Structural and molecular biology of Acheta domesticus segmented densovirus, the first parvovirus to harbor a bipartite genome

*†Judit J. Penzes[1,2], Hanh T. Pham[1], Paul Chipman[2], ¥Emmanuel W. Smith[3], Robert McKenna[2], Peter Tijssen[1]

[1]Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique, Laval, QC H7V 1B7, Canada;

[2]The McKnight Brain Institute, University of Florida, Gainesville, FL 32610, USA;

[3]Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA

*Correspondence: judycash08@gmail.com

†Current address: Institute for Quantitative Biomedicine, Rutgers, the Sate University of New Jersey, Piscataway, NJ 08854, USA

¥Current address: JEOL USA Inc., Peabody, MA 01960, USA

Abstract

Parvoviruses (family *Parvoviridae*) are defined by their linear monopartite ssDNA genome, T=1 icosahedral capsid, and distinct structural (VP) and non-structural (NS) protein expression cassettes within their genome. Here, we report the first parvovirus with a segmented genome, Acheta domesticus segmented densovirus (AdSDV), a house cricket (*Acheta domesticus*) pathogen. The AdSDV harbors its NS and VP cassettes on two separate segmented genomes. Its VP segment acquired a phospholipase A2-encoding gene via inter-subfamily recombination, which is absent from its capsid. The AdSDV evolved a transcription profile in response to its multipartite replication strategy that has diverged from its *Brevihamaparvovirus* ancestors. Furthermore, AdSDV assembles three capsid populations, which package one genome segment per particle. The cryoEM structures of these three capsids (2.3 to 3.3 Å resolution) reveal a genome packaging mechanism, which differs from other parvoviruses. This study provides a new perspective on ssDNA genome segmentation and on the plasticity of parvovirus biology.

Introduction

Parvoviruses (PVs) are small, non-enveloped icosahedral viruses, which infect vertebrate animals as well as protostome and deuterostome invertebrates [1]. Densoviruses (DVs) are autonomously replicating, invertebrate-infecting members of the *Parvoviridae* family, classified into two subfamilies [2]. Members of subfamily *Densovirinae* infect a wide array of terrestrial and aquatic invertebrates, in which they are pathogenic [3-20]. The PVs of subfamily *Hamaparvovirinae* infect either vertebrates or invertebrates, with hamaparvoviral DVs classified into three genera. Members of genera *Penstylhamaparvovirus* and *Hepanhamaparvovirus* infect penaeid shrimps, in which they are highly pathogenic [21-24]. Members of the genus *Brevihamaparvovirus* infect exclusively mosquitos and are closely related to penstylhamaparvoviruses, suggested by their genome organization, protein homology and transcription strategy [25-29].

Members of the *Parvoviridae* have linear, monopartite ssDNA genomes of 3.6 to 6.2 kb [12, 30-32], flanked by partially double-stranded, hairpin-like DNA secondary structures, which can form ~~of~~ inverted terminal repeats (ITRs) [32, 33]. The termini are essential for replication and genome packaging [32, 33]. The parvovirus genome includes two expression cassettes, one of which encodes a varied number of non-structural proteins (NS). At least one of these proteins, conventionally designated NS1, encompasses a superfamily 3 (SF3) helicase domain, which is the only highly conserved protein sequence motif throughout the entire family [32, 34]. The other expression cassette, designated *cap*, encodes one to four structural proteins (VPs). These are usually N-terminal extensions of one another, sharing an overlapping C-terminal segment, responsible for comprising the capsid shell [32, 35]. In case of subfamilies *Parvovirinae* and *Densovirinae*, the unique N-terminal extension of minor capsid protein 1 (VP1u), the largest of the VPs, typically encodes highly conserved phospholipase A2 (PLA2) domain, essential for endosomal egress [36, 37]. Following receptor-mediated endocytosis, in order to reach the nucleus to replicate, PVs have been found to traffic through the endo-lysosomal system of the host cell, exposing the viral particle to increasing acidity from pH 7.4 to 4.0 [37-42].

To date, there have been more than one hundred parvoviral capsid structures resolved at or near atomic resolution, the majority belonging to the *Parvovirinae*, as opposed to only five derived from DVs. Three of these capsid structures belong to members of the *Densovirinae*, including Galleria mellonella densovirus (GmDV) at 3.7 Å resolution (PDB ID: 1DNV) [43] of genus *Protoambidensovirus*, Acheta domesticus densovirus (AdDV) at 3.5 Å resolution (PDB ID: 4MGU) of genus *Scindoambidensovirus* [44] and Bombyx mori densovirus (BmDV) at 3.1 Å resolution (PDB ID: 3P0S) of genus *Iteradensovirus* [45]. The capsid structure of Penaeus stylirostris densovirus (PstDV) at 2.5 Å resolution (PDB ID: 3N7X) of genus *Penstylhamaparvovirus* [46] is the only high-resolution structure of the *Hamaparvovirinae* to date. The fifth structure belongs to the divergent Penaeus monodon metallodensovirus (PmMDV) at 3 Å resolution (PDB ID: 6WH3), lacking a subfamily affiliation [42]. All PV structures possess T=1 icosahedral symmetry,

71 comprised by 60 VP subunits. Each subunit displays an eight-stranded (βB to βI) jellyroll fold [47], in which
72 variable loops link the β-strands together to compose the variable capsid surface morphology [35]. With
73 the exemption of PmMDV, the luminal (BIDG) jellyroll sheet is complemented by an additional N-terminal
74 β-strand, canonically designated βA [35, 42]. The fivefold symmetry axis of the PV capsid characteristically
75 includes a pore-like opening that continues in a channel, a portal to aid genome packaging and uncoating,
76 as well as for PLA2 domain externalization [48-50].

77 The common house cricket (*Acheta domesticus*) is a host to 2 DVs of subfamily *Densovirinae*. AdDV
78 is highly pathogenic, causing mass mortality at cricket rearing facilities worldwide [8, 51]. The AdDV
79 harbors an ambisense genome, which includes a "split" VP gene, a *Scindoambidensovirus* characteristic.
80 Consequently, its VP1 is encoded by a spliced transcript, in which the PLA2-containing VP1u is expressed
81 by a separate ORF (cap1) upstream from cap2, the latter ultimately giving rise to VPs 2 to 4 via leaky
82 scanning [8]. Acheta domesticus mini ambidensovirus (AdMDV), of genus *Miniambidensovirus*, also
83 harbors an ambisense expression strategy, but only includes one cap, encoding a PLA2-encompassing VP1
84 [17].

85 Here, we report the discovery, complete genome sequence, transcription strategy and near-
86 atomic 3D structure of a novel DV infecting the common house cricket, designated Acheta domesticus
87 segmented densovirus (AdSDV). The AdSDV is the first, hitherto, PV to harbor a bipartite genome, a result
88 of recombination between subfamilies *Densovirinae* and *Hamaparvoviriane*. We show that each segment
89 is packaged into a separate viral particle, to maintain particle size and integrity of its *Brevihamaparvovirus*
90 ancestors. The AdSDV has a complex transcription strategy, which is distinct from other members of its
91 genus *Brevihamaparvovirus*, a potential adaptation to the multipartite replication strategy. Furthermore,
92 AdSDV relies on a novel DNA-packaging model, which involves the threefold and twofold axes and results
93 in increased thermostability of the full virions, by reinforcing the twofold axis via direct stacking

interactions between the lumen wall and the ssDNA genome. This study of AdSDV provides a new perspective on parvoviral genome and transcription evolution as well as on capsid architecture.

Results

Virus discovery and pathogenesis

Common house crickets at an insect rearing facility in Ontario, Canada, exhibited clinical signs consistent with a viral infection i.e., erratic, uncontrolled movement, followed by complete paralysis and eventual death. Icosahedral viral particles, ~220 Å in diameter, could be visualized in homogenized fat bodies of affected specimen by negative staining electron microscopy (Fig 1A). The presence of DVs or known CRESS DNA viruses so far known of similar size and morphology were excluded by PCR testing. Following CsCl density gradient purification, the particles were introduced per os via contaminated food to healthy house cricket nymphs. These developed identical signs in 14 to 20 days, whereas injection of the purified particles directly into the abdominal fat body accelerated the progression of the disease by about 7 days. The experiments were repeated involving two more commercially reared cricket species i.e., *Gryllus bimaculatus* and *Grylloides sigillatus*, neither of which displayed signs of infection or harbored viral particles in homogenized fat bodies. Isolated DNA could not be amplified by rolling circle amplification, which suggested a linear genome. Consequently, the extracted DNA was blunt-ended, cloned and sequenced.

Complete genome characterization and phylogeny inference of a new densovirus with a segmented genome

Sequencing identified two cloned populations, both approx. 3.3 kb in size. To verify whether both were present in the extracted viral DNA, native, blunted DNA was subjected to restriction endonuclease (RE) digestion by enzymes with recognition sites only in one (HindIII, SpeI) or in both (XbaI) obtained

sequences (Fig. 1B). The resulting restriction profiles supported the presence of a heterogenous, bipartite DNA isolate. Both segments were flanked by T-shaped hairpin-like secondary structures, which did not form ITRs but were identical at the corresponding termini of both sequences, implying a common genome origin (Fig 1C). The first segment (3316 nt in length) harbored three complete and two partial ORFs (Table S1, Fig 1D). The derived amino acid (aa) sequence of two of these, at the length of 796 and 379 aa, respectively, displayed significant similarity with the NS1 and NS2 of various mosquito-infecting brevihamaparvoviruses according to a BLASTP (basic local alignment search tool protein) search (NS1: 41% identity at 69% coverage, Haemagogus equinus DV; NS2: 41% identity at 97% coverage, Aedes albopictus DV 2), hence designating it the NS segment. The longer ORF, now referred to as NS1, harbored an SF3 helicase domain. The second segment (3332 nt) included three ORFs (Table S1, Fig 1D). ORF1 is capable of encoding a 378-aa-long protein, a homologue of a *Brevihamaparvovirus* VP from Aedes albopictus DV 2 (40% identity at 83% coverage). The putative 104-aa-long product of the small central ORF, ORF2, displayed weak similarity to bacterial SF1 DNA and RNA helicases (Beacteroides uniformis, 63% identity on 25% coverage). The 347-aa-long putative product of ORF3 was identified as a homologue of the VP1u-encoding *cap1* of AdDV (45% identity at 67% coverage). Consequently, ORF3 also harbors a PLA2 domain, similarly to the *Scindoambidensovirus* VP1u. Segment 2 was designated the VP segment. Due to the bipartite genome, the new DV was named Acheta domesticus segmented densovirus (AdSDV) and both segments were deposited to the NCBI GenBank, under accession numbers OP436269 and OP436270, respectrively.

Based on the phylogeny inference of the family-wide conserved SF3 helicase domain, AdSDV clustered to genus *Brevihamaparvovirus* of subfamily *Hamaparvovirinae*, becoming its first non-mosquito infecting member and the first one to harbor a PLA2 domain (Fig 2).

Transcription Strategy

142 To characterize the transcriptome of the bipartite genome, total RNA was isolated from infected house
143 crickets, three days following direct fat body inoculation. Both segments harbored two promoters and
144 two polyadenylation signals, respectively (Fig 1D and E). The upstream NS segment promoter at map unit
145 12 (P12) yielded two transcripts and both are polyadenylated at the proximal polyadenylation signal
146 (positioned at 2551 nt, tail is added at 2612 nt). Transcript 1 is capable of expressing the complete NS2 in
147 its entire length and did not undergo splicing. The spliced transcript 2 utilized the upstream donor site
148 (D1, ATC|CA), uniting it with the proximal acceptor site (A1, TTATCAA|AA), removing NS intron 1 (Table
149 S2, Fig S3). This results in an ORF, NS1-N1, which can express the N-terminal region of NS1, terminating
150 the frame directly upstream of the SF3 domain, by receiving a 7-aa-long tail from a small ORF without an
151 ATG start codon. There are four NS transcripts transcribed from the downstream promoter, P21,
152 terminating at the distal polyA signal (positioned at 2865 nt, polyA tail is added at 2880 nt). The unspliced
153 transcript 3 is capable of expressing an N-terminally truncated NS1 ORF, translated from the second ATG
154 of the original frame, located in a strong Kozak context (CCGCC**ATG**G). With a predicted molecular weight
155 of 74 kDa, this is the only NS segment-derived protein, which includes the complete SF3 helicase domain.
156 Either from this transcript, or from a putative transcript 4, an N-terminally truncated NS2 product (NS2-
157 C), could also be expressed, using the second ATG codon of the NS2 frame, located in a weaker Kozak
158 context (ACAC**ATG**A). The spliced transcript 5 utilized a distinct set of donor and acceptor sites from those
159 of transcript 2, namely D2 (GCA|AG) and A2 (GGGAGCA|GAG) (Table S2, Fig S3). The removal of NS-intron
160 2 puts in frame a small auxiliary ORF, ORF3, with the C-terminal portion of the NS1 ORF, potentially
161 encoding the 20-kDa-sized NS1-C. Similarly to transcript 4, the existence of transcript 6 as a separate
162 mRNA population is debatable. The removal of the same intron results in the union of the NS1 ORF with
163 a 25-aa-long tail from another ATG-lacking small ORF, giving rise to NS1-N2. Based on the transcription
164 profile, the AdSDV NS segment has a coding capacity for six putative NS proteins.

The VP segment also encoded two promoters (P12, P42) and two polyadenylation signals, resulting in the separation of the *Brevihamaparvovirus*-like ORF1 and the *Scindoambidensovirus*-like ORF3, together with auxiliary ORF2, into two separate expression units. The expression of VP ORF1, from unspliced transcript 1 is under the control of the upstream P12 and is polyadenylated at the PolyA1 site, directly upstream from the ORF3 ATG start codon (at position 1844 nt, polyA tail added at pos. 1852 nt). Transcript 2, the only unspliced transcript of the downstream P42, is potentially capable of expressing the PLA2-encoding ORF3. The spliced transcripts 3 and 4 shared the same donor site (GAA|GA) but used separate acceptor sites, A1 (TATTATA|AAC) and A2 (CAAAAAA|GAC), respectively (Fig S3). While transcript 3 unites ORF2 with an almost complete ORF3, the removal of the longer intron from transcript 4 only preserves the C-terminal portion of this ORF, removing the PLA2-encoding region.

Structural Proteins

Using the Bac-to-Bac baculovirus expression system, Sf9 cultures were transfected by three recombinant bacmid constructs, namely AdSDV-Bac-VP-ORF1 and AdSDV-Bac-VP-ORF3, with a polyhedrin promoter-linked VP ORF1 or ORF3, respectively, as well as AdSDV-VP-P42, containing the entire P42-associated expression unit in a polyhedrin promoter knock-out construct. Virus-like particle (VLP) formation was only observed in the AdSDV-Bac-VP-ORF1-transfected culture. The VLPs were purified utilizing a sucrose step gradient, with particle accumulation in the 20% fraction (Fig 3A). The particles from AdSDV-infected house crickets showing advanced signs were next purified. The gradient had two distinct particle bands in the aforementioned 20% fraction (high buoyancy – HB - capsids) as well as in the 25% fraction (low buoyancy – LB – capsids) (Fig 3A). As the true sedimentation value of these capsid populations remains undetermined, naming them only mirrors their buoyancy properties in the sucrose step gradient. When subjected to isopycnic centrifugation in a continuous CsCl gradient, two HB bands could be observed, with a buoyant density of 1.132 g/cm^3 (HB1) and 1.191 g/cm^3 (HB2), respectively, yet

maintaining a single, well-defined LB band with a buoyant density of 1.459 g/cm^3 (Fig 3B, typical for full PV particles).

Analysis by SDS-PAGE revealed a variance in the presence, absence, and incorporation ratio of protein bands at sizes of approx. 55, 50, 43, 40 and 38 kDa (Fig 3C). The bands were excised and analyzed by Nano-liquid chromatography tandem mass spectrometry (Nano-LC/MS/MS), and the protein sequences were searched against the NCBI non-redundant protein database as well as against the AdSDV genome, revealing that all bands comprised solely products of VP-ORF1 (Table S4). To furthermore investigate the absence of the PLA2-including ORF3 products, a colorimetric PLA2 assay was performed involving each capsid population and it was shown that none of these displayed PLA2 activity, in concordance with the VP-ORF3 absence suggested by the Nano-LC/MS/MS (Fig S4). The 43-kDa-sized SDS-PAGE band corresponded with the predicted weight of VP-ORF1 and was the only one with coverage throughout the complete ORF (Fig 3D). Consequently, this protein was designated VP1, the major component of the HB, HB1 and HB2 capsids and comprises about 50% of the LB population. All other bands displayed the same coverage profile, being N-terminally truncated versions of VP1. We designated the 38-kDa-sized protein VP2, representing a minority fraction of the HB, HB1 and HB2 capsids, yet accounting for half of the VPs comprising the LB particles. VP2 was also the component of the 55 and 50 kDa minor bands, the size of which exceeds the coding capacity of the AdSDV genome.

The HB and LB fractions varied significantly in genome content, yet each contained a similar ratio of NS and VP segments (Fig 3E). Cryoelectron microscopy (cryoEM) micrographs revealed that the consistently high genome count (range of 10^{16} to 10^{17} genome particles of each segment) of the LB capsids is due to the presence of almost exclusively full, genome-packaging particles (Fig 3F). This number was approx. 4-logs lower in case of the HB capsids, which displayed an overwhelmingly large proportion of empty capsids. The CsCl-purified HB1 and HB2 bands were composed of exclusively empty particles (Fig 3F).

