV-II has a pyramidal to spherical shape (Fig. 3C) and is surrounded by six arginine residues (R84 and R270) that resemble an octahedral coordination geometry (Fig. 3A). The distance from the center of the density to surrounding nitrogen atoms on the arginine sidechains is 3.7 to 4.0 Å. The long distance and the fact that arginine residues would be in the coordination sphere suggest that a metal ion at this position is unlikely (35). However, the blob is the last density that remains when the contour level is increased, which strongly suggest that it is a metal ion. A calcium ion has been reported in nodaviruses, such as Pariacoto virus and Flock House virus (36), at a close to identical location (2.6 Å apart) to the unmodeled blob in CtenDNAV-II (Fig. 3B), which is interesting from an evolutionary perspective, as described later. Considering

the location of the unmodeled density, close to the inside of the capsid (Fig. 3C), and the positively charged arginine sidechains, another possibility could be that the density blob originates from a piece of genome. However, the unmodeled blob appears isolated and structurally obstructed by R84 from the capsid interior (Fig. 3C). A multiple sequence alignment (Fig. S5) shows that a positively charged residue (either arginine or lysine) is rather common (16 out of 24 sequences) among bacilladnaviruses at the corresponding position of R84. The second arginine, R270, is very unusual at the corresponding position; however, several sequences have a histidine one amino acid upstream, which potentially could serve the same function. Future research will have to discern the true nature of the blob.

Structure-based phylogeny reveal evolutionary relationship to ssRNA viruses. Previous structures of so-called CRESS DNA viruses (phylum *Cressdnaviricota*) (18) include viruses from families *Ciroviridae* (e.g., 3R0R and 5ZJU), *Geminiviridae* (e.g., 6F2S and 6EK5), and *Nanoviridae* (6S44), whose capsids follow T = 1 symmetry. The capsid proteins of these three families also contain a jelly-roll domain but lack the third surface-exposed β-sheet and C-terminal tail found in CtenDNAV-II. The DALI web server (37) was used as an initial step to investigate the previous theory that ssDNA viruses have acquired their capsid protein from ssRNA viruses (6, 21). Indeed, the result indicated that the CtenDNAV-II capsid protein is more similar to capsid proteins of ssRNA viruses than to other ssDNA viruses (Table S2). The closest ssDNA virus was that of Beak and feather disease virus (*Circoviridae*), which ended up oin 11th place (z-score 8.7) behind 10 RNA viruses. Highest similarity was found between CtenDNAV-II and ssRNA viruses from families *Carmotetraviridae*, *Alphatetraviridae*, and *Nodaviridae*, which had z-scores of 14.4 to 15.3 (Table S2). Notably, only nodaviruses from the genus *Alphanodavirus* were included and none from *Betanodavirus* or the informal group of gamma-nodaviruses. Z-scores below 2 are considered insignificant, scores between 2 and 8 are a gray zone, scores between 8 and 20 indicate that two structures probably are homologous, and with a z-score above 20 they are definitely homologous. For an in-depth explanation of the DALI method and z-scores, see Holm (38).