Structural Studies

The AdSDV capsid populations and VLPs were subjected to data collection by cryoEM followed by single-single particle image reconstruction [52] (summarized by Table S6), obtaining capsid structures at a near atomic resolution for the ORF1-VLPs at 3.3 Å (PDB ID: 8EU7), for the HB (PDB ID: 8ERK), HB1 (PDB ID: 8EU6) and HB2 (PDB ID: 8EU5) capsids at 2.5 Å, 3 Å and 3.1 Å, respectively, as well as for the LB capsids at 2.3 Å (PDB ID: 8ER8) (Fig 4A). Each population possessed a T=1 icosahedral symmetry and a small particle size of 20 to 25 nm, with the smallest *Parvoviridae* lumen volume to date (Fig S7). AdSDV displays an overall smooth capsid surface with the only protruding area surrounding the fivefold symmetry axis. The lumen of the HB1 and HB2 capsids was devoid of density, apart from a small amount of dust in the proximity of the threefold axis. The same location was occupied by a larger amount of disordered density in the ORF1-VLPs. The lumen of the LB capsids was filled with density attributed to the ssDNA genome. The slight amount of genome-like density of HB capsid lumen confirmed the genome quantification and EM results, namely that this population is heterogenous and is composed of HB1, HB2 and LB capsids (detailed by Fig S8).

The VP-ORF1 sequence could be modelled into the LB density map from Thr47 to Leu377, the final C-terminal residue (Fig 4A). The AdSDV VP subunit displayed the canonical eight-stranded jellyroll fold, complemented by an N-terminal βA, located on an elongated N-terminal region. The arrangement and approximate size of each loop from AB to HI corresponded to that of PstDV the most, supported by a DALI Z-score of 16.8 [35]. The HB, HB1, HB2 and the ORF1-VLPs monomer lacked two N-terminally ordered residues as well as the unusually elongated C-terminal tail of the LB capsids, making Gln366 the last ordered residue (Fig 4C). When superimposing the monomer VP model of all capsid populations, the only region showing conformational difference was the DE loop, which comprises the fivefold channel and wall (Fig 4C). The VP capsid models built into the HB1 and HB2 density, respectively, were essentially identical (Fig S9).

Although AdSDV displays a distinct morphology within the *Parvoviridae*, the overall impression of its surface resembled that of the PstDV capsid (Fig 4D). With a root-mean square deviation (RMSD) of 2.8 Å for the Cα coordinates for 340 residues, the two structures are superimposable (Fig 4E). The AdSDV surface morphology is significantly different from that of AdDV, with only their jellyroll cores superimposable (Fig 4E).

Multimer interactions

The AdSDV fivefold channel displayed two distinct conformation states i.e., either basically empty for the HB capsids and the ORF1-VLPs, or filled with "column-like" density, as observed in the LB capsids (Fig 5A). Via the two additional ordered residues, the density column displayed clear connection to the LB capsid shell, inferring it to be part of the disordered 46-aa-long N-terminal region. While the channel of the ORF1-VLPs is covered by a hydrophobic plug, small portions of disordered density occupy the channel in the HB1 and 2 capsids. In a ~10 Å resolution map, the HB2 channel is revealed to be significantly narrower than its HB1 counterpart (Fig S9). To accommodate the density column, the peak of the LB DE loop bends away from the channel, which is occupied by large aromatic and hydrophobic sidechains, such as Tyr159 and Ile152 (Fig 5B). In contrast, both HB capsids displayed DE loops bending inwards, narrowing the pore from 16.0 Å (LB) to 10.1 Å, with Tyr159 and Ile152 retracted underneath the DE loop peak. The ORF1-VLPs possess an identical conformation to that of the LB capsids (Fig 5B).

The AdSDV capsid harbors another opening at its threefold axis, which is created by three interwoven β-strands, forming a β-annulus (Fig 5C), lined by large, hydrophobic sidechains, covering a ring of three histidines (His232). This opening is 10.8 Å wide.

The long N-terminal segment of the AdSDV capsids is located in a domain-swapping conformation at the twofold symmetry axis, with its outstretched βA interacting with the βG of its twofold-neighboring subunit, comprising the luminal surface of five-stranded β-sheets in a BIDGA order (Fig 5D).

261

262 *Capsid-genome interactions*

263 The elongated C-terminal portion of the AdSDV-LB enters the capsid lumen from the threefold symmetry
264 axis, connecting to the inner surface throughout the twofold interface. Each subunit interacts with the C-
265 terminal tail of the fivefold neighbor to their twofold neighboring subunit (Fig 6A). This interaction is
266 absent from all the other capsid populations, resulting in the last ordered C-terminal residue to enter the
267 lumen directly below the threefold symmetry axis, where the free-hanging C-terminus manifests as
268 various amounts of disordered density (Fig 6B, Fig 4A). The ordered density of six nucleotides per each LB
269 subunit occupies the luminal twofold axis. Five of these are interlinked and interact directly with the
270 capsid surface via π-stacking interactions (Fig 4C). The first stack (stack 1) also incorporates the sixth free-
271 standing purine nucleotide. The ordered ssDNA displays the sequence of purine-purine-pyrimidine-
272 purine-purine, a GC-rich motif, which is especially abundant in the AdSDV termini as well as at various
273 intervals throughout the entire sequence of both segments (85 times for the NS segment, 75 times for the
274 VP segment). The genome density displayed an increasing amount of order when it is located closer to
275 the lumen surface, as the result of the C-terminal tails “grabbing” and “sticking” the genome to the inner
276 twofold axis, by establishing π-stacks via the GC-rich pentanuclear motifs (Fig 6D).

277 Genome packaging also altered the thermostability of the AdSDV capsid, characterized by
278 differential scanning fluorometry (DSF) (Fig 6E, melt curve profiles shown by Fig S10). At neutral pH the
279 LB capsids already possessed a 2-4°C higher unfolding temperature compared to the other populations.
280 This increases to 5-6°C at pH 6.0, simulating the environment of the early endosome, and to 6-8°C at pH
281 5.5, modeling the environment of the late endosome. At the lysosomal pH of 4.0, the difference is still 3-
282 4°C. Regardless of genome content, each AdSDV capsid population displayed peak stability at pH 5.5,
283 which declined radically at pH 4.0 of the mature lysosomes.

284

Discussion

To date, all *Parvoviridae* members have been identified as containing a ssDNA monopartite genome, which is one of the key characteristics of the family [32]. However, despite of its identified bipartite genome, AdSDV is still a PV, as it displays a (i) linear ssDNA genome, flanked by hairpin-like DNA secondary structures, (ii) assembles a T=1 icosahedral, non-enveloped capsid, (iii) encodes an SF3 helicase domain in its NS1 protein, and (iv) possesses distinct NS and VP expression cassettes, which are uniquely located on two separate genome segments. Furthermore, phylogeny inference and NS homology show that AdSDV is a member of a well-established *Hamaparvovirinae* genus, *Brevihamaparvovirus*. Unlike other hamaparvoviruses, yet similarly to the other two *A. domestica*-infecting parvoviral lineages of the *Densovirinae*, AdSDV harbors a PLA2-including VP ORF of *Scindoambidensovirus* origin. Acquiring VP-ORF3 could have extended its original dipteran host spectrum to orthopterans, especially when considering the narrow host spectrum of AdSDV. While AdDV causes subclinical infection in other commonly reared orthopteran species [53], AdSDV is host-specific to *A. domesticus*.

All hitherto parvoviral PLA2 domains are either located within structural protein-encoding genes, or obtain the common C-terminal VP region via alternative splicing, so that they can be assembled into the capsid shell [8, 35, 54]. In both instances the PLA2 domain is located upstream of the major VP-encoding region. VP-ORF3 is not only placed downstream of the major VP-ORF1, but is also included in a completely separate transcription unit, with its own promoter and polyadenylation signal. Consequently, AdSDV might be a unique parvovirus to express a nonstructural PLA2. Homology search results indicate that VP-ORF2, a small auxiliary ORF of the VP segment, might have a role in the AdSDV or host cell DNA- or RNA metabolism. As this ORF is spliced to VP-ORF3, a non-structural role of VP-ORF3 expressed proteins is even more likely.

Genome segmentation is scarce among DNA viruses [55] especially in animal hosts, unlike multipartite DNA viruses of fungi and plants, which package each DNA segment into a separate viral

particle [56]. The AdSDV capsid volume is basically half of that of GmDV, the PV with the largest genome (6 kb) for which the capsid structure has been resolved [43]. As the united length of its two segments would only be slightly larger, it is reasonable to assume that AdSDV is also a multipartite-multicompartment virus. The ssRNA-packaging foot- and mouth disease virus (*Picornaviridae*), which possesses both a monopartite and multipartite variant, has been found to package shorter segments not only to maintain its particle stability, but also for the viral genome to reach an energetically favorable density [57]. Members of genera *Brevi-* and *Penstylhamaparvovirus* harbor the smallest genomes within the *Parvoviridae*, which, in case of PstDV, are packaged into similarly small particles [46]. Another arthropod-infecting linear ssDNA virus family, the *Bidnaviridae*, also harbors a bipartite, recombination-heavy genome, with genes derived from four distinct viral lineages [58]. The AdSDV genome is the first to display inter-subfamily recombination within the *Parvoviridae,* by incorporating the *Densovirinae*-originated VP-ORF3 into its ancestral hamaparvoviral genome. Acquiring distant ORFs, which may increase viral fitness by occupying new niches via extending viral host spectrum, may lead to eventual genome segmentation and a multipartite replication strategy in linear ssDNA viruses to maintain particle stability and optimal genome density.

Parvoviral transcription strategies vary significantly even within each subfamily, with vertebrate-infecting PVs and ambisense DVs of the *Densovirinae* relying heavily on alternative splicing [6, 32, 33, 59]. Brevihamaparvoviruses, however, display a simpler transcription profile, relying exclusively on leaky scanning [60]. AdSDV, in contrast, harbors a complex transcription strategy, which employs alternative splicing and could potentially express six NS proteins as opposed to the only two brevihamaparvoviral ones. Interestingly, ambisense members of the *Densovirinae* also possess a higher number of NS proteins, sometimes as many as four [4, 6-8, 30, 54]. Ambisense and multipartite genomes both may overcome the limitations of the typical parvoviral temporal monosense promoter expression order by potentially allowing the transcription machinery simultaneous access to the VP and NS expression cassettes. The large

number of AdSDV NS proteins hence might be an adaptation to the multipartite replication strategy, which suggests extreme flexibility in parvoviral expression evolution, maximizing the coding capacity of the small genome.

The surface morphology and size of the AdSDV capsid, unlike its transcription profile, remained reminiscent of PstDV. This similarity, however, is limited only to the basic fold of each subunit as the AdSDV capsid displays multimer interactions unlike other PVs. The N-termini of invertebrate-infecting PVs VPs are arranged in a domain-swapped conformation, where the βA interacts with the βB of the twofold-neighboring subunit and not with its βB, as in case of the *Parvovirinae* [43-46]. In both instances, however, the capsid lumen surface is composed of the five-stranded ABIDG sheets [35], with the exception of PmMDV, where only BIDG sheets exist [42]. The AdSDV dimer represents yet a third type of lumen architecture, which also relies on five-stranded β-sheets, but of BIDGA conformation instead. Its N-terminus still utilizes the fivefold channel comprised by its twofold-neighboring subunit for N-terminus externalization, similarly to all DVs thus far, regardless of subfamily affiliation [42-46].

The DV threefold axis is covered by the β-annulus, which creates an opening on the capsid surface of various size, similarly to the T=3 ssRNA virus family, *Tombusviridae* [61] rather than the *Parvovirinae*, where this area is covered by spikes and protrusions [35]. This opening was previously suggested to be the location of DV genome packaging, because of its size and flexibility in the GmDV capsid [43]. The AdSDV annulus is also lined by large hydrophobic sidechains as well as bulky, positively charged residues, as opposed to the abundant negative charge of the fivefold pore entrance, the canonical location of DNA entry and uncoating [32, 33, 49, 62]. Moreover, the elongated 12-aa-long portion of the C-termini are also located here, in the absence of a packaged genome. This region is not involved in comprising the AdSDV shell, yet forms stacking interactions with the genome to “pin” it to the luminal surface at the twofold axis. These characteristics might point toward a threefold axis-related genome-packaging model, instead of a fivefold-involving one.

Little is known about how the genome interacts with the PV lumen. Protoparvoviruses harbor large portions of icosahedral-ordered ssDNA, arranged in π-stacks, while adeno-associated viruses and AdDV display only a couple of ordered nucleotides, enclosed in a small pocket near the threefold axis [39, 44, 63-67]. Interactions between the lumen and these nucleotides, however, are scarce and limited to potential hydrogen bonds. The AdSDV LB capsid structure displayed the highest number of ordered nucleotides thus far, which directly interact with the luminal capsid surface. This lumen-genome interaction, which involves four subunits along with two separate regions of the ssDNA genome, provides additional stability to the otherwise weak twofold axis, observed in the absence of a packaged genome. This mechanism may be responsible for the increased thermostability of the LB capsids. Regardless of genome content, the AdSDV capsids display an identical pH-linked thermostability profile to PLA2-including members of the *Parvovirinae*, which suggested a similar endo-lysosomal trafficking pathway, even in the absence of a capsid-bound PLA2 domain [39, 48, 50, 68].

The electron density occupying the parvoviral fivefold channel has been consistently associated with N-terminal externalization, following either exposure to low endosomal pH or as the result of genome packaging [48, 63, 69]. The tightly closed channel of the HB capsids only contains traces of electron density, as opposed to the completely filled fivefold channel of the LB capsids. The AdSDV is the first parvovirus, where the “density column” harbors a direct connection to the ordered N-terminus, confirming that N-terminal externalization is an evolutionary conserved mechanism and independent from the presence of a functional PLA2.

The genome-packaging LB particles, in contrast with the HB populations, comprised a VP1-to-VP2 ratio of 1:1. It is possible that some of the externalized VP N-termini undergo proteolytic cleavage, which would explain the VP incorporation ratio shift. This mechanism, provided it exists, might have evolved to ensure that only matured virions reach the eventual replication site, as several virus families require cellular proprotein convertases to mature [70]. Unlike mature capsids of the LB population, the HB capsids

segregated into two populations of buoyant density. As the structure of the HB1 and HB2 capsids only differ in the conformation of the fivefold channel, these populations might differ in stages of particle maturation. Alternatively, the HB2 capsids might chelate ions, which could not be averaged icosahedrally, similarly to the GmDV capsid [43]. Although the ORF1-VLPs lack a packaged genome, their fivefold pore displays the open conformation of the LB particles. This suggests that the conformation changes leading up to opening the fivefold pore might happen even prior to packaging-induced N-terminus externalization, the mechanism of which remains enigmatic.

Taken together, AdSDV provides the first evidence of a multipartite replication strategy within the *Parvoviridae*, as the consequence of maintaining a *Brevihamaparvovirus*-like capsid size and morphology yet harnessing the fitness gain from a recombinant ORF from another subfamily. Adaptations imposed by the multipartite replication strategy of AdSDV may manifest in its expression profile, requiring a higher number of NS proteins, which could have led to the incorporation of alternative splicing. AdSDV also demonstrates a unique DNA-packaging mechanism, which might involve the threefold annulus for genome entry, after which the genome is attached to the luminal twofold axis by the VP C-termini. When packaging is complete, the VP N-termini undergo externalization through a possibly already open fivefold channel, which subjects them to proteolytic cleavage at some point of their life cycle. These findings alter the perspective of parvoviral traits previously deemed conservative, including a monopartite genome, a capsid-associated PLA2 domain, parvoviral capsid-DNA interactions and DV lumen architecture.

Methods

Virus Detection, Infection, DNA isolation and Cloning

Deceased common house crickets from the rearing facility were mechanically homogenized in 1x phosphate buffered saline (PBS). The homogenate was cleared up by low-speed centrifugation and PBS-diluted supernatant was applied on a glow-discharged carbon-covered Cu grid (Electron Microscopy

Sciences) and stained with 2% uranyl acetate. The dried grids were visualized at an acceleration voltage of 120 kV.

From this sample, virus particles were purified by cesium chloride gradient ultracentrifugation to obtain viral DNA for cloning. By chloroform/butanol (1:1 volume) extraction, followed by low-speed centrifugation, a clear supernatant containing viral particles was obtained. Virus stock was concentrated from the supernatant by ultracentrifugation at 40,000 rpm in a type 60Ti rotor for 2 h at 4 °C. Pellets were resuspended in small volume of PBS and was again checked by negative staining EM to verify particle presence. This virus stock was applied to dried cricket feed in a 0.1 mg/ml protein concentration. Using a 1 ml insulin syringe and a delicate needle, approx. 5 µl was injected into ~10-mm-long house crickets, targeting the fat bodies inside the abdomen but avoiding the puncture of abdominal organs.

Viral DNA was extracted by the High Pure Viral Nucleic Acid Kit (Roche) and eluted in 40 µM (µL) of distilled water. The extracted DNA was subjected to amplification by Phi29 DNA polymerase (New England Biolabs). The isolated DNA was blunt-ended utilizing T4 DNA polymerase and Large Klenow fragment of DNA polymerase I (New England Biolabs) in the presence of 33 µM of each dNTP and cloned into the EcoRV restriction site of a pBluescript KS+ vector and sequenced by primer walking. The complete sequence of both segments was cloned, but without intact termini. To obtain the sequences of the termini, single-stranded adaptors (5'- Phos – ATCCACAACAACTCTCCTCCTC – 3') were linked to the AdSDV genome segments using T4 RNA ligase I (New England Biolabs). Using the reverse adaptor primer paired with another primer, targeting the AdSDV genome in proximity of the termini, PCR amplification was performed. The 25 µL amplification reaction utilized Phusion® high-fidelity polymerase (New England Biolabs), supplemented by 10% dimethyl sulfoxide and 3 µL 2 mM EDTA. The obtained amplicons were blunt-cloned into an EcoRV-digested pBluscript KS+ plasmid and transformed into the Sure Escherichia coli strain (Stratagene), cultured at 30 °C.

To construct recombinant bacmid DNA, the VP regions to express were PCR-amplified, using Phusion® high-fidelity polymerase (New England Biolabs). The primers included the sequences of the RE sites to utilize of the pFastBac1 multiple cloning site. We obtained a polyhedrin and P10 knock-out pFastBac Dual vector by subjecting it to restriction digestion at the XhoI and BamHI sites, which flank the two promoters. The obtained pFastBac1 or Dual knock-out clones were verified by Sanger sequencing and the insert was transferred into DH10Bac competent bacteria via transposition. The recombinant baculovirus genome i.e., bacmid, was isolated from the obtained colonies and the presence of the insert was verified by PCR.