A structure-based phylogenetic analysis was used to further investigate the relationship between the capsid proteins of CtenDNAV-II and ssRNA viruses. In summary, structures with DALI-scores >8 and three additional CRESS DNA viruses were used together with the CtenDNAV-II capsid protein structure as inputs to the program MUSTANG (39). A complete list of structures used for the analysis is shown in Table S2. The root mean square deviation (RMSD) values provided by MUSTANG were then used to create phylogenetic trees (Fig. 4 and Fig. S6A). The RMSD matrices created by MUSTANG that were used to create the phylogenetic trees are available in Data Set S1. The compared structures have various polypeptide lengths, ranging from 172 amino acids for Faba bean necrotic stunt virus (*Nanoviridae*) to 644 amino acids for Nudaurelia capensis omega virus (*Alphatetraviridae*), and thus the longer structures e.g., from CtenDNAV-II, carmotetraviruses, alphatetraviruses, nodaviruses, and birnaviruses have additional folds and domains in addition to their jelly-roll fold core (Fig. S7). The phylogenetic analysis was therefore performed using only the jelly-roll fold core, for which the result is presented in Fig. 4. The selected amino acid numbers for each structure in the jelly-roll fold comparison are listed in Table S2. Figure 4 displays three clades with different ssDNA viruses in each clade. CtenDNAV-II is placed together with RNA viruses from families *Nodaviridae*, *Alphatetraviridae*, *Carmotetraviridae*, and *Birnaviridae* (Clade 1), thus corroborating the result from the DALI search (Table S2) as well as the previous HGT theory by Kazlauskas et al. (6). The ssDNA circoviruses were organized in a second clade together with ssRNA viruses from families *Tombusviridae*, *Hepaviridae*, and *Solemoviridae*, whereas the other two ssDNA families *Geminiviridae* and *Nanoviridae* were placed alone in the third clade. As a comparison, the phylogenetic analysis was also performed by using full-length chains (Fig. S6A). The result is to some extent similar: CtenDNAV-II and nodaviruses still share structural similarities, and CtenDNAV-II and other ssDNA viruses are organized into different clades; however, the details in the trees differ between the two approaches. For example, only CtenDNAV-II and the nodaviruses appear related when using the full-length chains, whereas carmotetraviruses, alphatetraviruses, and birnaviruses are
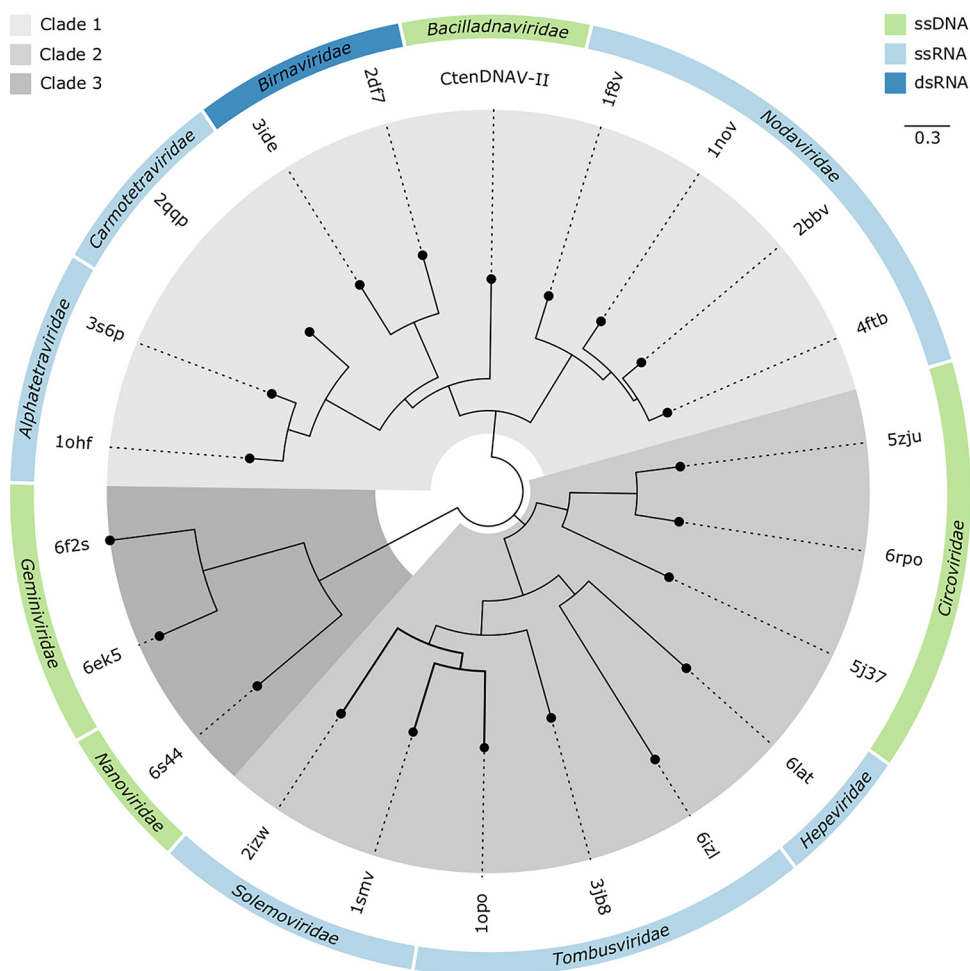
FIG 4 Structure-based phylogenetic tree of the jelly-roll folds. The three major clades are highlighted in different shades of gray. The tree is reconstructed based on RMSD values from superpositions performed on the common jelly-roll fold core. The chain identity and residues included for each structure are listed in Table S2. A phylogenetic tree on full-length chains is shown in Fig. S6A. The RMSD tables of the jelly-roll fold and full-length superpositions are available in Data Set S1.

placed in the other clades. However, our opinion is that Fig. 4 better represents the evolutionary relationship between these viruses, since the jelly-roll fold, which is the conserved structural core and a better representation of the primordial fold, is in some cases misaligned when using the full-length chains (see example in Fig. S6B).

The results presented support the theory of an acquisition of the capsid protein gene in a bacilladnavirus ancestor from a ssRNA noda-like virus via a HGT event, thus supporting the so-called RNA-to-DNA jump scenario for the origin of DNA viruses with jelly-roll capsid proteins (40). Other alternative scenarios include the independent emergence (i.e., convergent evolution) and the gradual transition scenarios (40). While the latter two cannot be completely ruled out, the first scenario is strongly supported for DNA jelly-roll fold containing viruses, for example, due the fact that several different DNA viruses, including bacilladnaviruses, apparently contain capsid proteins that are more similar, on both sequence and structural level, to different RNA viruses, while their rolling-circle replication proteins are polyphyletic (6, 13, 21–23, 41). Findings supporting that ssRNA and ssDNA recombination is plausible include the fact that nucleic acid packing can be unspecific (21, 42) and that single-stranded genomes have similar persistence lengths (43). For more detailed discussions on this topic, see Holmes (44), Krupovic (40), and Moreira and López-García (45).

Putative emergence of ssRNA, dsRNA, and ssDNA viruses. The organization of ssDNA viruses into three different clades supports a polyphyletic origin of ssDNA
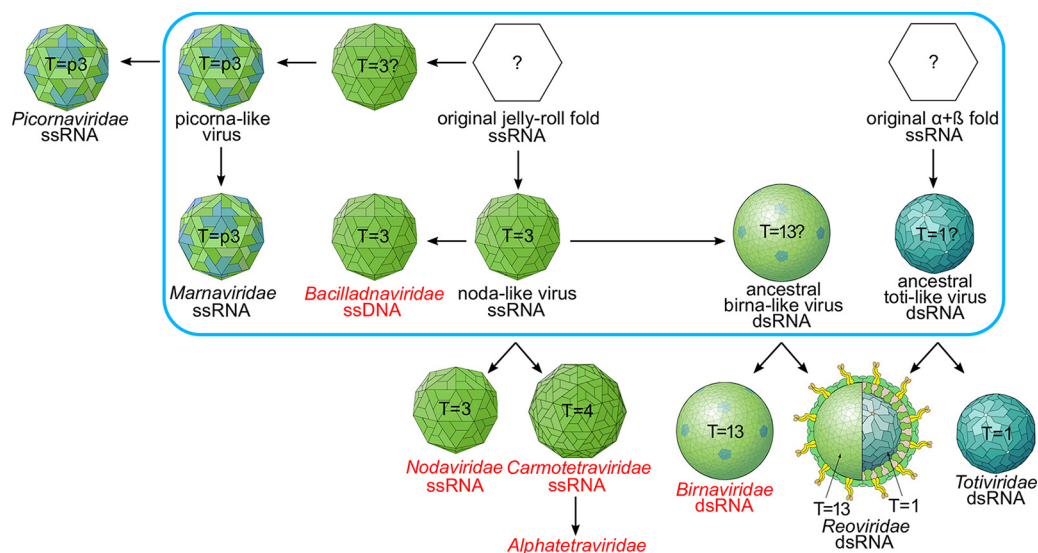
FIG 5 A scheme depicting the putative emergence of virus capsids carrying different nucleic acid genomes. The figure illustrates, based on current knowledge and hypotheses, the most relevant virus groups and the evolutionary relationship between their virus capsids. Viruses in Clade 1 (Fig. 4) are highlighted in red. Events that likely took place at an early stage of evolution in ancient algal pools are circled in blue. The figure is based on results described in this paper as well as previous results: the relationship between bacilladnaviruses, nodaviruses, carmotetraviruses, alphatetraviruses, and binaviruses is described herein as well as in references 6, 21, 46, 47; the relationship between the dsRNA viruses is described in reference 46; the evolution of picornaviruses were described in reference 8; the link between the ssRNA, ssDNA, and dsDNA is described in references 15, 19, 21; and the presence of birna-like and toti-like viruses in the oceans is described in references 78–80. The viron pictures were derived from ViralZone, SIB Swiss Institute of Bioinformatics (81) (https://viralzone.expasy.org/) licensed under a Creative Commons Attribution 4.0 International License.

viruses and the independent acquisition of capsid protein genes from different RNA viruses (21). Our findings (Fig. 4) also support the previously described close evolutionary relationship between virus families *Carmoteraviridae*, *Alphatetraviridae*, *Nodaviridae*, and *Birnaviridae* (21, 46, 47). The capsid proteins of the T = 4 tetraviruses have been proposed to be derived from the T = 3 nodaviruses, and in particular, there is a higher degree of similarity between carmotetraviruses and nodaviruses than between alphatetraviruses and nodaviruses, suggesting that carmotetraviruses are closer than alphatetraviruses to the primordial T = 4 virus (47). The dsRNA birnaviruses (T = 13) have been suggested to have a noda/tetravirus-like ancestor (46), which exemplifies the evolutionary link between ssRNA and dsRNA viruses. In turn, an ancestral birna-like virus could have been the precursor of the *Reoviridae* outer capsid layer, whereas the inner layer that has T = 1 symmetry, might have been acquired from an ancestral toti-like virus (46). To put our new structural insights into a broader context, we have updated the putative emergence of viruses harboring different genomes types (Fig. 5).