Cell Lines, Transfection, VLP Expression and Culturing Conditions

Sf9 (ATCC CRL-1711) cultures were maintained in SF900 II medium (Gibco) in a serum-free system at 28 °C. Cellfectin II Reagent (Invitrogen) was used for DNA transfection at a cell density of 8×10^5 per well. The culturing medium was aspired and replaced by seeding medium of Grace's complete insect medium supplemented with 5% FBS (Gibco) and Graces's unsupplemented insect medium, mixed at a ratio of 1:6, respectively. After adding the transfection reagent–DNA mixture to the wells, cells were incubated for 5 h. The aspired transfection medium was replaced with SF900 II medium. Cells were checked daily for signs of cytopathic effects (CPE) and the whole culture was collected when 70% of the cells detached from the dish or showed granulation. This was followed by three cycles of freeze–thaws on dry ice and 200 µL of this passage 1 (P1) stock was transferred to 25 mL of fresh Sf9 suspended cell culture in polycarbonate Erlenmeyer flasks (Corning) at the density of 2.5×10^6 cells/mL, to create the P2 stock.

Transcription Studies

Purified AdSDV-injected house crickets were collected three days post inoculation and total RNA was extracted using the Direct-zol RNA MiniPrep Kit (Zymo Research), where the denaturation step was

452 executed by adding TRIzol Reagent (Thermo Fisher Scientific). RNA was treated by digestion with the
453 TURBO DNA-free Kit (Ambion) to get rid of residual DNA contamination, as well as subjected to a control
454 PCR for the remaining DNA fragments. Reverse transcription was performed only on entirely DNA-
455 negative preparations using the SuperScript IV or the SuperScript III enzymes (Thermo Fisher Scientific),
456 supplemented with random nonamers (Sigma-Aldrich). To avoid false detection of splicing, the isolated
457 RNA was subjected to dephosphorylation by adding Antarctic phosphatase (New England Biolabs) and
458 incubated for 30 min at 37 °C. Primers were designed at the following nt positions of the AdSDV NS
459 segment: 594 (forward), 750 (reverse), 941 (reverse), 1269 (forward), 1358 (reverse), 2202 (forward and
460 reverse), 2313 (forward), 2538 (reverse). Primers targeting mRNA of the VP segment were positioned at
461 the following nt locations: 614 (reverse), 560 (forward), 1249 (forward), 1614 (forward and reverse), 1800
462 (forward and reverse), 1971 (reverse), 2611 (forward and reverse).

463 Anchored oligo(dT) primers were used with the 2202 and 2313 forward primers for the NS
464 segment, and the 1249 and 2611 forward primers in case of the VP segment for 3' RACE (rapid
465 amplification of cDNA ends). To perform 5' RACE to map transcription start sites, we designed adaptors
466 with the sequence of 5' – Phos-GCUGAUGGCGAUGAACACUGCGUUUGCUGGCUUUGAUGAAA – 3'. RNA
467 was subjected to dephosphorylation by alkaline calf intestinal phosphatase (New England Biolabs),
468 followed by phenol-chloroform extraction of the dephosphorylated RNA. We utilized tobacco acid
469 pyrophosphatase (Ambion) to remove 5' RNA caps. After the ligation of adaptors using T4 RNA ligase 1
470 (New England Biolabs), reverse transcription was executed. PCR was performed with the re-adaptor
471 primers together with oligos 750 and 941 reverse in case of the NS segment and 614, 1614 and 1971
472 reverse for the VP segment. To find out which polyadenylation signal belongs to which promoter, the
473 same re-adaptor primer was used with 20-nt-long oligos, of which 15 nucleotides corresponded with
474 those located directly upstream the polyA tail, also including a 5 nt-long polyT sequence. All PCRs were
475 performed using Phusion Hot Start Flex DNA Polymerase (New England Biolabs) in a 25 µL final reaction

476 volume, including 2 µL of purified cDNA target, 0.5 µL of both primers in 50 pmol concentration, 0.5 µL
477 dNTP mix with 8 µmol of each nucleotide, 0.75 µL of 50 mM MgCl2 solution, and 0.25 µL of enzyme. PCR
478 reactions were executed under a program of 5 min denaturation at 95°C followed by 35 cycles of 30 s
479 denaturation at 95°C, 30 s annealing at 48°C, and 1 or 2 min of elongation at 72°C. The final elongation
480 step was 8 min long at 72°C. In case of the 5'RACE reactions, 0.5 µL of enzyme was used and the number
481 of cycles was reduced to 25. For the 3'RACE, the reaction was supplemented with 1 µL of 50 mM MgCl2
482 and the annealing step was left out.

483 Total mRNA was also purified from Sf9 cell cultures transfected by the AdSDV-Bac-VP-ORF1,
484 AdSDV-Bac-VP-ORF3 and AdSDV-VP-P42 bacmid constructs, respectively. Purified mRNA was reverse-
485 transcribed and subjected to PCR amplification, so that mRNA expression of these constructs could be
486 verified, even in the absence of VLPs.

487

488 *Protein Expression and Purification of VLPs and Infectious Virus*

489 The AdSDV-Bac-VP-ORF1 P2 baculovirus stocks were incubated for at least five days and
490 monitored for CPE every day. When at least 70% of the cells showed signs of CPE, the culture was
491 collected, centrifuged at 3,000 × g, and the pelleted cells disrupted by three cycles of freeze–thaws on dry
492 ice. This lysed cell pellet was then resuspended in 1 mL of 1×TNTM pH8 (50 mM Tris pH8, 100 mM NaCl,
493 0.2% Triton X-100, 2 mM MgCl2) and centrifuged again. Supernatant was mixed back with the cell culture
494 supernatant and was subjected to treatment with 250 units of Benzonase Nuclease (Sigma-Aldrich) per
495 every 10 mL. The liquid was mixed with 1× TNET pH8 (50 mM Tris pH8, 100 mM NaCl, 0.2% Triton X-100,
496 1 mM EDTA) in a 1:1 ratio and concentrated on a cushion of 20% sucrose in TNET, using a type 60 Ti rotor
497 for 3 h at 4 °C at 45,000 rpm on a Beckman Coulter S class ultracentrifuge. The pellet was resuspended in
498 1 mL of 1×TNTM pH8 and after overnight incubation purified on a 5 to 40% sucrose step gradient for 3 h
499 at 4°C at 35,000 rpm on the same instrument in an SW 41 Ti swinging bucket preparative ultracentrifuge

rotor. The visible single band was then collected by needle puncture and a 10 mL volume syringe. For purifying infectious virus, AdSDV-inoculated crickets were mechanically homogenized in 1x PBS then subjected to the same freeze-thaw cycles and lysate clearing steps. Following Benzonase treatment the cleared lysate was subjected to the very same purification steps, detailed above. To establish the buoyant density of the AdSDV capsids, the 1xTNTM-suspended pellet of the sucrose cushion step was mixed into a 1xTNTM solution, in which CsCl was previously dissolved at a concentration of 419.5 mg/ml. The CsCl suspension was then centrifuged for 24 h in an SW 41 Ti swinging bucket preparative ultracentrifuge rotor at 10°C at 35000 rpm. The buoyant density of the obtained fractions was established using a refractometer. The aspirated fractions were dialyzed into 1× PBS at pH 7.4 to remove the sucrose or the cesium chloride.

Genome Particle Quantification

Quantification of the VP and NS segments was carried out by real-time PCR amplification (qPCR), using a Bio-Rad CFX96 instrument. A 300-bp-long target sequence was amplified of both segments (nt positions 885 to 1188 for NS and 738 to 1035 for VP). For dsDNA quantification the Bio-Rad SsoAdvanced Universal SYBR Green Supermix was used, with an amplification program of 5-min denaturation at 95 °C followed by 45 cycles of 30 s denaturation at 95°C, 15 s annealing at 55°C, and 30 s of elongation at 72°C. Results were analyzed by the CFX Maestro Software (Bio-Rad).

Differential Scanning Fluorometry (DSF) and PLA2 Assay

Capsid populations at 0.1 mg/mL concentration were dialyzed into 1x universal buffer (20 mM Hepes, 20 mM MES, 20 mM sodium acetate, 0.15 M NaCl, 3.7 mM CaCl2) at pHs 7.4, 6.0, 5.5 and 4.0. 22.5 µL capsid suspension was supplemented with 2.5 µL 1% SYPRO orange dye (Invitrogen) and subjected to DSF in a Bio-Rad CFX96 qPCR instrument. From 30°C to 99°C, the specimen was screened at a ramp rate

of 1°C/min in steps of 0.5°C. Fluorescence was measured as the function of temperature, plotted as -dRFU/dT vs. temperature, which was multiplied by -1 and normalized to the highest RFU value. Each run was performed in triplicates.

For the PLA2 assay, the same capsid concentration was used in 1x universal buffer, using VLPs of parvovirus B19 as the positive control, a generous gift from Renuk Lakshmanan (University of Florida). As heating the AdDV capsid to 65°C drastically increases PLA2 activity [44], each AdSDV capsid population was subjected to this treatment for 10 min. We performed the assay using Cayman's PLA2 Colorimetric Assay Kit. The assay was run in triplicates at 28°C for 1 hour and the absorbance was measured at 414 nm.

Protein Identification by Nano-LC/MS/MS

Excised gel bands were digested with sequencing grade trypsin (Promega) and were dehydrated with 1:1 v/v acetonitrile: 50 mM ammonium bicarbonate, followed by rehydration with dithiothreitol (DTT) solution (25 mM in 100 mM ammonium bicarbonate) and the addition of 55 mM iodoacetamide in 100 mM ammonium bicarbonate solution. The protease was driven into the gel pieces by rehydrating them in 12 ng/mL trypsin in 0.01% ProteaseMAX Surfactant for 5 minutes. The bands were then overlaid with 40 µL of 0.01% ProteaseMAX surfactant:50 mM ammonium bicarbonate and gently mixed on a shaker for 1 hour. The digestion was stopped with addition of 0.5% TFA.

Nano-LC/MS/MS was performed on a Thermo Scientific Q Exactive HF Orbitrap mass spectrometer equipped with EASY Spray nanospray source (Thermo Scientific) operated in positive ion mode. The LC system was an UltiMate™ 3000 RSLCnano system from Thermo Scientific. The mobile phase A was water containing 0.1% formic acid and the mobile phase B was acetonitrile with 0.1 % formic acid. The mobile phase A for the loading pump was water containing 0.1 % trifluoracetic acid. Five microliters of the sample was injected on to a PharmaFluidics µPAC C18 trapping column at 10 µL/ml flow rate. This was held for 3 minutes and washed with 1% B to desalt and concentrate the peptides. PharmaFluidics 50

548 cm µPAC was used for chromatographic separations with the column temperature at 40°C. A flow rate of
549 750 nl/min was used for the first 15 minutes and then the flow was reduced to 300 nl/min. Peptides were
550 eluted directly off the column into the Q Exactive system using a gradient of 1 to 20% B over 100 minutes
551 and then to 45% B in 20 minutes for a total run time of 150 minutes. The scan sequence of the mass
552 spectrometer was based on the original TopTen™ method; the analysis was programmed for a full scan
553 recorded between 375–1575 Da at 60,000 resolution, and an MS/MS scan at resolution 15,000 to
554 generate product ion spectra to determine amino acid sequence in consecutive instrument scans of the
555 fifteen most abundant peaks in the spectrum. Singly charged ions were excluded from MS2. A siloxane
556 background peak at 445.12003 was used as the internal lock mass. All MS/MS samples were analyzed
557 using Proteome Discoverer 2.4 (Thermo Fisher Scientific).
558
559 *In Silico Analyses*
560 The complete sequence of both AdSDV segments was assembled using Staden package v4.11.2
561 [71]. The assembled genome was annotated, as well as the transcripts assembled and investigated in
562 Artemis Genome Browser by the Sanger Institute [72]. To determine sequential similarity, the Blast
563 algorithms were applied [73]. To investigate conserved domains with known homologues in the derived
564 aa sequences the SMART web application was used [74]. Structural similarity of the resolved capsid
565 structures with those available in the RCSB Protein Data Bank (PDB) was investigated using the DALI server
566 [75].
567 For phylogeny inference, alignments, incorporating the outputs of pairwise, multiple, and structural
568 aligners, were constructed using the Expresso algorithm of T-Coffee, ran in PDB mode [76]. The
569 constructed alignment was edited using Unipro Ugene [77]. Model selection was executed by ProtTest
570 v2.4, suggesting the LG + I + G + F substitution model based on both the Bayesian and Akaike information
571 criteria [78]. Bayesian inference was executed by the BEAST v1.10.4 package, using a log-normal relaxed

clock with a Yule speciation prior, throughout 50,000,000 generations [79]. Convergence diagnostics were carried out using Tracer v1.7.1 of the same package, which indicated the Markov-chain Monte Carlo runs to have converged. Phylograms were edited and displayed in the FigTree 1.4.1 program of the Beast package.

Structural Studies

Three-microliter aliquots of all AdSDV capsid populations (~1 mg/mL) were applied to glow-discharged Quantifoil holey carbon grids with a thin layer of carbon (Quantifoil Micro Tools GmbH) and vitrified using a Vitrobot Mark IV (FEI) at 95% humidity and 4°C. The quality and suitability of the grids for cryo-EM data collection were determined by screening with a 16-megapixel charge-coupled device camera (Gatan) in a Tecnai G2 F20-TWIN transmission electron microscope operated at 200 kV in low-dose mode (~20 e−/Å2) prior to data collection. For collecting the low-resolution HB1 and 2 datasets, the same microscope was used at 50 frames per 10 s with a K2 Direct Electron Detector (DED) at the University of Florida Interdisciplinary Center for Biotechnology Research electron microscopy core (RRID:SCR_019146).

High-resolution data collection was carried out at two locations: the Florida State University (FSU) for the ORF1-VLPs, HB1 and HB2 populations, and the University of California, Los Angeles (UCLA) for the HB and LB capsids. In both cases a Titan Krios electron microscope (FEI) was used, operating at 300 kV, equipped with a Gatan K3 DED at FSU and Gatan K2 DED at UCLA. At UCLA, the scope also contained a Gatan postcolumn imaging filter and a free-path slit width of 20 eV. Movie frames were recorded using the Leginon semiautomated applications at both sites [80]. At FSU, the frame rate was 50 per 10 s with ~60 e−/Å2 electron dosage. At UCLA, images were collected at 50 frames per 10 s with a ~75 e−/Å2 electron dosage. All movie frames were aligned using the MotionCor2 application with dose weighting [81].

595 Single-particle image reconstruction was carried out by cisTEM v1.0. [82]. Micrograph quality was
596 assessed by CTF estimation using a box size of 512. A subset of micrographs with the best CTF fit values
597 was included in further processing. Boxing particles was performed by the particle selection subroutine,
598 at a threshold value of 2.0 to 4.0. Boxed particles were curated by 2D classification, imposing icosahedral
599 symmetry at 50 classes. Particle classes, which failed to display a clear 2D-average were eliminated from
600 the reconstruction. *Ab initio* model generation was carried out in 40 iterations under icosahedral
601 symmetry constraints. The obtained startup volume was subjected to automatic refinement with imposed
602 icosahedral symmetry and underwent iterations until reaching a stable resolution. The final maps were
603 achieved by sharpening at a post cutoff B-factor of 10 to 20. The resolution of the obtained maps was
604 calculated based on a Fourier shell correlation (FSC) of 0.143. Each map was resized to the voxel size
605 determined in Chimera (by maximizing correlation coefficient) using the e2proc3D.py subroutine in
606 EMAN2 and then converted to the CCP4 format using the program MAPMAN [83]. The atomic model of
607 the ORF1-VLPs map was built directly into the density, without an initial docked model, using Coot [84].
608 This atomic model was used to build the density of the other four reconstructions. Lastly, each model was
609 refined against the map utilizing the rigid body, real space, and B-factor refinement subroutines in Phenix
610 [85].
611

612 **Acknowledgements**

613 The authors would like to thank the late Dr. Mavis Agbandje-McKenna for her pioneering studies of
614 parvovirus capsid structures. The study was funded by the Natural Sciences and Engineering Research
615 Council of Canada Discovery grant (to P.T.) and an NIH grant R01 NIH GM082946 (to RM). The authors
616 thank Micheline Letarte at INRS IAF microscopy and the UF-ICBR Electron microscopy core for access to
617 electron microscopes utilized for negative stain electron microscopy and cryo-electron micrograph
618 screening. The Spirit and TF20 cryo-electron microscopes were provided by the UF College Of Medicine

(COM) and Division of Sponsored Programs (DSP). We thank Dr. Hong Zhou (University of California Los Angeles) and the NIH "West/Midwest Consortium for High-Resolution Cryo Electron Microscopy" project for access to the Electron Imaging Center for Nanomachines's Titan Krios and K2 DED utilized for high-resolution data collection. Data collection at Florida State University was made possible by NIH grants S10 RR025080-01 (PI Taylor), and U24 GM116788 The Southeastern Consortium for Microscopy of MacroMolecular Machines (PI Taylor). The authors also thank Kari Basso at the University of Florida Mass Spectrometry Research and Education Center for the mass spectrometry work, funded from NIH S10 OD021758-01A1. We thank for the insect rearing help provided by Martin Holm.

Author contributions

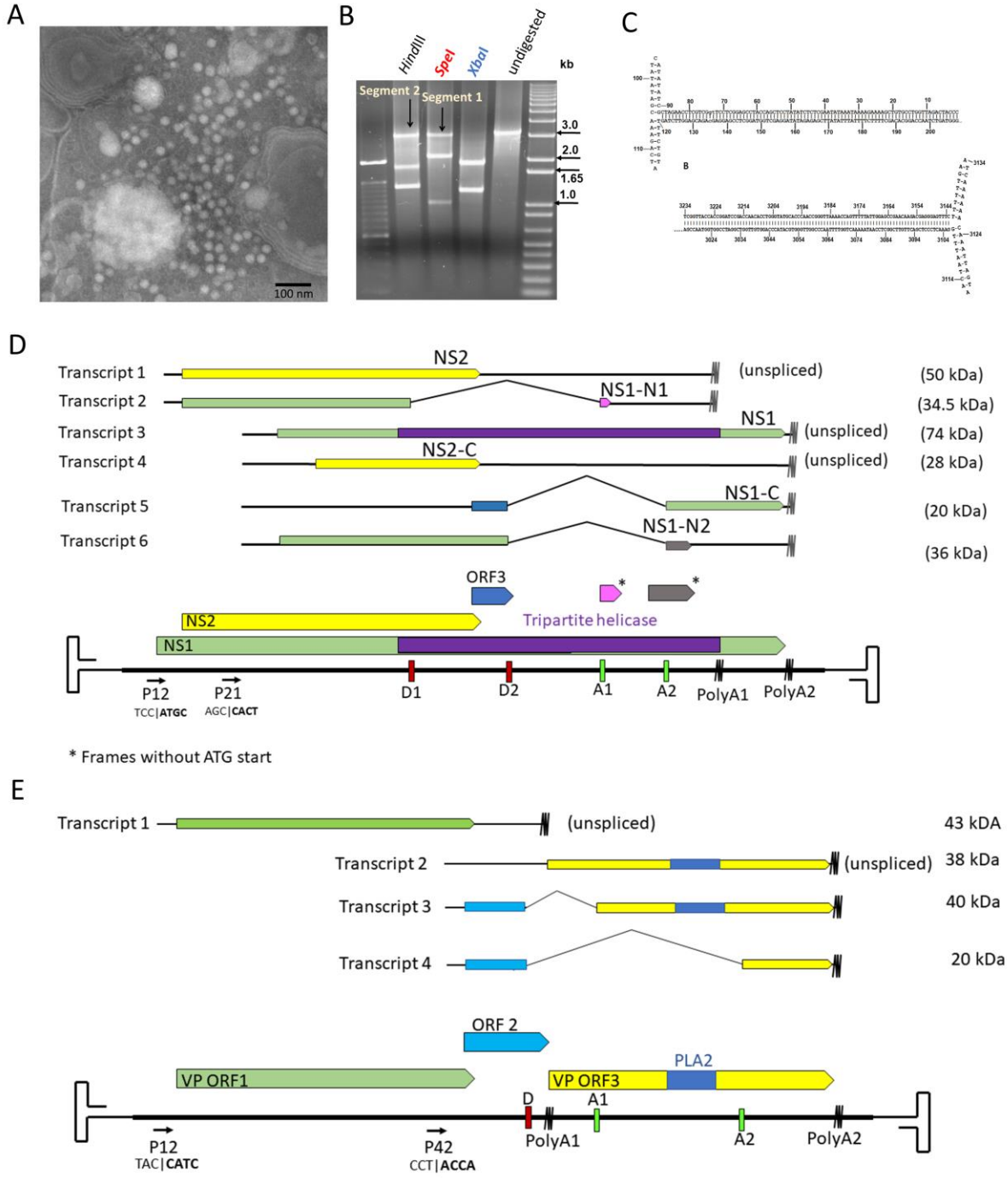
Conceived & designed by: PT, JJP, RM

Data was collected by: JJP, HTP, PC, EWS

Analysis was performed by: JJP, HTP, PT, RM

Manuscript was written by: JJP, RM, PT

Figure Captions



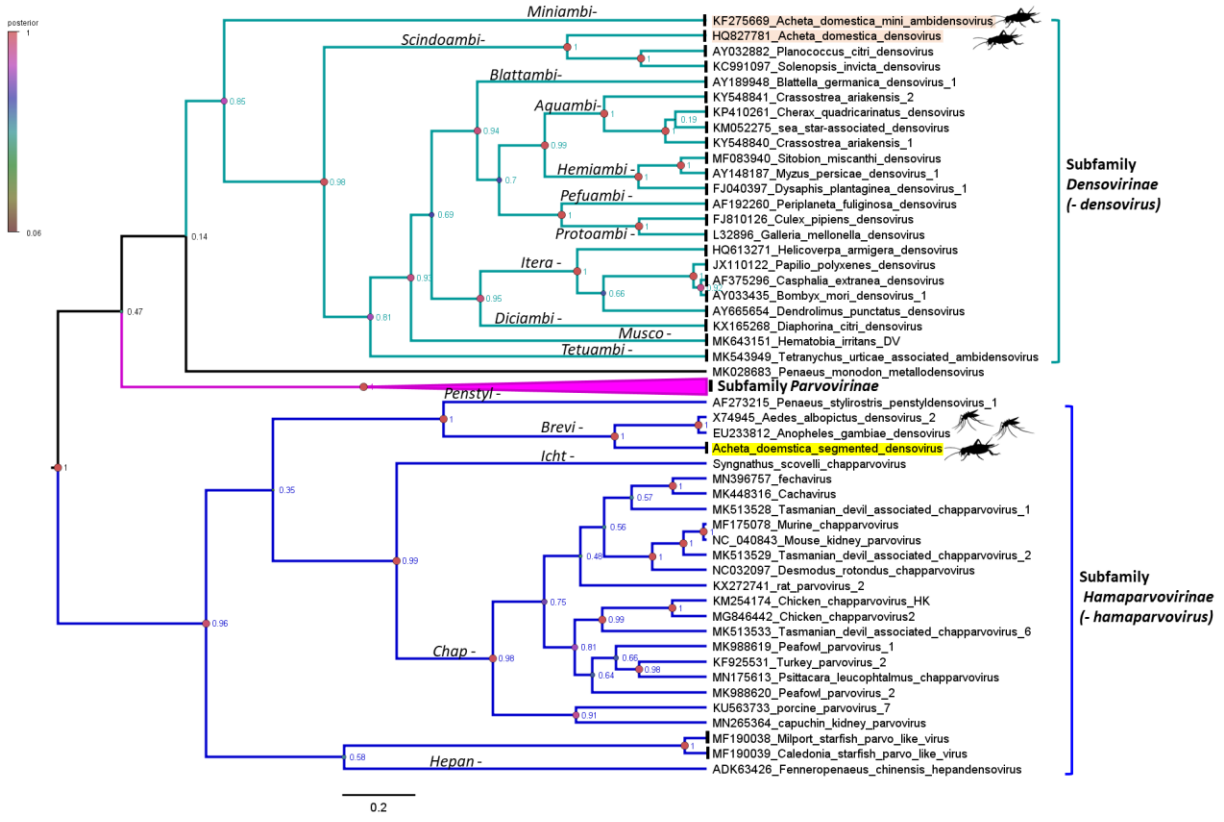
635

636 **Figure 1** Discovery, genome organization and transcription strategy of Acheta domesticus segmented

637 densovirus (AdSDV). (A) Icosahedral virus particles visualized in the homogenized fat bodies of infected

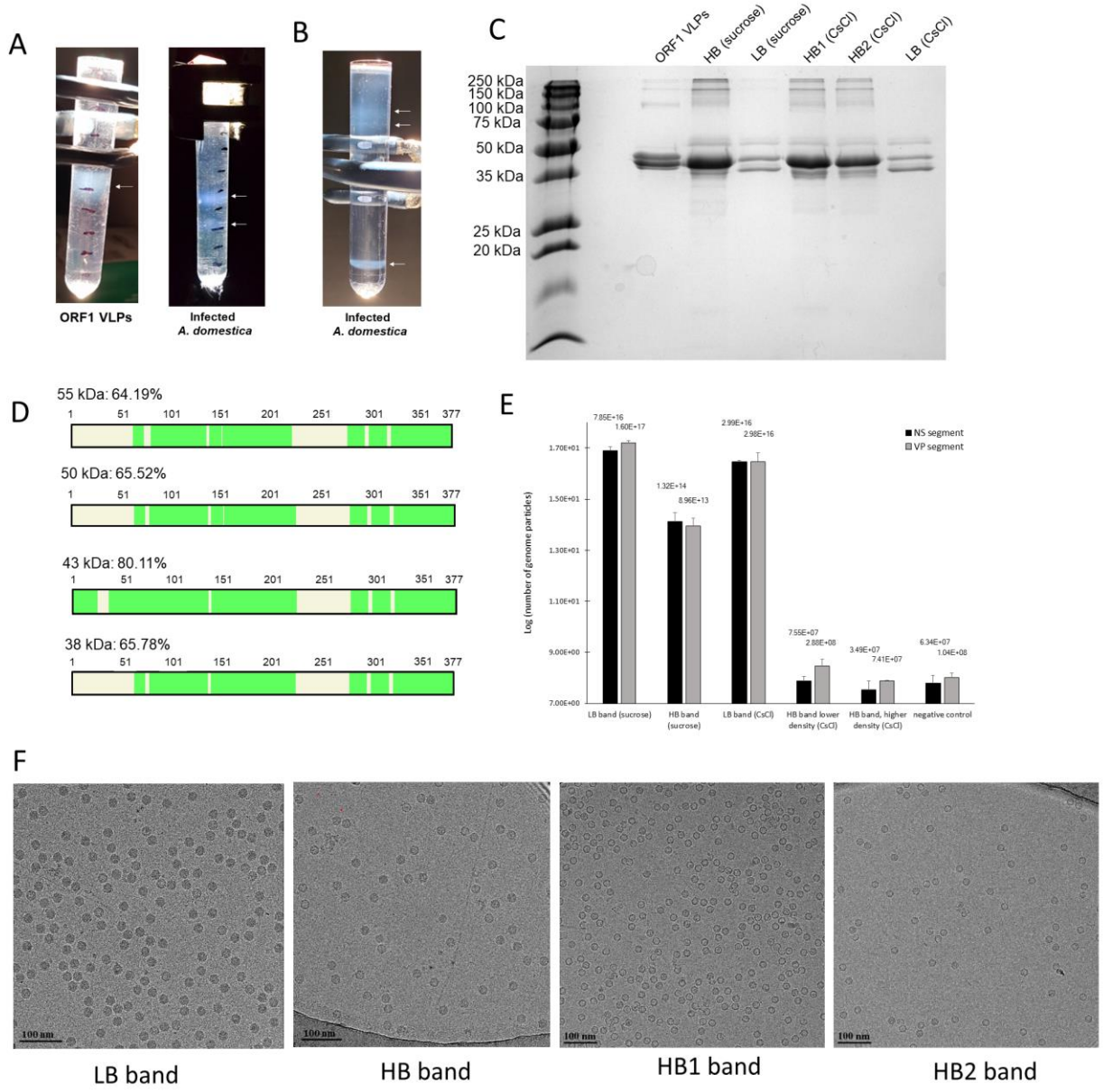
638 common house crickets (*Acheta domesticus*) by negative staining transmission electron microscopy. (B)

639 Ethidium bromide-stained agarose gel image displaying the results of digestion by restriction
640 endonucleases cutting only the NS-, VP segment, or both. (C) DNA secondary structure predictions of the
641 T-shaped genome termini, flanking both the NS and VP segment. (D) Genome organization and
642 transcription strategy of the NS segment. Open reading frames (ORFs) are marked by the colored arrows
643 and boxes, mRNA is stylized by the black lines. The wavy lines mark polyadenylation signals. Promoters
644 are labeled as "P," donor- and acceptor sites as "D" and "A" respectively. Below the promoters the
645 sequence of the transcription start site is shown, with the actual mRNA 5' end presented in bold behind
646 the vertical line. The estimated molecular weights of the protein products to be expressed by each
647 transcript are shown to the right. (E) Genome organization and transcription strategy of the VP segment,
648 using the same display and labeling scheme as in panel D.
649



650

651 **Figure 2** Bayesian phylogeny inference based on a 162-aa-long region of the NS1 protein, corresponding
652 with the superfamily 3 helicase domain, the only protein sequence conserved throughout the *Parvoviridae*
653 family. Each sequence represents one species, while the *Parvovirinae* subfamily is collapsed for
654 visualization. Genera names are shown on the branches and the posterior probability values to evaluate
655 the reliability of the topology are shown as node labels. A color coding of the node shapes is also displayed,
656 according to the posterior probability values. House cricket infecting densoviruses are highlighted in
657 apricot and labeled with the silhouette of the animal. Acheta domesticus segmented densovirus clusters
658 within the mosquito-infecting genus *Brevihamaparvovirus* and is highlighted in yellow.
659



660

661 **Figure 3** Purification and protein analyses of the Acheta domesticus segmented densovirus (AdSDV)
662 capsids. (A) Sucrose step gradients visualized under fluorescent light, following ultracentrifugation.
663 Fractions occupied by purified AdSDV virus-like particles (left) or capsids (right) are marked by the small
664 white arrows. (B) Purification of AdSDV capsids directly from the deceased house crickets by a continuous
665 CsCl gradient, which was also used to assess the buoyant density of each capsid population. The small
666 arrows mark the AdSDV capsid fractions. (C) SDS-PAGE in 15% polyacrylamide of the obtained sucrose-

667 and CsCl gradient-purified fractions. HB stands for “high-buoyancy” capsids, LB marks “low-buoyancy” (re-
668 visit?) capsids, ORF1 VLPs were expressed recombinantly from the VP-ORF1 structural protein gene. (D)
669 Coverage map of the protein sequencing reads, obtained by Nano-liquid chromatography tandem mass
670 spectrometry (Nano-LC/MS/MS) and previously excised from the polyacrylamide gel in panel C. Regions
671 that were represented among the MS reads are colored green. (E) Quantification of AdSDV viral DNA in
672 each purified capsid fraction, using real time PCR (qPCR). The histogram shows the natural logarithm of
673 the actual quantification results, with the exact unaltered values shown on top of each column. (F)
674 Electron micrographs of each AdSDV capsid fraction.
675

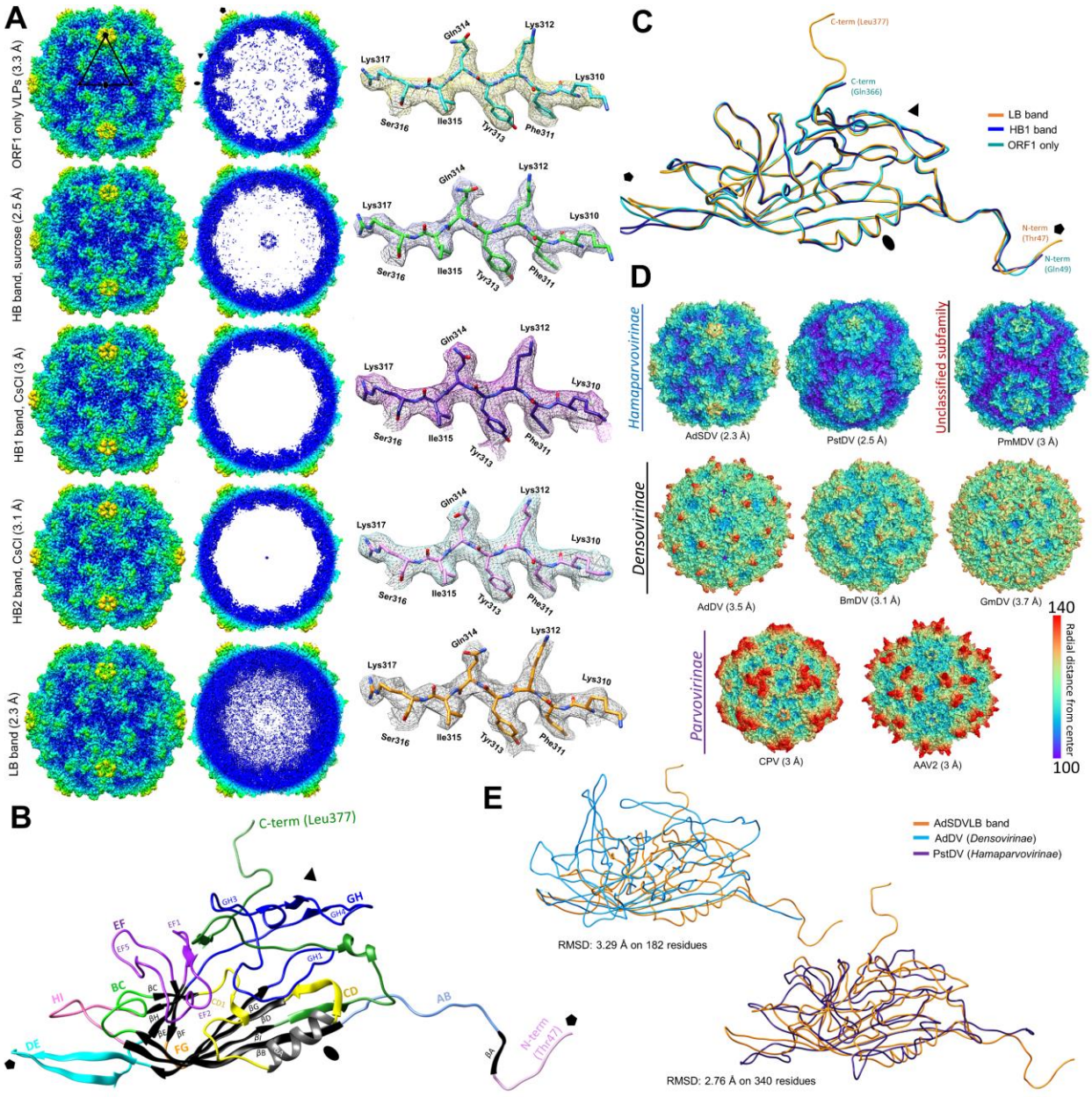


Figure 4 CryoEM single particle structural studies of the Acheta domesticus segmented densovirus (AdSDV) capsids and VP-ORF1 virus-like particles (VLPs). (A) surface (left) and cross section (right) views of the obtained AdSDV capsid and VP-ORF1 VLPs electron density maps, rendered at σ=1, with the example density and atomic model shown next to each map at σ=4. The maps are colored radially, orientated in the I1 icosahedral convention (twofold axis in z plane). The five- three- and twofold symmetry axes are marked by a pentagon, triangle and ellipse, respectively. (B) Ribbon diagram showing