Acquired and retained structures in capsids of CtenDNAV-II and ssRNA viruses. The jelly-roll fold of CtenDNAV-II aligns well with the other jelly-roll folds in Clade 1, with RMSD values of 1.4 to 1.7, compared to the jelly-roll folds in the other clades that had RMSD values of 2.2 to 3.6 (Data Set S1). As expected, the differences between the capsid proteins from CtenDNAV-II and carmotetraviruses, alphatetraviruses, nodaviruses, and birnaviruses lie primarily in surface exposed areas and to some extent in the regions facing the capsid interior (Fig. 6 and Fig. S7), i.e., structural elements that have evolved based on the host and on genome packing requirements, respectively. All capsid proteins in Clade 1 have specific surface structures in addition to the jelly-roll fold (Fig. 6 and Fig. S7). While the surface structures of CtenDNAV-II and nodaviruses are formed by extensions of loops in the jelly-roll, carmotetraviruses, alphatetraviruses, and birnaviruses instead have larger separate domains (called Ig-like and P-domains) inserted between two jelly-roll strands (Fig. 6). Unique for CtenDNAV-II is the long C-terminal tail in subunits A and B that extends to the capsid surface, while the C terminus of the capsid protein from the other viruses can be found on the
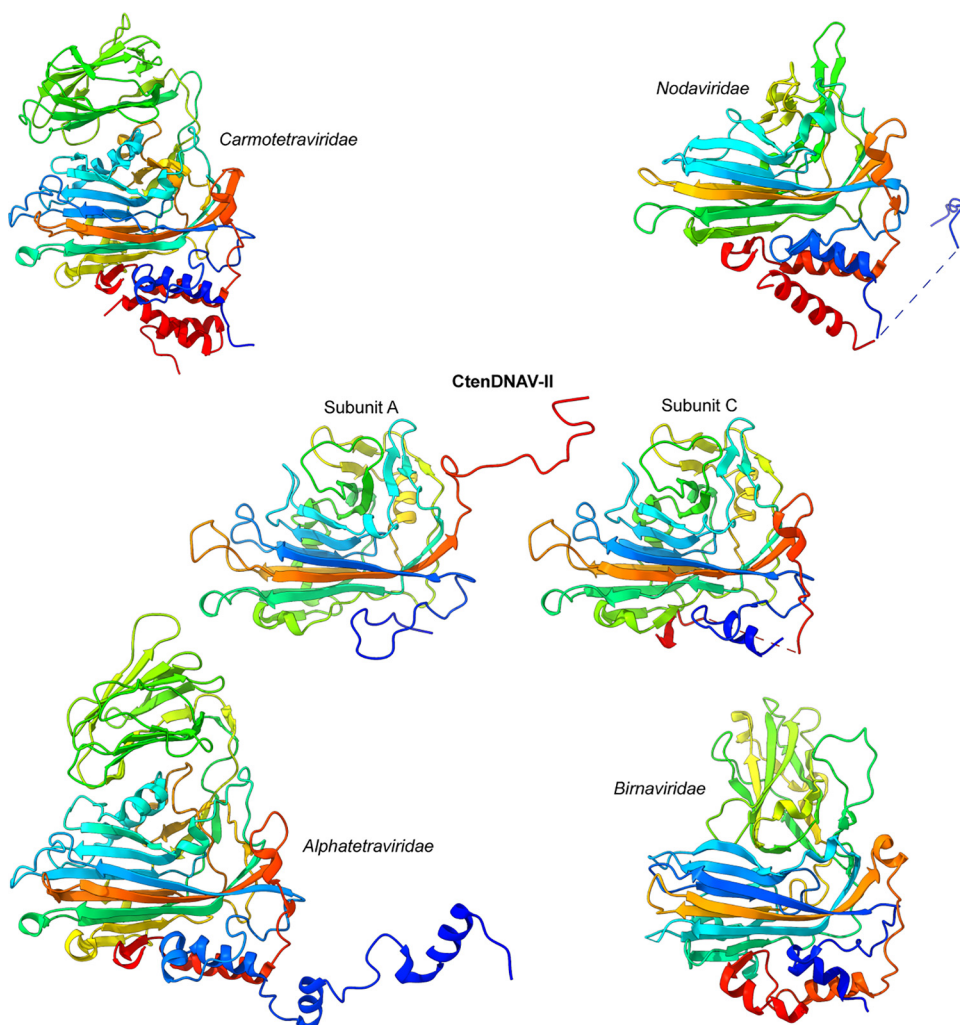
FIG 6 Comparison of capsid proteins in Clade 1. Representatives of capsid proteins in Clade 1 viewed from the side with surface and interior structures on top and bottom, respectively. The chains are colored as rainbow from blue (N terminus) to red (C terminus). Figure S7 shows the same images, but colored according to domain.

capsid interior similarly as the C terminus of the C subunit in CtenDNAV-II (Fig. 6). The acquired surface features for each virus family (Fig. S7), such as the Ig-like domain of tetraviruses and the P domain of birnaviruses, have been described as putative receptor-binding domains of those viruses (46–50). The unique surface features of CtenDNAV-II include the third β-sheet and the long C-terminal tails of the A and B subunits. Similar to RNA viruses, these acquired features of CtenDNAV-II could potentially be important for host recognition; however, future studies will have to confirm this theory.

Carmotetraviruses, alphatetraviruses, nodaviruses, and birnaviruses have α-helices located on the capsid interior that are formed by the capsid protein termini, which in some structures have been shown to interact with the viral RNA (47). Of the three CtenDNAV-II subunits, the C subunit shows highest similarity to the capsid proteins of the other viruses in Clade 1: both termini are located on the capsid interior and the termini forms α-helices (Fig. 6). The model of the C subunit has one α-helix in the N terminus; however, based on the cryo-EM map and the predicted model by AlphaFold, also the C terminus is likely to form an α-helix (Fig. S4). Superpositions of the capsid proteins from families *Carmotetraviridae*, *Alphatetraviridae*, *Nodaviridae*, and *Birnaviridae* with the predicted model of the CtenDNAV-II capsid protein show that the two predicted terminal α-helices align well with α-helices from the experimentally determined structures. More and longer helices are found among the RNA viruses (Fig. 6 and
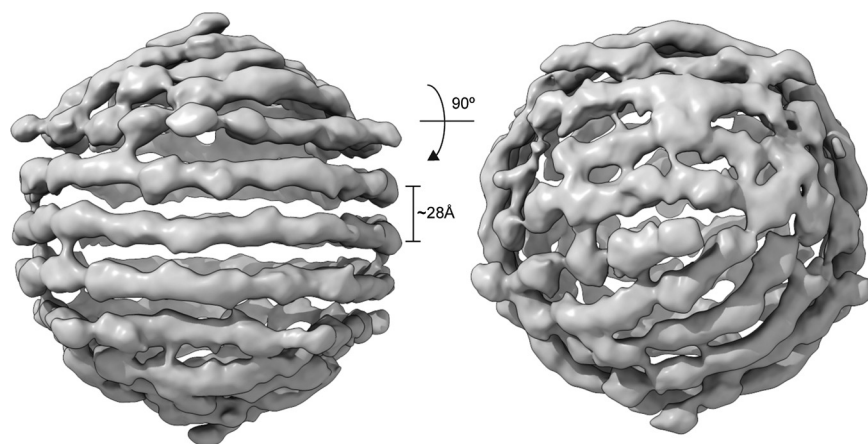
FIG 7 Outer genome layer of CtenDNAV-II. Left: the coil of three turns is visualized in the center of the image, where the individual parallel turns are separated ~28 Å. Right: the image on the left has been rotated 90°.

Fig. S7), and the two below the C subunit could thus be remnants from the noda-like virus ancestor, whereas the C terminus in the A and B subunits has evolved to form other structural elements on the capsid surface as a host adaptation. Thus, the C subunit of CtenDNAV-II presumably resembles the primordial bacilladnavirus fold more than subunit A and B do. Finally, an additional similarity is found in the N terminus of CtenDNAV-II and the ssRNA viruses in Clade 1. Due to disorder, up to 63 residues of the highly positively charged terminus could not be modeled for CtenDNAV-II (Fig. 2C). Likewise, the first ~40 to 70 residues, also heavily populated with arginine and lysine residues, are unmodeled for the ssRNA viruses. This is in contrast to the dsRNA birnaviruses, whose N terminus is largely modeled (missing 5 to 10 residues) and less charged. While differences in surface features between the viruses in Clade 1 could reflect variations in host specificity, the differences on the capsid interior could possibly reflect variations in genome organization. Structural information on the genomes of the RNA viruses in Clade 1 is limited to short nucleotide (nt) segments, which have been revealed even when the icosahedral symmetry has not been broken during the structure determination (36, 47). Despite similarities in the capsid interior among viruses in Clade 1, no DNA segments were found inside CtenDNAV-II during the capsid reconstruction, suggesting that noda-, tetra-, and birnaviruses interact more specifically with their genome than CtenDNAV-II does and that their genomes are organized structurally different. Perhaps the observed adaptations of the CtenDNAV-II capsid interior, with fewer and shorted α-helices, were enough to allow packing a genome that interact nonspecifically with the capsid.

CtenDNAV-II genome is partially spooled. The reconstruction of the outer genome layer (EMDB-12555) displays a coil of three turns (Fig. 7, left), which were readily visible also in 2D classes (Fig. S2C), and on each side of the three turns are additional DNA fragments that do not follow the same spooling arrangement (Fig. 7, right). The spooled genome packaging has previously only been described in viruses with double-stranded genomes (28, 51, 52), whereas the genomes of ssRNA viruses forms secondary structures that organize into branched networks (53–55) or a dodecahedral cage (36, 56).

Considering that spooled genome arrangements have only been observed among viruses with double-stranded genomes and that the distance between the parallel turns are about 28 Å (Fig. 7, left), similar to double-stranded genomes (28, 51, 57), the outer genome layer of CtenDNAV-II could potentially be at least partially formed by double-stranded DNA. In fact, the circular ssDNA genome of CtenDNAV-II includes a partially double-stranded region of 669 bp (17). The low resolution of the outer layer reconstruction prevented precise modeling; however, by manually combining short (5 to 20 nt) A-T base pairs of ideal B-form dsDNA into the density of the spool with three turns, a model with 630 bp could be

obtained (Fig. S2D). Thus, hypothetically, the observed spool could correspond to the double stranded region of the genome. However, future bacilladnavirus genome structures at higher resolution will have to confirm present observations and hypotheses as well as if additional base-pairing is formed that produce the semispooling pattern observed on each side of the three central turns.

Previous cryo-EM studies have revealed different types of genome organization for non-enveloped icosahedral viruses that often reflect the physical properties of the genome and the viron assembly mechanisms. Viruses containing double-stranded genomes (dsDNA and some dsRNA) use an NTP-driven motor to form spooled genome structures within a preassembled capsid (28, 51, 52). The spooled genomes are packaged by flexible interactions between the capsid protein and the genome. The interactions are mediated by small contacts with hydrophobic and/or positively charged amino acid residues of their capsid proteins. Segmented dsRNA reoviruses instead form nonspooled or partially spooled genomes with pseudo-D3 symmetry that interact with the RNA-dependent RNA polymerase (57–60). Genomes of ssRNA viruses form secondary structures that can either form a branched network, such as in *Leviviridae* viruses (53–55), or a dodecahedral cage, such as in *Secoviridae* and *Nodaviridae* viruses (36, 56). Contrary to the large double-stranded genomes that arrange within a preformed capsid, the small ssRNA genomes allow a simultaneous and cooperative genome packing and capsid assembly that is governed by specific interactions between the genome and capsid protein (55, 56). Less is known about the genomes of ssDNA viruses; however, similar to viruses with other genome types, the genomes of ssDNA viruses are strongly connected to the packing mechanism (61–63). Some ssDNA viruses, such as phages from the *Microviridae* family, have been shown to combine genome packing properties of both double-stranded viruses and ssRNA viruses by performing genome replication while simultaneously packing the newly synthesized strand into preformed capsids (64). Unlike double-stranded viruses, microviruses do not require additional energy from NTP-driven packaging motors since their genomes do not require as dense packaging (65). Previous structural information on the genomes of ssDNA viruses have been limited to short (<10 nt) nucleotide segments but have revealed specific interactions between the capsid and DNA of single-stranded type (61, 62, 66). The genomes of ssDNA viruses have the capability to form biologically functional secondary structures similar to ssRNA viruses (67) and since the capsid gene of bacilladnaviruses has been horizontally transferred from ssRNA viruses, a similar structural arrangement of the genome as in ssRNA viruses is imaginable also for ssDNA viruses. However, the structure of the CtenDNAV-II genome is much more similar to the spooled genomes found in double-stranded viruses (Fig. 7), thus raising the question whether the CtenDNAV-II capsid preassembles before packing the genome, such as the ssDNA microviruses (64). Interestingly, rod-shaped virus-like particles have been found together with CtenDNAV-II particles in infected host cells and were suggested to be precursors of mature virons (17). Future studies will be needed to unravel the exact mechanisms behind bacilladnaviron assembly and genome packing.