683 the atomic model of a single subunit from the AdSDV low-buoyancy (LB) capsid. Structurally conserved
684 elements are shown in black and grey, the loops connecting these are highlighted in their respective
685 colors. Symmetry axes are labeled the same as in panel A, to indicate orientation. (C) Superimposition of
686 the subunit atomic model ribbon diagrams, obtained from the genome-filled infectious capsid population
687 (LB capsids), from the empty capsid fraction with the highest-buoyancy (HB1) and the VP-ORF1,
688 recombinantly expressed VLPs. (D) Capsid surface comparison of the AdSDV LB capsid with those of other
689 members of the *Parvoviridae* family i.e., invertebrate-infecting Penaeus stylirostris densovirus (PstDV),
690 Penaeus monodon metallodensovirus (PmMDV), Acheta domesticus densovirus (AdDV), Bombyx mori
691 densovirus (BmDV) and Galleria mellonella densovirus (GmDV) as well as vertebrate-infecting canine
692 parvovirus (CPV) and adeno-associated virus 2 (AAV2). Each surface model map is orientated the same
693 way as in panel A and is radially-colored. (E) Ribbon diagram superimposition of the AdSDV LB capsid
694 subunit atomic model with those of another house cricket infecting densovirus from the *Densovirinae*
695 subfamily (AdDV) as well as with PstDV, its closest relative for which the high-resolution capsid structure
696 has been resolved, of subfamily *Hamaparvovirinae*.
697

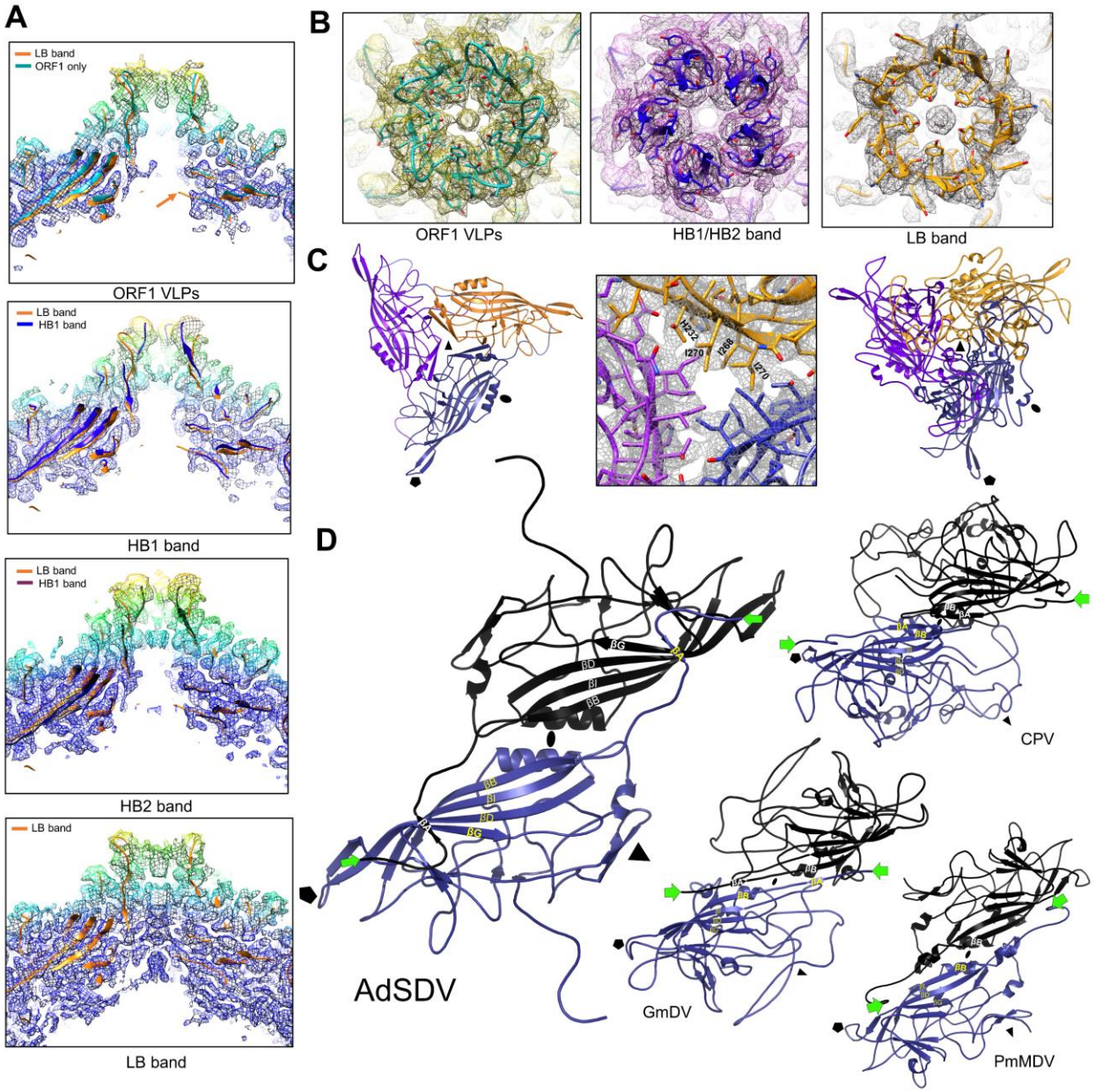
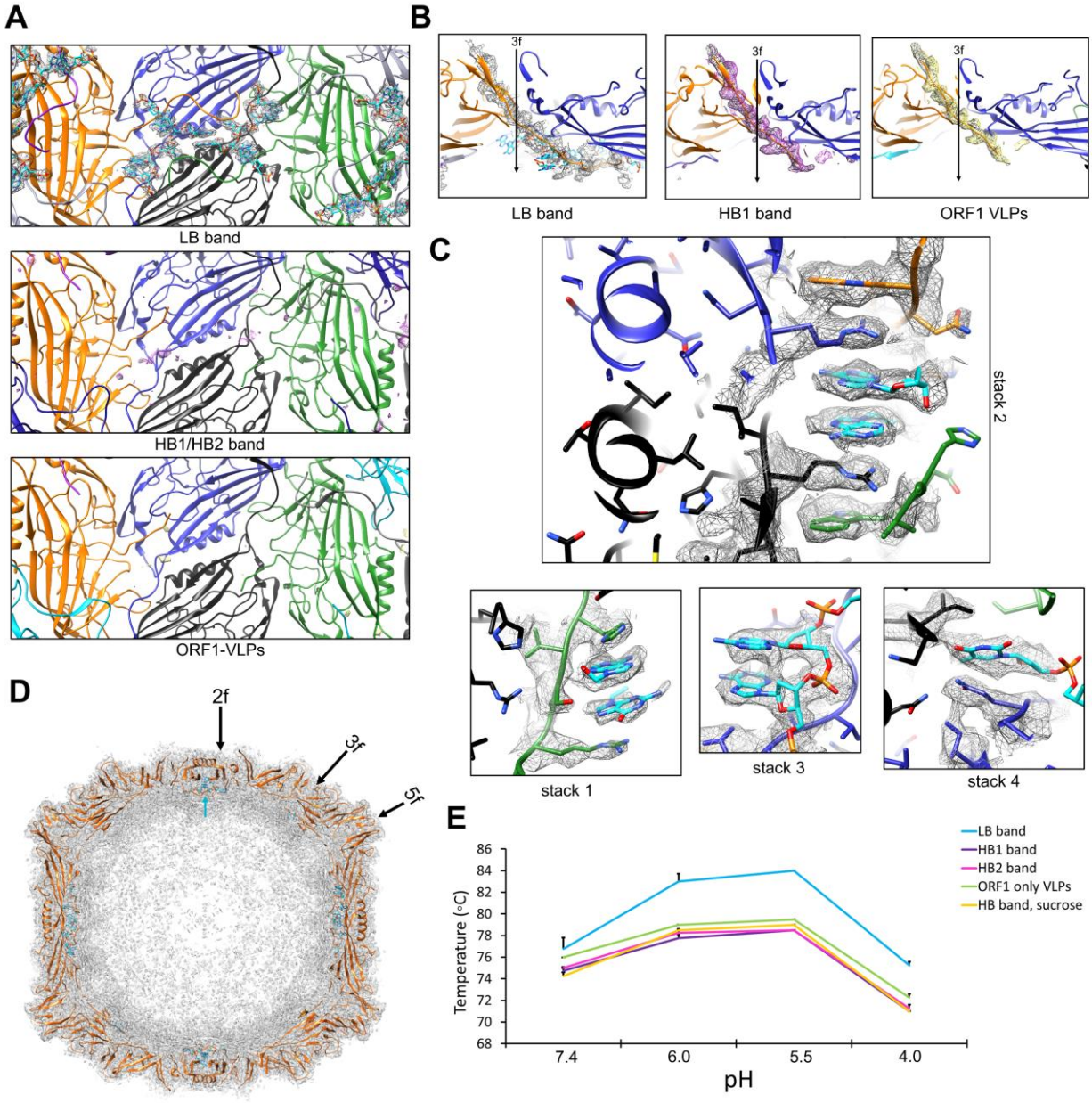


Figure 5 Multimer interactions of the Acheta domesticus segmented densovirus (AdSDV) capsid. (A) Cross-section view of the fivefold channel in case of the VP-ORF1 virus-like particles (VLPs), both high-buoyancy (HB) capsid populations and the low-buoyancy (LB) genome-filled particles. The electron density is radially-colored from the map center and is shown as a mash at σ(σ?)=2. Each map is fitted with its corresponding atomic model as ribbon diagrams as well as with the ribbon diagram of the LB capsid structure. The orange arrow indicates the LB capsid N-terminus. (B) Top-down view of the opening of the

fivefold channel, the fivefold pore, with the electron density rendered at σ=2. The atomic model fitted shows the ribbon diagram as well as the sidechains of the corresponding residues. Note the drastic conformation change of the five DE loops in opening up the channel from the closed conformation of the empty HB capsids vs. the open conformation of the genome-filled LB capsids. Note the difference between the hydrophobic plug covering the channel in case of the VP-ORF1 VLPs, as opposed to the actual N-terminus externalization observed in the LB capsids. (C) Ribbon diagrams of the AdSDV LB capsid trimer (left panel), displaying the opening of the β-annulus, typical of densoviruses. The middle panel shows the hydrophobic and positively-charged sidechains occupying the annulus, cf this threefold axis architecture with those of vertebrate-infecting parvoviruses, represented by canine parvovirus, in the right panel. (D) Ribbon diagrams of the AdSDV dimer, shown from capsid lumen, in comparison with the three other types of dimer assembly strategy, described in the *Parvoviridae* thus far. Symmetry axes are labeled by a pentagon (fivefold axis), a triangle (threefold axis) and an ellipse (twofold axis). The N-terminus of each subunit is marked by the green arrow. Note the differences between the vertebrate-infecting parvoviruses, represented here by canine parvovirus (CPV) in comparison to the domain swapping conformation of the invertebrate-infecting members of the family, represented by Galleria mellonella densovirus (GmDV) and Penaeus monodon metallodensovirus (PmMDV).



722

723 **Figure 6** Interactions between the Acheta domesticus densovirus (AdSDV) ssDNA genome and the capsid
724 lumen. (A) Luminal view of the AdSDV twofold axis, with the representation of electron density zoned to
725 the six ordered nucleic acid bases (σ=3), only present in case of the genome-filled low buoyancy (LB)
726 capsids. The ribbon diagram representation of the chain A atomic model is shown in blue and its twofold-
727 neighboring subunit in black. The fivefold-neighboring subunit of chain A is presented in green, while the
728 fivefold-neighbor of its twofold neighboring subunit in yellow. The atomic model of the nucleic acid is

729 shown in cyan. (B) Side cross section view of the β-annulus, occupying the AdSDV threefold axis. The actual
730 threefold axis is marked by an arrow. Electron density is shown zoned to the final C-terminally ordered
731 residues, indicating that the ordered portion of the LB capsid C-terminus stretches far and bends
732 underneath the twofold symmetry axis. In the absence of a packaged genome, the ordered region ends
733 directly underneath the threefold axis. (C) The AdSDV genome interacts with the luminal surface via π-
734 stacking interactions, underneath the twofold symmetry axis. Note how these stacks are the result of
735 interactions between two ssDNA regions and four subunits, colored the same as in panel A. (D) Cross
736 section of the AdSDV LB capsid map, with density shown as a mash at σ=1.5. The closer the luminal
737 genome density is located to the twofold symmetry axis, the more ordered it appears. The cyan arrow
738 points to the location of the ordered nucleotides under the twofold symmetry axis. (E) Differential
739 scanning fluorometry melting temperature profile of the AdSDV capsids and VLPs at the four pHs
740 parvoviruses may encounter during endo-lysosomal trafficking. Note the difference in thermostability
741 between the empty particles and VLPs (HB1, HB2, HB, ORF1 only VLPs) vs. the genome-containing LB
742 capsids.
743

744 **References**

745 1. Cotmore, S.F., et al., *The family Parvoviridae.* Arch Virol, 2014. **159**(5): p. 1239-47.
746 2. Pénzes, J.J., et al., *Reorganizing the family Parvoviridae: proposal for a revised taxonomy*
747 *independent from the canonical approach based on host affiliation.* Archives of Virology, 2020.
748 In Press.
749 3. Guo, H., J. Zhang, and Y. Hu, *Complete sequence and organization of Periplaneta fuliginosa*
750 *densovirus genome.* Acta Virol, 2000. **44**(6): p. 315-22.
751 4. Thao, M.L., et al., *Genetic characterization of a putative Densovirus from the mealybug*
752 *Planococcus citri.* Curr Microbiol, 2001. **43**(6): p. 457-8.

5. Fediere, G., et al., *Genome organization of Casphalia extranea densovirus, a new iteravirus.* Virology, 2002. **292**(2): p. 299-308.

6. Baquerizo-Audiot, E., et al., *Structure and expression strategy of the genome of Culex pipiens densovirus, a mosquito densovirus with an ambisense organization.* J Virol, 2009. **83**(13): p. 6863-73.

7. Kapelinskaya, T.V., et al., *Expression strategy of densonucleosis virus from the German cockroach, Blattella germanica.* J Virol, 2011. **85**(22): p. 11855-70.

8. Liu, K., et al., *The Acheta domesticus densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses.* J Virol, 2011. **85**(19): p. 10069-78.

9. Gudenkauf, B.M., et al., *Discovery of urchin-associated densoviruses (family Parvoviridae) in coastal waters of the Big Island, Hawaii.* J Gen Virol, 2014. **95**(Pt 3): p. 652-8.

10. Bucci, C., et al., *Sea Star Wasting Disease in Asterias forbesi along the Atlantic Coast of North America.* PLoS One, 2017. **12**(12): p. e0188523.

11. Kang, Y.J., et al., *Densoviruses in oyster Crassostrea ariakensis.* Arch Virol, 2017. **162**(7): p. 2153-2157.

12. Francois, S., et al., *A New Prevalent Densovirus Discovered in Acari. Insight from Metagenomics in Viral Communities Associated with Two-Spotted Mite (Tetranychus urticae) Populations.* Viruses, 2019. **11**(3).

13. Ribeiro, J.M., et al., *An insight into the sialome, mialome and virome of the horn fly, Haematobia irritans.* BMC Genomics, 2019. **20**(1): p. 616.

14. Richard, J.C., et al., *Mass mortality in freshwater mussels (Actinonaias pectorosa) in the Clinch River, USA, linked to a novel densovirus.* Sci Rep, 2020. **10**(1): p. 14498.

15. Jackson, E.W., et al., *A Highly Prevalent and Pervasive Densovirus Discovered among Sea Stars from the North American Atlantic Coast.* Appl Environ Microbiol, 2020. **86**(6).

16. Li, Y., et al., *Genome organization of the densovirus from Bombyx mori (BmDNV-1) and enzyme activity of its capsid.* J Gen Virol, 2001. **82**(Pt 11): p. 2821-5.

17. Pham, H.T., et al., *A Novel Ambisense Densovirus, Acheta domesticus Mini Ambidensovirus, from Crickets.* Genome Announc, 2013. **1**(6).

18. Nigg, J.C. and B.W. Falk, *Diaphorina citri densovirus is a persistently infecting virus with a hybrid genome organization and unique transcription strategy.* J Gen Virol, 2020. **101**(2): p. 226-239.

19. Bochow, S., et al., *First complete genome of an Ambidensovirus; Cherax quadricarinatus densovirus, from freshwater crayfish Cherax quadricarinatus.* Mar Genomics, 2015. **24 Pt 3**: p. 305-12.

20. Hewson, I., et al., *Densovirus associated with sea-star wasting disease and mass mortality.* Proc Natl Acad Sci U S A, 2014. **111**(48): p. 17278-83.

21. Bonami, J.R., et al., *Characterization of hepatopancreatic parvo-like virus, a second unusual parvovirus pathogenic for penaeid shrimps.* J Gen Virol, 1995. **76 (Pt 4)**: p. 813-7.

22. Tang, K.F., et al., *Geographic variations among infectious hypodermal and hematopoietic necrosis virus (IHHNV) isolates and characteristics of their infection.* Dis Aquat Organ, 2003. **53**(2): p. 91-9.

23. Tang, K.F., C.R. Pantoja, and D.V. Lightner, *Nucleotide sequence of a Madagascar hepatopancreatic parvovirus (HPV) and comparison of genetic variation among geographic isolates.* Dis Aquat Organ, 2008. **80**(2): p. 105-12.

24. Lightner, D.V., et al., *Historic emergence, impact and current status of shrimp pathogens in the Americas.* J Invertebr Pathol, 2012. **110**(2): p. 174-83.

25. Shike, H., et al., *Infectious hypodermal and hematopoietic necrosis virus of shrimp is related to mosquito brevidensoviruses.* Virology, 2000. **277**(1): p. 167-77.

26. Roekring, S., et al., *Comparison of penaeid shrimp and insect parvoviruses suggests that viral transfers may occur between two distantly related arthropod groups.* Virus Res, 2002. **87**(1): p. 79-87.