MATERIALS AND METHODS

Virus production and purification. CtenDNAV-II was produced as previously described (17). The crude virus suspension was loaded onto 15% to 50% (wt/vol) 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 (68), and contrast transfer function (CTF) parameters were estimated using Gctf 1.06 (69). The RELION 3.1 package (70) was used for particle picking, 2D and 3D classifications, *de novo* 3D model generation and refinement. The capsid reconstruction was further sharpened with DeepEMhancer (71) by using the two half-maps from the refinement. The 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) (25, 26). The data set and image processing are summarized in Table S1.

To reconstruct the genome without icosahedral symmetry a similar procedure to what has been described by Ilca et al. (28) 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.

Model building and refinement. The atomic model of CtenDNAV-II capsid protein was manually built into the density map using Coot (72). The model was further improved through cycles of real-space refinement in PHENIX (73) with geometric and secondary structure restraints, and subsequent manual corrections by Coot were carried out iteratively. Refinement and validation statistics are summarized in Table S1. The model of the outer genome layer was constructed by manually combining short (5 to 20 nt) A-T base pairs of ideal B-form dsDNA into the density using UCSF Chimera X (74). The structure prediction was carried out using AlphaFold 2.0 (29).

Structure analysis. Structural comparison of the CtenDNAV-II capsid protein was initially carried out by the DALI web server as a heuristic search against all structures (as of 2021-07-05) in the PDB (37). Unique viruses (i.e., not unique PDB entries) from the DALI search with z-scores >8 were used for structure-based multiple alignments using the program MUSTANG (39). In addition, three other CRESS DNA viruses with known structure (6S44, 6EK5 and 6F2S) were included in the MUSTANG analysis. Detailed information on the structures used for the comparison can be found in Table S2. The RMSD values provided by MUSTANG (Data Set S1) were used to create phylogenetic trees using the Neighbor-joining method (75) in MEGA X (76, 77), which were further aesthetically modified with FigTree. Figures were prepared using UCSF Chimera X (74).

Data availability. The atomic model and map of the CtenDNAV-II capsid was deposited in the PDB (7NS0) and EMDB (12554). The EMDB entry provides the full map, as well as the half-maps, mask, and map sharpened by DeepEMhancer (71). The map of the cryo-EM reconstruction of the outer genome layer was deposited in the EMDB (12555).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, TXT file, 0 MB.
FIG S1, PDF file, 2.8 MB.
FIG S2, PDF file, 1.9 MB.
FIG S3, EPS file, 2.7 MB.
FIG S4, EPS file, 1.8 MB.
FIG S5, EPS file, 0.1 MB.
FIG S6, TIF file, 1.2 MB.
FIG S7, EPS file, 1.5 MB.
TABLE S1, DOCX file, 0.01 MB.
TABLE S2, EPS file, 0.2 MB.

REFERENCES

1. Fuhrman JA. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548. https://doi.org/10.1038/21119.
2. Suttle CA. 2007. Marine viruses—major players in the global ecosystem. Nat Rev Microbiol 5:801–812. https://doi.org/10.1038/nrmicro1750.
3. Wilhelm SW, Suttle CA. 1999. Viruses and nutrient cycles in the sea. Bioscience 49:781–788. https://doi.org/10.2307/1313569.
4. Brussaard CPD. 2004. Viral control of phytoplankton populations–a review. J Eukaryot Microbiol 51:125–138. https://doi.org/10.1111/j.1550-7408.2004.tb00537.x.

5. Abrescia NGA, Bamford DH, Grimes JM, Stuart DI. 2012. Structure unifies the viral universe. Annu Rev Biochem 81:795–822. https://doi.org/10.1146/annurev-biochem-060910-095130.

6. Kazlauskas D, Dayaram A, Kraberger S, Goldstein S, Varsani A, Krupovic M. 2017. Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses. Virology 504:114–121. https://doi.org/10.1016/j.virol.2017.02.001.

7. Krupovic M, Koonin EV. 2017. Multiple origins of viral capsid proteins from cellular ancestors. Proc Natl Acad Sci U S A 114:E2401–E2410. https://doi.org/10.1073/pnas.1621061114.

8. 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. https://doi.org/10.1128/JVI.01855-19.

9. Nasir A, Caetano-Anollés G. 2017. Identification of capsid/coat related protein folds and their utility for virus classification. Front Microbiol 8:380. https://doi.org/10.3389/fmicb.2017.00380.

10. 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. https://doi.org/10.1016/j.str.2020.04.016.

11. 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. https://doi.org/10.1038/srep33170.

12. Wang X, Ren J, Gao Q, Hu Z, Sun Y, Li X, Rowlands DJ, Yin W, Wang J, Stuart DI, Rao Z, Fry EE. 2015. Hepatitis A virus and the origins of picornaviruses. Nature 517:85–88. https://doi.org/10.1038/nature13806.

13. 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. https://doi.org/10.1186/1745-6150-7-13.

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

15. Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology 479–480:2–25. https://doi.org/10.1016/j.virol.2015.02.039.