27. Sivaram, A., et al., *Isolation and characterization of densonucleosis virus from Aedes aegypti mosquitoes and its distribution in India.* Intervirology, 2009. **52**(1): p. 1-7.

28. Chen, S., et al., *Genetic, biochemical, and structural characterization of a new densovirus isolated from a chronically infected Aedes albopictus C6/36 cell line.* Virology, 2004. **318**(1): p. 123-33.

29. Li, J., et al., *A Novel Densovirus Isolated From the Asian Tiger Mosquito Displays Varied Pathogenicity Depending on Its Host Species.* Front Microbiol, 2019. **10**: p. 1549.

30. Tijssen, P., et al., *Organization and expression strategy of the ambisense genome of densonucleosis virus of Galleria mellonella.* J Virol, 2003. **77**(19): p. 10357-65.

31. Sukhumsirichart, W., et al., *Complete nucleotide sequence and genomic organization of hepatopancreatic parvovirus (HPV) of Penaeus monodon.* Virology, 2006. **346**(2): p. 266-77.

32. Cotmore, S.F., et al., *ICTV Virus Taxonomy Profile: Parvoviridae.* J Gen Virol, 2019. **100**(3): p. 367-368.

33. Cotmore, S.F. and P. Tattersall, *Parvoviruses: Small Does Not Mean Simple.* Annu Rev Virol, 2014. **1**(1): p. 517-37.

34. Pénzes, J.J., et al., *Reorganizing the family Parvoviridae: a revised taxonomy independent of the canonical approach based on host association.* Arch Virol, 2020. **165**(9): p. 2133-2146.

35. Mietzsch, M., J.J. Penzes, and M. Agbandje-McKenna, *Twenty-Five Years of Structural Parvovirology.* Viruses, 2019. **11**(4).

36. Zadori, Z., et al., *A viral phospholipase A2 is required for parvovirus infectivity.* Dev Cell, 2001. **1**(2): p. 291-302.

824 37. Farr, G.A., L.G. Zhang, and P. Tattersall, *Parvoviral virions deploy a capsid-tethered lipolytic*
825 *enzyme to breach the endosomal membrane during cell entry.* Proc Natl Acad Sci U S A, 2005.
826 **102**(47): p. 17148-53.
827 38. Bartlett, J.S., R. Wilcher, and R.J. Samulski, *Infectious entry pathway of adeno-associated virus*
828 *and adeno-associated virus vectors.* J Virol, 2000. **74**(6): p. 2777-85.
829 39. Nam, H.J., et al., *Structural studies of adeno-associated virus serotype 8 capsid transitions*
830 *associated with endosomal trafficking.* J Virol, 2011. **85**(22): p. 11791-9.
831 40. Mani, B., et al., *Low pH-dependent endosomal processing of the incoming parvovirus minute*
832 *virus of mice virion leads to externalization of the VP1 N-terminal sequence (N-VP1), N-VP2*
833 *cleavage, and uncoating of the full-length genome.* J Virol, 2006. **80**(2): p. 1015-24.
834 41. Vendeville, A., et al., *Densovirus infectious pathway requires clathrin-mediated endocytosis*
835 *followed by trafficking to the nucleus.* J Virol, 2009. **83**(9): p. 4678-89.
836 42. Pénzes, J.J., et al., *Molecular biology and structure of a novel penaeid shrimp densovirus*
837 *elucidate convergent parvoviral host capsid evolution.* Proc Natl Acad Sci U S A, 2020. **117**(33): p.
838 20211-20222.
839 43. Simpson, A.A., et al., *The structure of an insect parvovirus (Galleria mellonella densovirus) at 3.7*
840 *A resolution.* Structure, 1998. **6**(11): p. 1355-67.
841 44. Meng, G., et al., *The structure and host entry of an invertebrate parvovirus.* J Virol, 2013. **87**(23):
842 p. 12523-30.
843 45. Kaufmann, B., et al., *Structure of Bombyx mori densovirus 1, a silkworm pathogen.* J Virol, 2011.
844 **85**(10): p. 4691-7.
845 46. Kaufmann, B., et al., *Structure of Penaeus stylirostris densovirus, a shrimp pathogen.* J Virol,
846 2010. **84**(21): p. 11289-96.

47. Rossmann, M.G., et al., *Structural comparisons of some small spherical plant viruses.* J Mol Biol, 1983. **165**(4): p. 711-36.

48. Penzes, J.J., et al., *Adeno-associated Virus 9 Structural Rearrangements Induced by Endosomal Trafficking pH and Glycan Attachment.* J Virol, 2021. **95**(19): p. e0084321.

49. Plevka, P., et al., *Structure of a packaging-defective mutant of minute virus of mice indicates that the genome is packaged via a pore at a 5-fold axis.* J Virol, 2011. **85**(10): p. 4822-7.

50. Venkatakrishnan, B., et al., *Structure and dynamics of adeno-associated virus serotype 1 VP1-unique N-terminal domain and its role in capsid trafficking.* J Virol, 2013. **87**(9): p. 4974-84.

51. Pham, H.T., et al., *Comparative Genomic Analysis of Acheta domesticus Densovirus Isolates from Different Outbreaks in Europe, North America, and Japan.* Genome Announc, 2013. **1**(4).

52. Guo, F. and W. Jiang, *Single particle cryo-electron microscopy and 3-D reconstruction of viruses.* Methods Mol Biol, 2014. **1117**: p. 401-43.

53. Duffield, K.R., et al., *Active and Covert Infections of Cricket Iridovirus and Acheta domesticus Densovirus in Reared Gryllodes sigillatus Crickets.* Front Microbiol, 2021. **12**: p. 780796.

54. Yamagishi, J., et al., *Genome organization and mRNA structure of Periplaneta fuliginosa densovirus imply alternative splicing involvement in viral gene expression.* Arch Virol, 1999. **144**(11): p. 2111-24.

55. Iranzo, J. and S.C. Manrubia, *Evolutionary dynamics of genome segmentation in multipartite viruses.* Proc Biol Sci, 2012. **279**(1743): p. 3812-9.

56. Sicard, A., et al., *The Strange Lifestyle of Multipartite Viruses.* PLoS Pathog, 2016. **12**(11): p. e1005819.

57. Ojosnegros, S., et al., *Viral genome segmentation can result from a trade-off between genetic content and particle stability.* PLoS Genet, 2011. **7**(3): p. e1001344.

58. Krupovic, M. and E.V. Koonin, *Evolution of eukaryotic single-stranded DNA viruses of the Bidnaviridae family from genes of four other groups of widely different viruses.* Sci Rep, 2014. **4**: p. 5347.

59. Tijssen, P., et al., *Diversity of small, single-stranded DNA viruses of invertebrates and their chaotic evolutionary past.* J Invertebr Pathol, 2016. **140**: p. 83-96.

60. Pham, H.T., et al., *Expression strategy of Aedes albopictus densovirus.* J Virol, 2013. **87**(17): p. 9928-32.

61. Sherman, M.B., et al., *Near-Atomic-Resolution Cryo-Electron Microscopy Structures of Cucumber Leaf Spot Virus and Red Clover Necrotic Mosaic Virus: Evolutionary Divergence at the Icosahedral Three-Fold Axes.* J Virol, 2020. **94**(2).

62. Subramanian, S., et al., *Cryo-EM maps reveal five-fold channel structures and their modification by gatekeeper mutations in the parvovirus minute virus of mice (MVM) capsid.* Virology, 2017. **510**: p. 216-223.

63. Mietzsch, M., et al., *Comparative Analysis of the Capsid Structures of AAVrh.10, AAVrh.39, and AAV8.* J Virol, 2020. **94**(6).

64. Halder, S., et al., *Structure of neurotropic adeno-associated virus AAVrh.8.* J Struct Biol, 2015. **192**(1): p. 21-36.

65. Halder, S., et al., *Structural characterization of H-1 parvovirus: comparison of infectious virions to empty capsids.* J Virol, 2013. **87**(9): p. 5128-40.

66. Llamas-Saiz, A.L., et al., *Structure determination of minute virus of mice.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 1): p. 93-102.

67. Tsao, J., et al., *The three-dimensional structure of canine parvovirus and its functional implications.* Science, 1991. **251**(5000): p. 1456-64.

68. Afione, S., et al., *Identification and mutagenesis of the adeno-associated virus 5 sialic acid binding region.* J Virol, 2015. **89**(3): p. 1660-72.

69. Kaufmann, B., et al., *Visualization of the externalized VP2 N termini of infectious human parvovirus B19.* J Virol, 2008. **82**(15): p. 7306-12.

70. Izaguirre, G., *The Proteolytic Regulation of Virus Cell Entry by Furin and Other Proprotein Convertases.* Viruses, 2019. **11**(9).

71. Staden, R., K.F. Beal, and J.K. Bonfield, *The Staden package, 1998.* Methods Mol Biol, 2000. **132**: p. 115-30.

72. Carver, T., et al., *Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data.* Bioinformatics, 2012. **28**(4): p. 464-9.

73. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.

74. Letunic, I. and P. Bork, *20 years of the SMART protein domain annotation resource.* Nucleic Acids Res, 2018. **46**(D1): p. D493-d496.

75. Holm, L., *DALI and the persistence of protein shape.* Protein Sci, 2019.

76. Armougom, F., et al., *Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W604-8.

77. Okonechnikov, K., O. Golosova, and M. Fursov, *Unipro UGENE: a unified bioinformatics toolkit.* Bioinformatics, 2012. **28**(8): p. 1166-7.

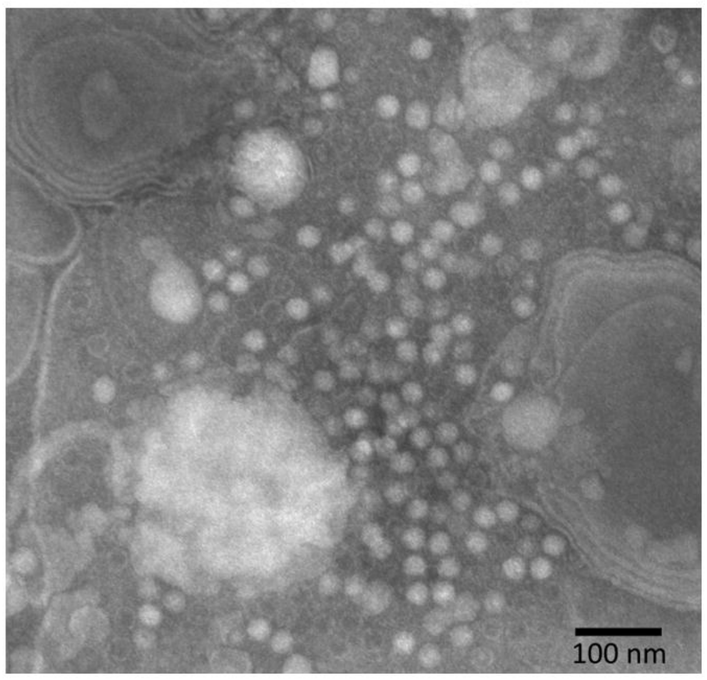
78. Darriba, D., et al., *ProtTest 3: fast selection of best-fit models of protein evolution.* Bioinformatics, 2011. **27**(8): p. 1164-5.

79. Suchard, M.A., et al., *Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10.* Virus Evol, 2018. **4**(1): p. vey016.

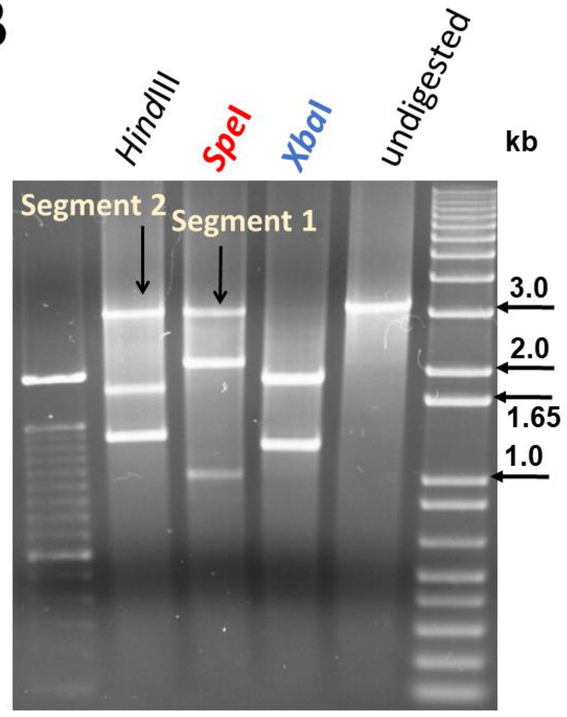
916 80. Suloway, C., et al., *Automated molecular microscopy: the new Leginon system.* J Struct Biol,
917 2005. **151**(1): p. 41-60.
918 81. Zheng, S.Q., et al., *MotionCor2: anisotropic correction of beam-induced motion for improved*
919 *cryo-electron microscopy.* Nat Methods, 2017. **14**(4): p. 331-332.
920 82. Grant, T., A. Rohou, and N. Grigorieff, *cisTEM, user-friendly software for single-particle image*
921 *processing.* Elife, 2018. **7**.
922 83. Kleywegt, G.J. and T.A. Jones, *xdlMAPMAN and xdlDATAMAN - programs for reformatting,*
923 *analysis and manipulation of biomacromolecular electron-density maps and reflection data sets.*
924 Acta Crystallogr D Biol Crystallogr, 1996. **52**(Pt 4): p. 826-8.
925 84. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010.
926 **66**(Pt 4): p. 486-501.
927 85. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure*
928 *solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