16. Koonin EV, Dolja VV. 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. Microbiol Mol Biol Rev 78:278–303. https://doi.org/10.1128/MMBR.00049-13.

17. Kimura K, Tomaru Y. 2015. Discovery of two novel viruses expands the diversity of single-stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine diatom. Appl Environ Microbiol 81:1120–1131. https://doi.org/10.1128/AEM.02380-14.

18. Krupovic M, Varsani A, Kazlauskas D, Breitbart M, Delwart E, Rosario K, Yutin N, Wolf YI, Harrach B, Zerbini FM, Dolja VV, Kuhn JH, Koonin EV. 2020. *Cressdnaviricota*: a virus phylum unifying seven families of Rep-encoding viruses with single-stranded, circular DNA genomes. J Virol 94:e00582-20. https://doi.org/10.1128/JVI.00582-20.

19. Kazlauskas D, Varsani A, Koonin EV, Krupovic M. 2019. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. Nat Commun 10:3425. https://doi.org/10.1038/s41467-019-11433-0.

20. Zhao L, Rosario K, Breitbart M, Duffy S. 2019. Eukaryotic Circular Rep-Encoding Single-Stranded DNA (CRESS DNA) viruses: ubiquitous viruses with small genomes and a diverse host range. Adv Virus Res 103:71–133. https://doi.org/10.1016/bs.aivir.2018.10.001.

21. Krupovic M. 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Curr Opin Virol 3:578–586. https://doi.org/10.1016/j.coviro.2013.06.010.

22. Roux S, Enault F, Bronner G, Vaulot D, Forterre P, Krupovic M. 2013. Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. Nat Commun 4:2700. https://doi.org/10.1038/ncomms3700.

23. 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. https://doi.org/10.7554/eLife.51971.

24. 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. https://doi.org/10.1038/s41564-020-0755-4.

25. 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. https://doi.org/10.1016/j.str.2011.12.014.

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

27. Kucukelbir A, Sigworth FJ, Tagare HD. 2014. Quantifying the local resolution of cryo-EM density maps. Nat Methods 11:63–65. https://doi.org/10.1038/nmeth.2727.

28. Ilca SL, Sun X, El Omari K, Kotecha A, de Haas F, DiMaio F, Grimes JM, Stuart DI, Poranen MM, Huiskonen JT. 2019. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. Nature 570:252–256. https://doi.org/10.1038/s41586-019-1229-9.

29. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2.

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

31. 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. https://doi.org/10.1038/317145a0.

32. Fisher AJ, Johnson JE. 1993. Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. Nature 361:176–179. https://doi.org/10.1038/361176a0.

33. Hogle J, Kirchhausen T, Harrison SC. 1983. Divalent cation sites in tomato bushy stunt virus. Difference maps at 2–9 A resolution. J Mol Biol 171:95–100. https://doi.org/10.1016/S0022-2836(83)80315-5.

34. Mathieu M, Petitpas I, Navaza J, Lepault J, Kohli E, Pothier P, Prasad BV, Cohen J, Rey FA. 2001. Atomic structure of the major capsid protein of rotavirus: implications for the architecture of the virion. EMBO J 20:1485–1497. https://doi.org/10.1093/emboj/20.7.1485.

35. Dokmanić I, Šikić M, Tomić S. 2008. Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination. Acta Crystallogr D Biol Crystallogr 64:257–263. https://doi.org/10.1107/S090744490706595X.

36. Johnson JE, Tang L, Johnson KN, Ball LA, Lin T, Yeager M. 2001. The structure of Pariacoto virus reveals a dodecahedral cage of duplex RNA. Nat Struct Biol 8:77–83. https://doi.org/10.1038/83089.

37. Holm L. 2020. DALI and the persistence of protein shape. Protein Sci 29:128–140. https://doi.org/10.1002/pro.3749.

38. Holm L. 2020. Using dali for protein structure comparison, p 29–42. *In* Gáspári Z (ed), Structural bioinformatics. Springer, New York, NY.

39. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. 2006. MUSTANG: A multiple structural alignment algorithm. Proteins Struct Funct Bioinforma 64:559–574. https://doi.org/10.1002/prot.20921.

40. Krupovic M. 2012. Recombination between RNA viruses and plasmids might have played a central role in the origin and evolution of small DNA viruses. Bioessays 34:867–870. https://doi.org/10.1002/bies.201200083.

41. Krupovic M, Ravantti JJ, Bamford DH. 2009. Geminiviruses: a tale of a plasmid becoming a virus. BMC Evol Biol 9:112. https://doi.org/10.1186/1471-2148-9-112.

42. Routh A, Domitrovic T, Johnson JE. 2012. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proc Natl Acad Sci U S A 109:1907–1912. https://doi.org/10.1073/pnas.1116168109.

43. Cuervo A, Daudén MI, Carrascosa JL. 2013. Nucleic acid packaging in viruses, p 361–394. *In* Mateu MG (ed), Structure and physics of viruses. Springer, Dordrecht, The Netherlands.

44. Holmes EC. 2011. What does virus evolution tell us about virus origins? J Virol 85:5247–5251. https://doi.org/10.1128/JVI.02203-10.

45. Moreira D, López-García P. 2009. Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol 7:306–311. https://doi.org/10.1038/nrmicro2108.

46. 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. https://doi.org/10.1016/j.cell.2005.01.009.

47. 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. https://doi.org/10.1016/j.str.2010.03.010.

48. 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. https://doi.org/10.1128/JVl.00368-06.

49. 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. https://doi.org/10.1016/j.virol.2003.08.045.

50. 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. https://doi.org/10.1016/j.virol.2016.08.028.

51. 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. https://doi.org/10.1038/s41586-019-1248-6.

52. 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 U S A 116:22591–22597. https://doi.org/10.1073/pnas.1908242116.

53. 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. https://doi.org/10.1038/nature20589.

54. 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 U S A 113:11519–11524. https://doi.org/10.1073/pnas.1609482113.

55. Koning RI, Gomez-Blanco J, Akopjana I, Vargas J, Kazaks A, Tars K, Carazo JM, Koster AJ. 2016. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. Nat Commun 7:12524. https://doi.org/10.1038/ncomms12524.

56. Hesketh EL, Meshcheriakova Y, Dent KC, Saxena P, Thompson RF, Cockburn JJ, Lomonossoff GP, Ranson NA. 2015. Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. Nat Commun 6:10113. https://doi.org/10.1038/ncomms10113.

57. 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 U S A 115:7344–7349. https://doi.org/10.1073/pnas.1803885115.

58. 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. https://doi.org/10.1038/s41467-019-10236-7.

59. Liu H, Cheng L. 2015. Cryo-EM shows the polymerase structures and a non-spooled genome within a dsRNA virus. Science 349:1347–1350. https://doi.org/10.1126/science.aaa4938.

60. 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. https://doi.org/10.1038/nature15767.

61. Chapman MS, Rossmann MG. 1995. Single-stranded DNA–protein interactions in canine parvovirus. Structure 3:151–162. https://doi.org/10.1016/S0969-2126(01)00146-0.

62. Hesketh EL, Saunders K, Fisher C, Potze J, Stanley J, Lomonossoff GP, Ranson NA. 2018. The 3.3 Å structure of a plant geminivirus using cryo-EM. Nat Commun 9:2369. https://doi.org/10.1038/s41467-018-04793-6.

63. Sarker S, Terrón MC, Khandokar Y, Aragão D, Hardy JM, Radjainia M, Jiménez-Zaragoza M, de Pablo PJ, Coulibaly F, Luque D, Raidal SR, Forwood JK. 2016. Structural insights into the assembly and regulation of distinct viral capsid complexes. Nat Commun 7:13014. https://doi.org/10.1038/ncomms13014.

64. Ogunbunmi ET, Roznowski AP, Fane BA. 2021. The effects of packaged, but misguided, single-stranded DNA genomes are transmitted to the outer surface of the ϕX174 capsid. J Virol 95:e0088321. https://doi.org/10.1128/JVl.00883-21.

65. Jardine P. 2008. Genome packaging in bacterial viruses, p 306–312. In Mahy BWJ, Regenmortel MHV (eds), Encyclopedia of virology. Elsevier, Amsterdam, The Netherlands.

66. McKenna R, Xia D, Willingmann P, Ilag LL, Krishnaswamy S, Rossmann MG, Olson NH, Baker TS, Incardona NL. 1992. Atomic structure of single-stranded DNA bacteriophage ΦX174 and its functional implications. Nature 355:137–143. https://doi.org/10.1038/355137a0.

67. 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. https://doi.org/10.1128/JVl.03031-13.

68. 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. https://doi.org/10.1038/nmeth.4193.

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

70. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SH. 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7:e42166. https://doi.org/10.7554/eLife.42166.

71. Sanchez-Garcia R, Gomez-Blanco J, Cuervo A, Carazo JM, Sorzano COS, Vargas J. 2021. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Commun Biol 4:874. https://doi.org/10.1038/s42003-021-02399-1.

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

73. 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. https://doi.org/10.1107/S0907444909052925.

74. 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. https://doi.org/10.1002/pro.3235.

75. Saitou N. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

76. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096.

77. Stecher G, Tamura K, Kumar S. 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol Biol Evol 37:1237–1239. https://doi.org/10.1093/molbev/msz312.

78. Charon J, Murray S, Holmes EC. 2021. Revealing RNA virus diversity and evolution in unicellular algae transcriptomes. Virus Evol 7:veab070. https://doi.org/10.1093/ve/veab070.

79. 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. Microb Environ 35:ME20016. https://doi.org/10.1264/jsme2.ME20016.

80. Zayed AA, Wainaina JM, Dominguez-Huerta G, Pelletier E, Guo J, Mohssen M, Tian F, Pratama AA, Bolduc B, Zablocki O, Cronin D, Solden L, Delage E, Alberti A, Aury J-M, Carradec Q, da Silva C, Labadie K, Poulain J, Ruscheweyh H-J, Salazar G, Shatoff E, Bundschuh R, Fredrick K, Kubatko LS, Chaffron S, Culley AI, Sunagawa S, Kuhn JH, Wincker P, Sullivan MB, Acinas SG, Babin M, Bork P, Boss E, Bowler C, Cochrane G, de Vargas C, Gorsky G, Guidi L, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Kandels S, Karp-Boss L, Karsenti E, Not F, Ogata H, Poulton N, Pesant S, et al. 2022. Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. Science 376:156–162. https://doi.org/10.1126/science.abm5847.

81. Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, Le Mercier P. 2011. ViralZone: a knowledge resource to understand virus diversity. Nucleic Acids Res 39:D576–D582. https://doi.org/10.1093/nar/gkq901.

82. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14:290–296. https://doi.org/10.1038/nmeth.4169.