929

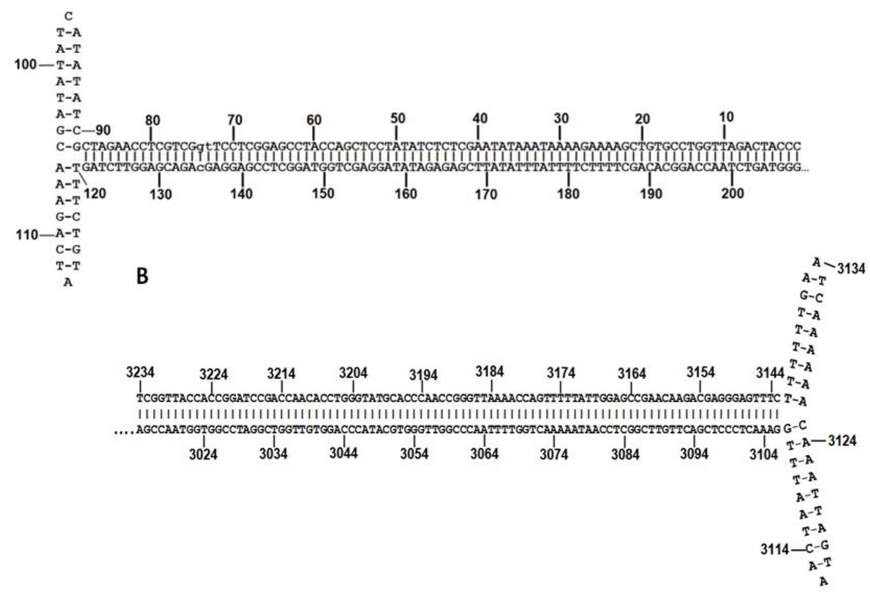
A



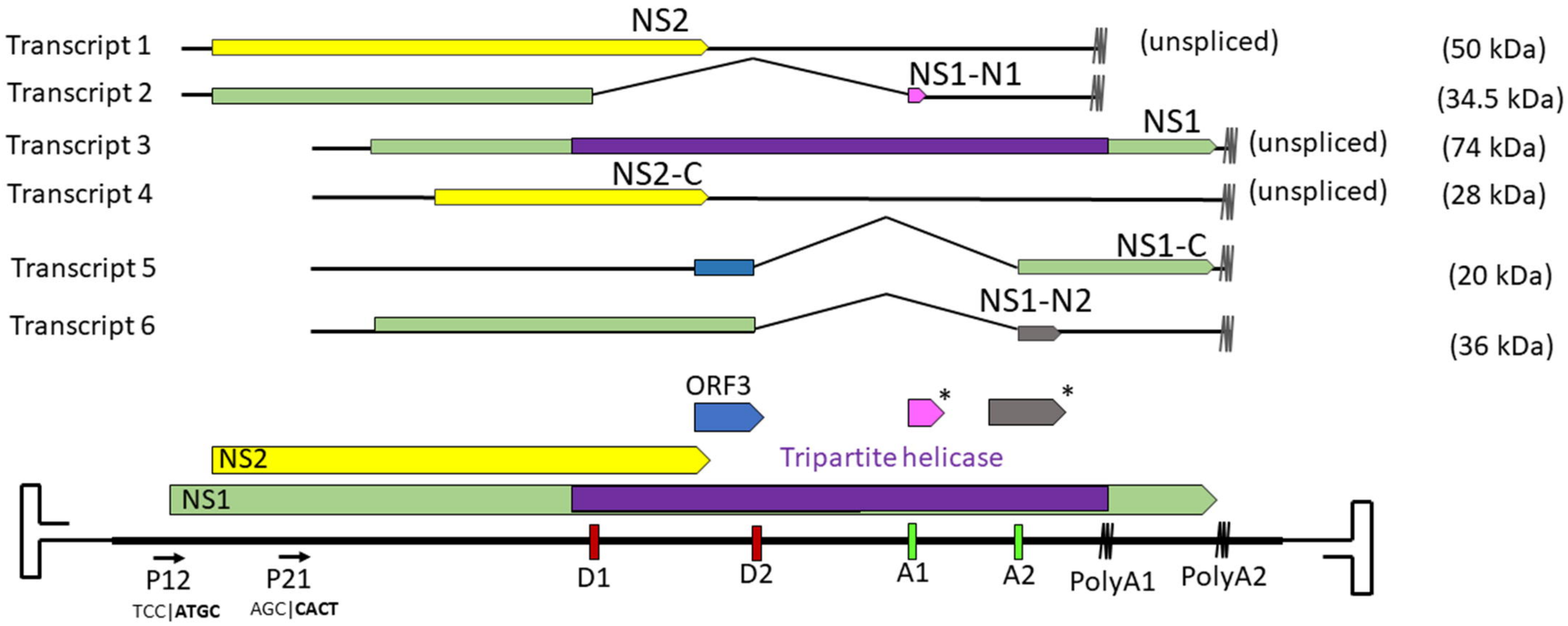
B



C

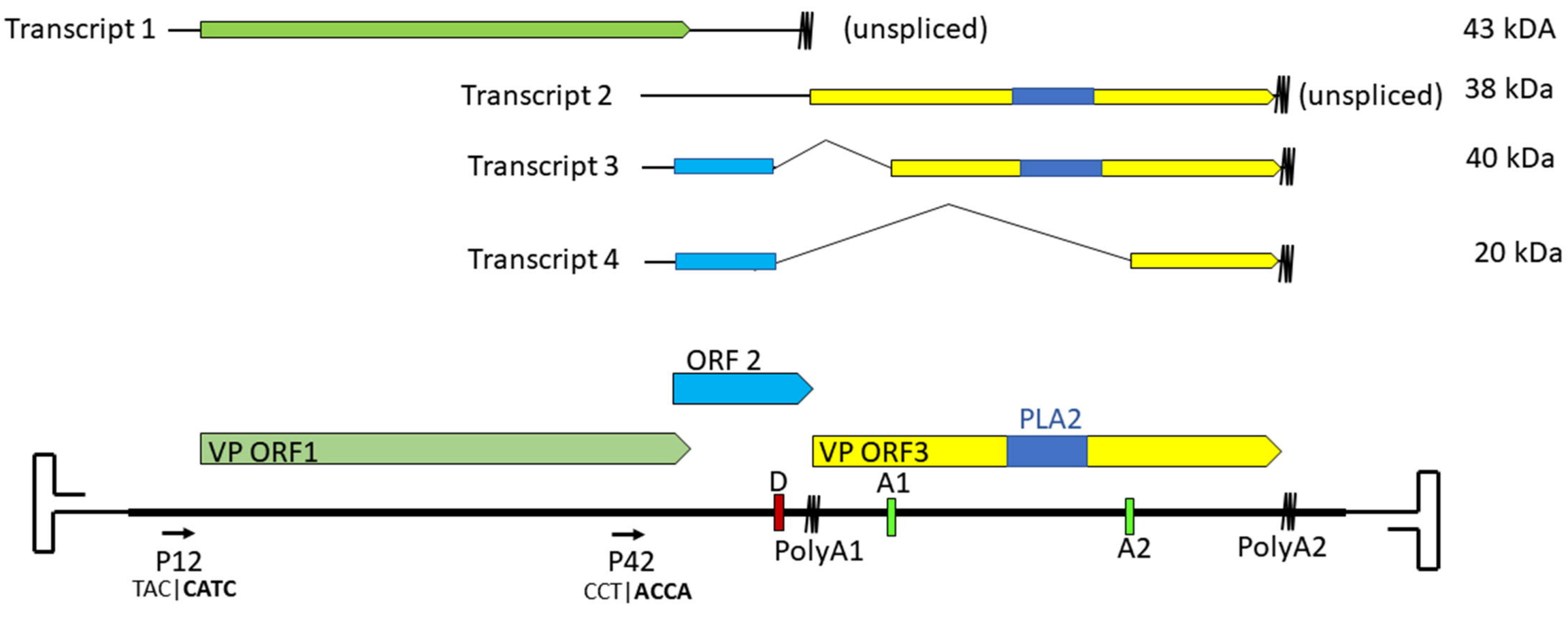


D



* Frames without ATG start

E



posterior
1
0.06

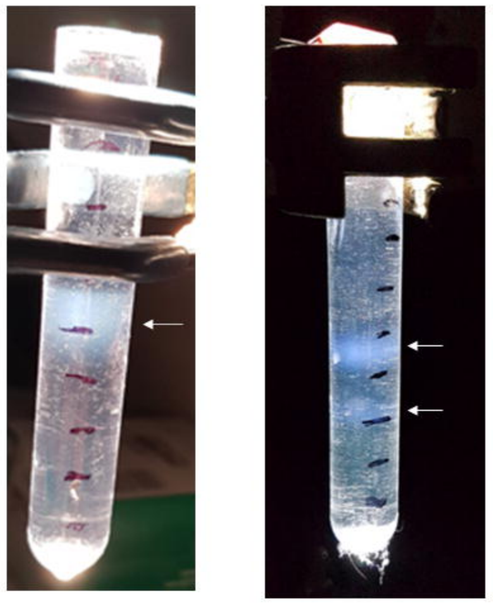
Miniambi-
Scindoambi-
Blattambi-
Aquambi-
Hemiambi -
Pefuambi -
Protoambi -
Itera -
Diciambi -
Musco -
Tetuambi -
Penstyl -
Brevi -
Icht -
Chap -
Hepan -

KF275669_Acheta_domestica_mini_ambidensovirus
HQ827781_Acheta_domestica_densovirus
AY032882_Planococcus_citri_densovirus
KC991097_Solenopsis_invicta_densovirus
AY189948_Blattella_germanica_densovirus_1
KY548841_Crassostrea_ariakensis_2
KP410261_Cherax_quadricarinatus_densovirus
KM052275_sea_star-associated_densovirus
KY548840_Crassostrea_ariakensis_1
MF083940_Sitobion_miscanthi_densovirus
AY148187_Myzus_persicae_densovirus_1
FJ040397_Dysaphis_plantaginea_densovirus_1
AF192260_Periplaneta_fuliginosa_densovirus
FJ810126_Culex_pipiens_densovirus
L32896_Galleria_mellonella_densovirus
HQ613271_Helicoverpa_armigera_densovirus
JX110122_Papilio_polyxenes_densovirus
AF375296_Casphalia_extranea_densovirus
AY033435_Bombyx_mori_densovirus_1
AY665654_Dendrolimus_punctatus_densovirus
KX165268_Diaphorina_citri_densovirus
MK643151_Hematobia_irritans_DV
MK543949_Tetranychus_urticae_associated_ambidensovirus
MK028683_Penaeus_monodon_metallodensovirus
AF273215_Penaeus_stylirostris_penstyldensovirus_1
X74945_Aedes_albopictus_densovirus_2
EU233812_Anopheles_gambiae_densovirus
Acheta_doemstica_segmented_densovirus
Syngnathus_scovelli_chapparvovirus
MN396757_fechavirus
MK448316_Cachavirus
MK513528_Tasmanian_devil_associated_chapparvovirus_1
MF175078_Murine_chapparvovirus
NC_040843_Mouse_kidney_parvovirus
MK513529_Tasmanian_devil_associated_chapparvovirus_2
NC032097_Desmodus_rotondus_chapparvovirus
KX272741_rat_parvovirus_2
KM254174_Chicken_chapparvovirus_HK
MG846442_Chicken_chapparvovirus2
MK513533_Tasmanian_devil_associated_chapparvovirus_6
MK988619_Peafowl_parvovirus_1
KF925531_Turkey_parvovirus_2
MN175613_Psittacara_leucophtalmus_chapparvovirus
MK988620_Peafowl_parvovirus_2
KU563733_porcine_parvovirus_7
MN265364_capuchin_kidney_parvovirus
MF190038_Milport_starfish_parvo_like_virus
MF190039_Caledonia_starfish_parvo_like_virus
ADK63426_Fenneropenaeus_chinensis_hepandensovirus

Subfamily *Densovirinae* (- densovirus)
Subfamily *Parvovirinae*
Subfamily *Hamaparvovirinae* (- hamaparvovirus)

0.2

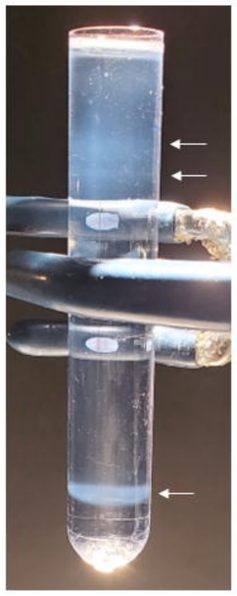
A



ORF1 VLPs

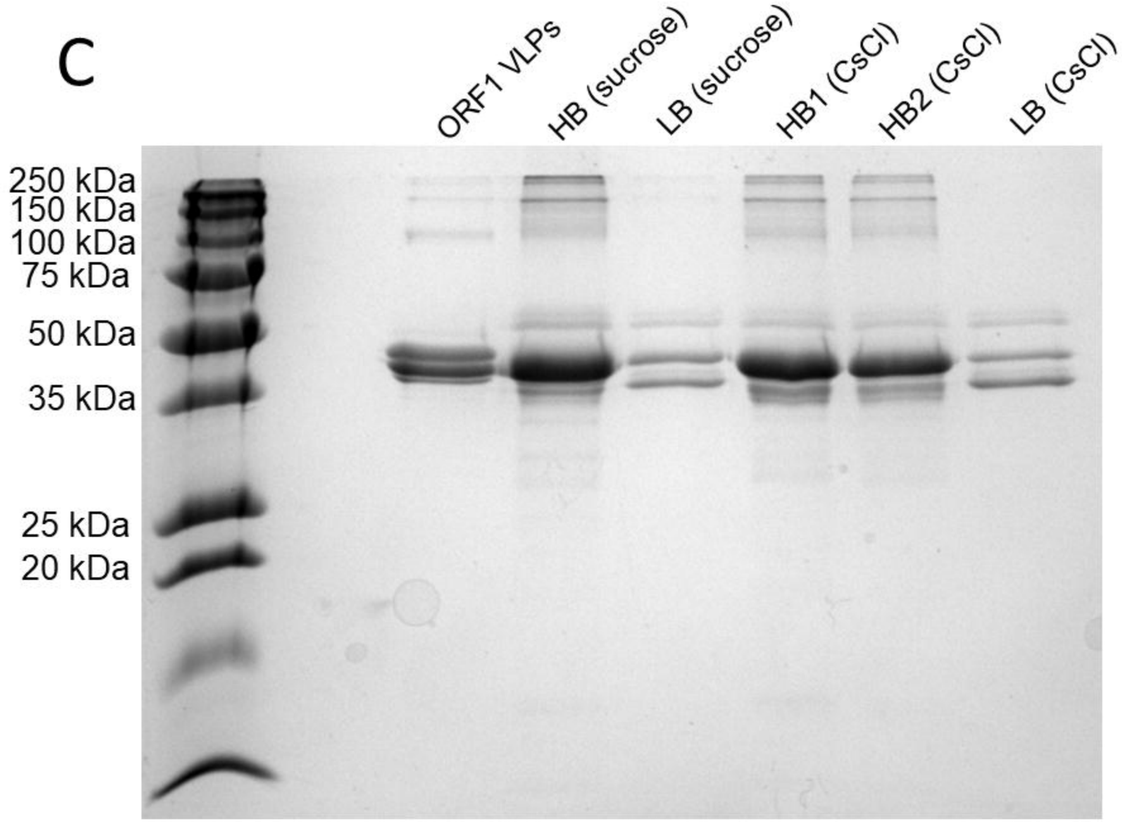
Infected
A. domestica

B



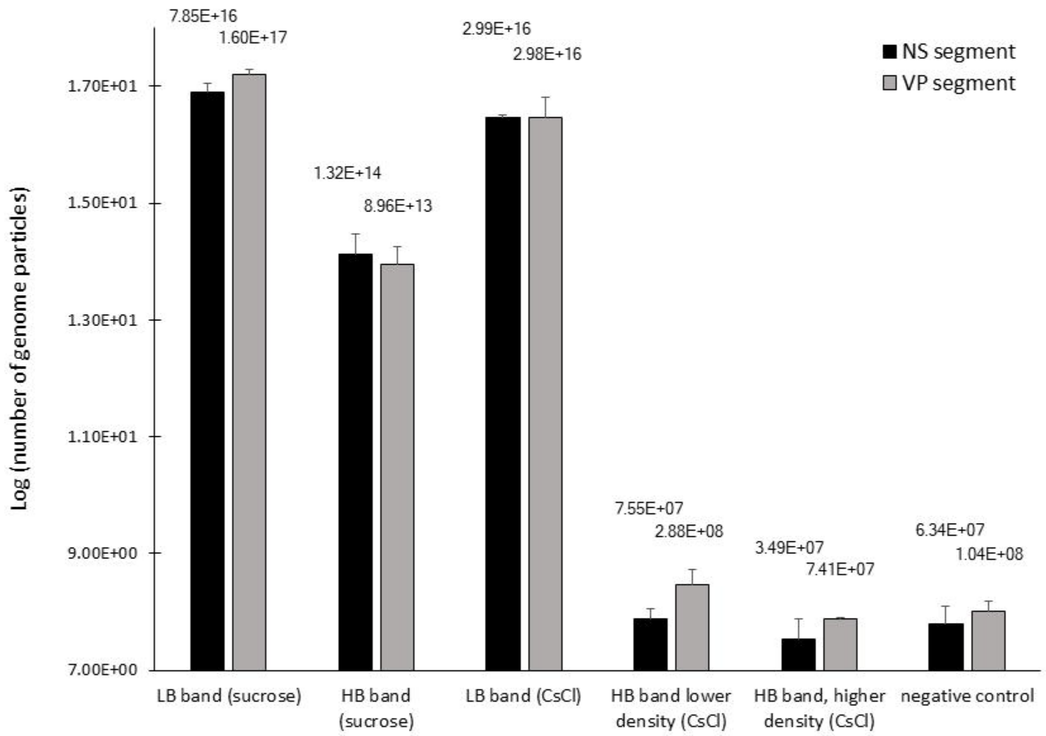
Infected
A. domestica

C

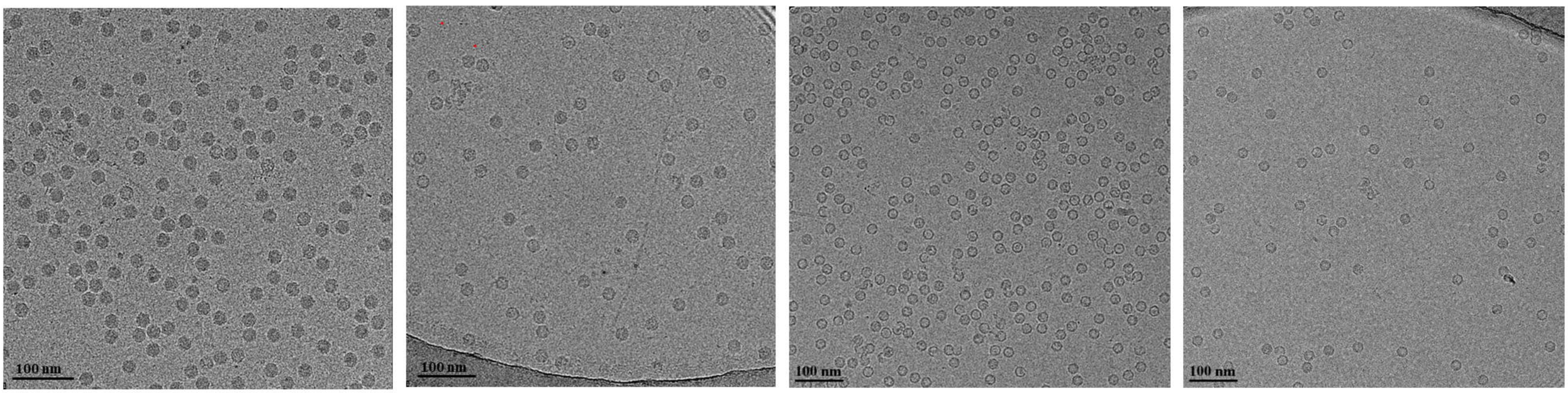


D

E



F



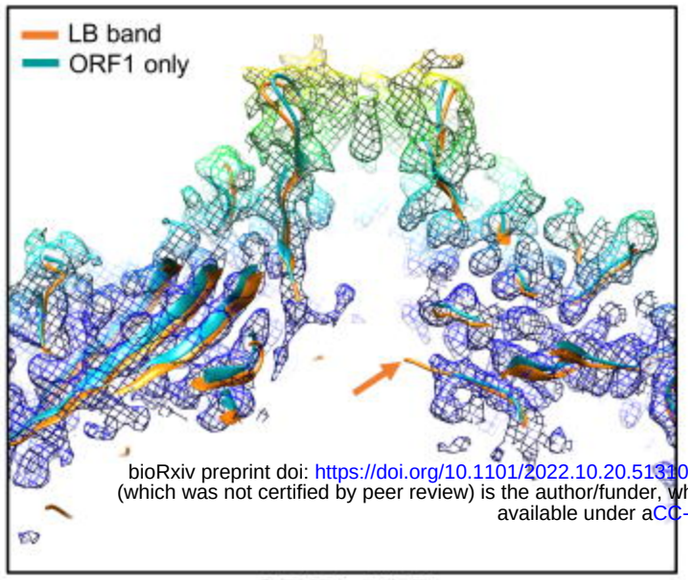
LB band

HB band

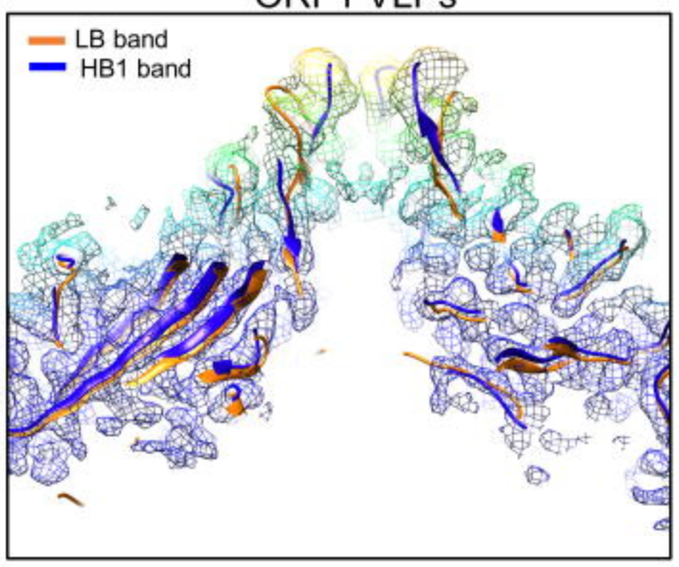
HB1 band

HB2 band

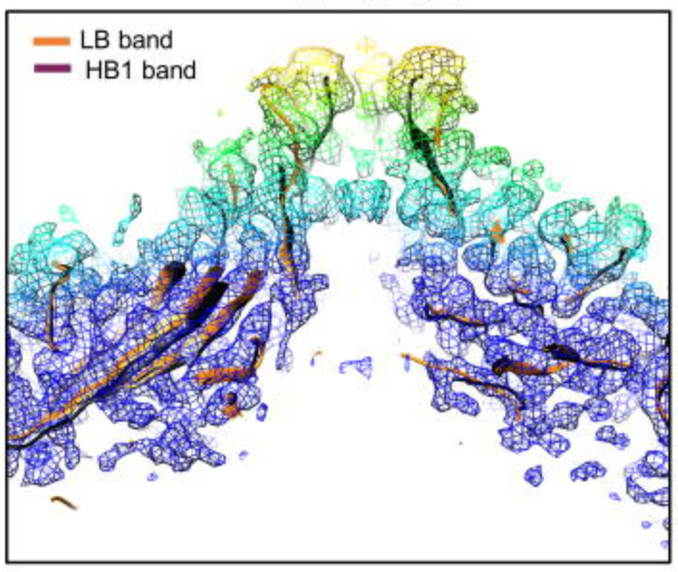
A



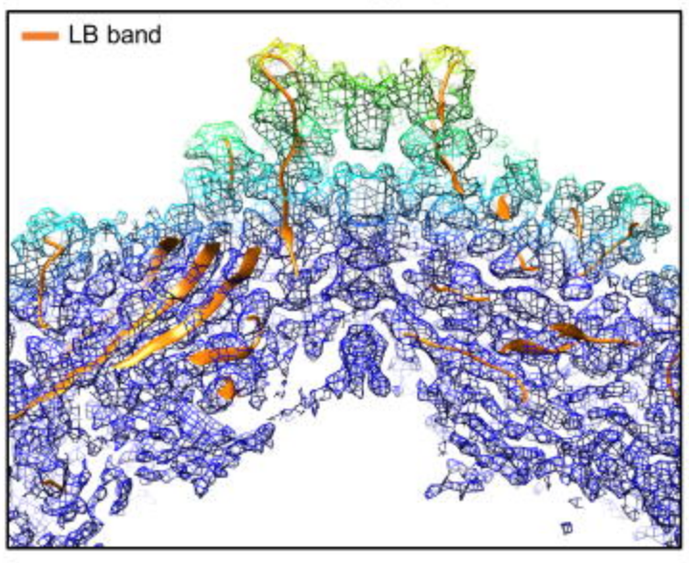
ORF1 VLPs



HB1 band

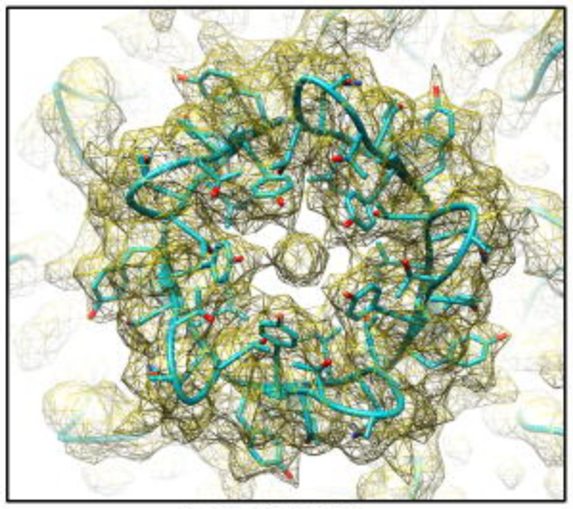


HB2 band

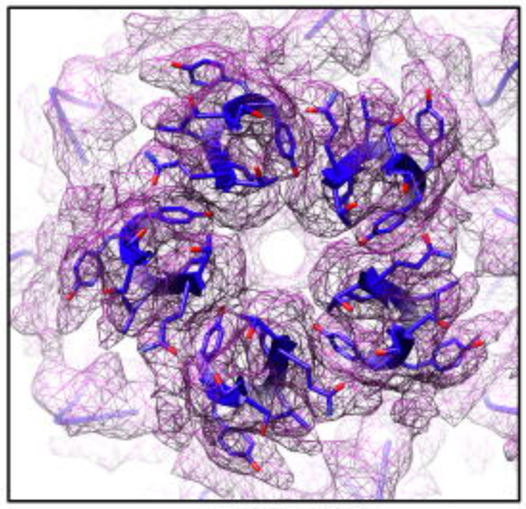


LB band

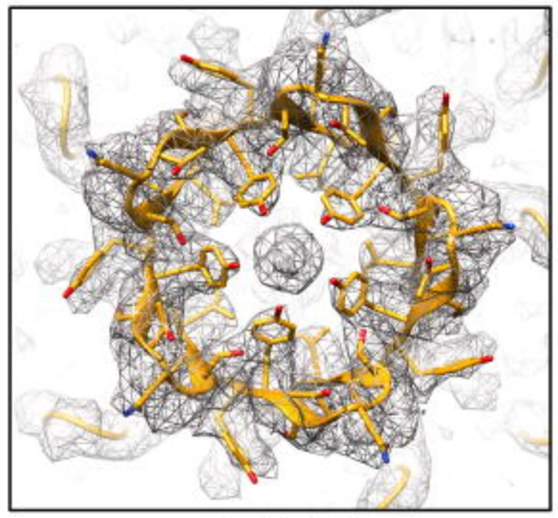
B



ORF1 VLPs

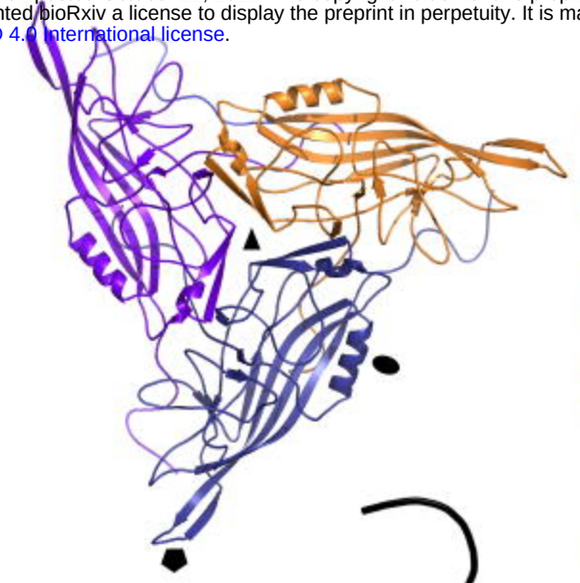


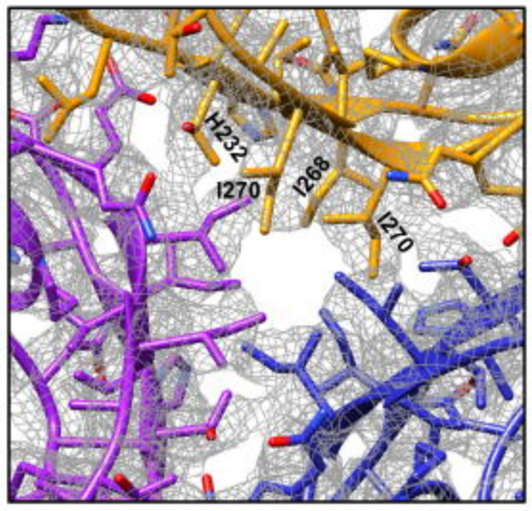
HB1/HB2 band

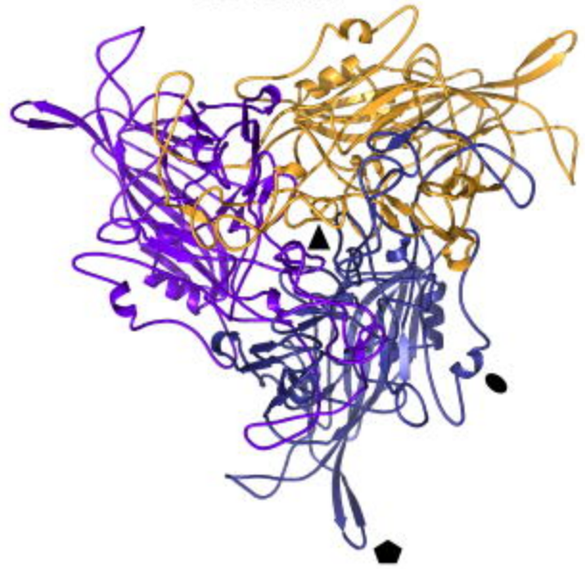


LB band

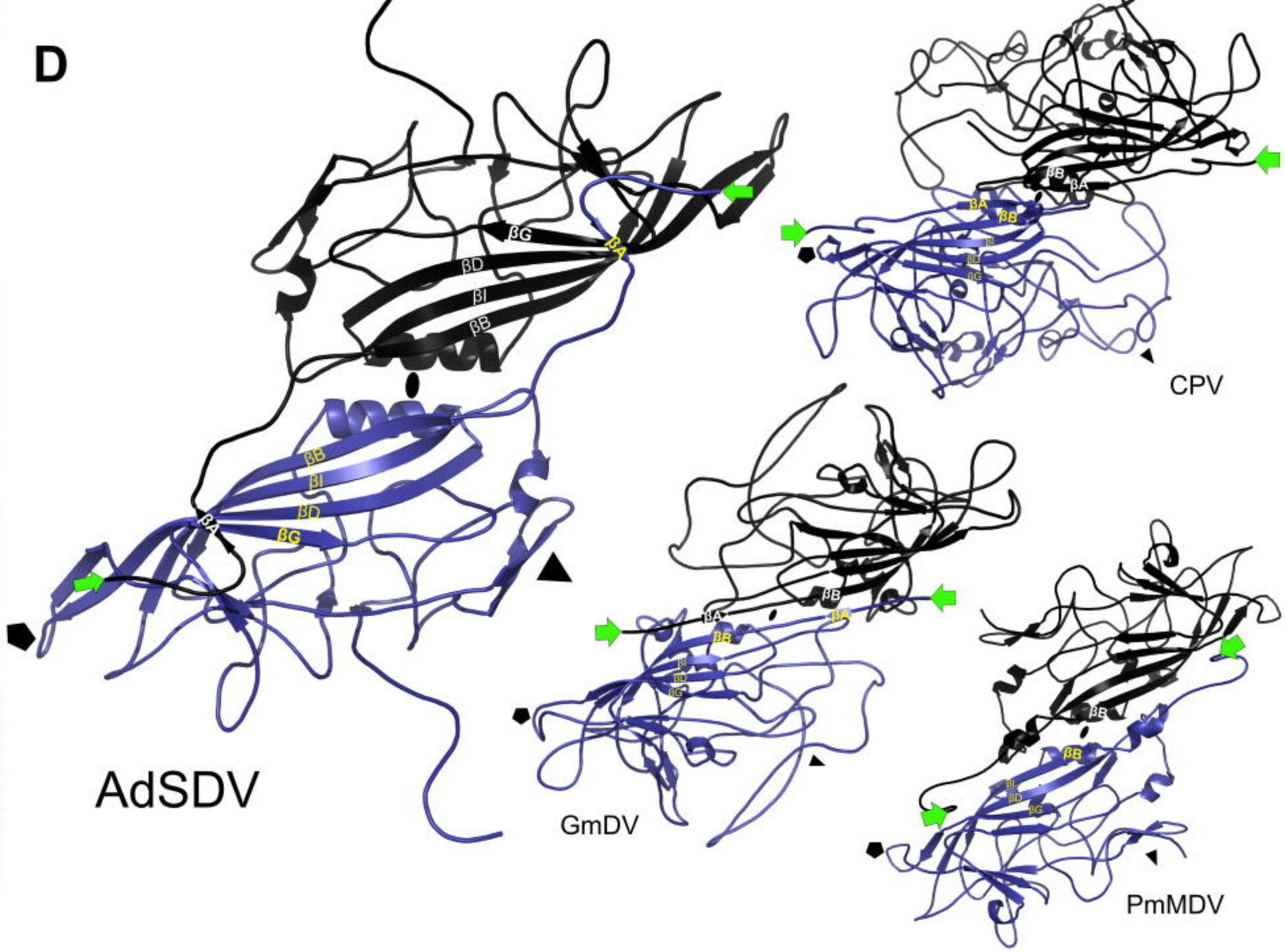
C



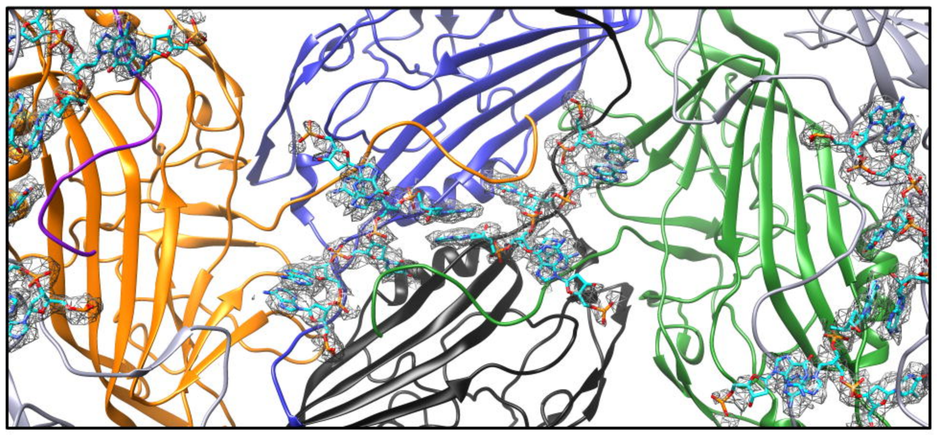




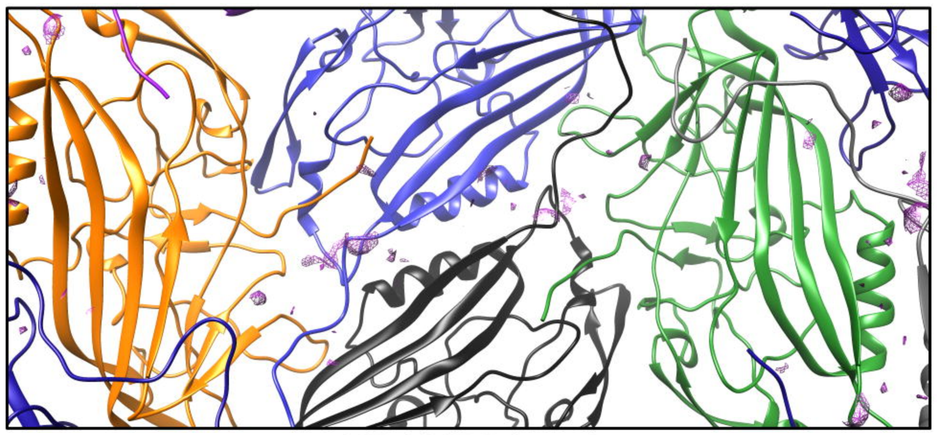
D



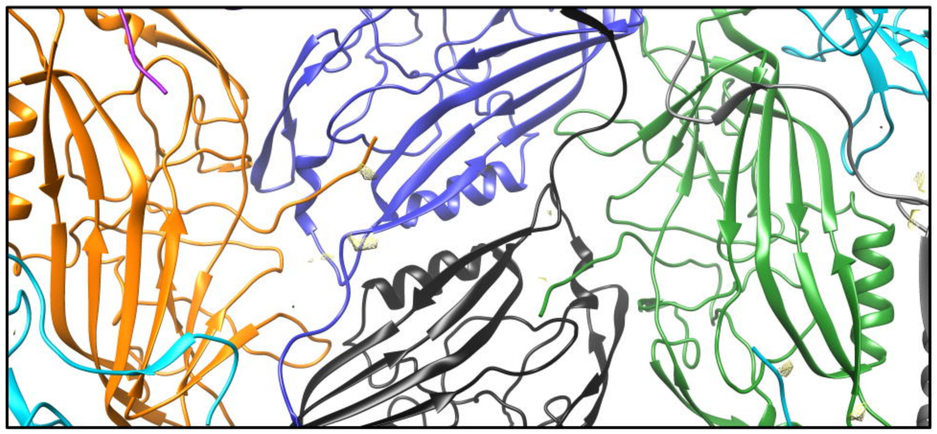
A



LB band

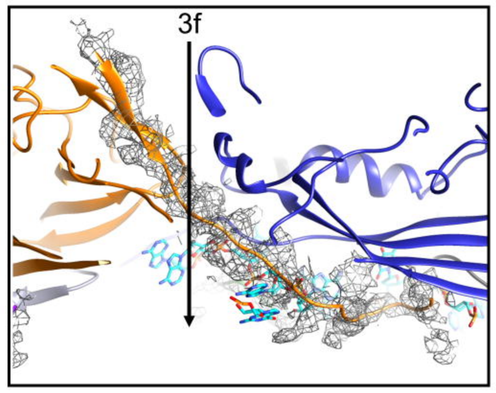


HB1/HB2 band

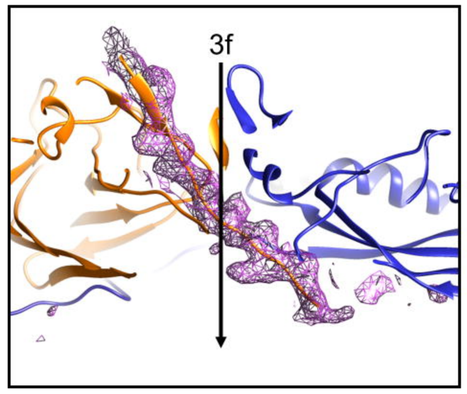


ORF1-VLPs

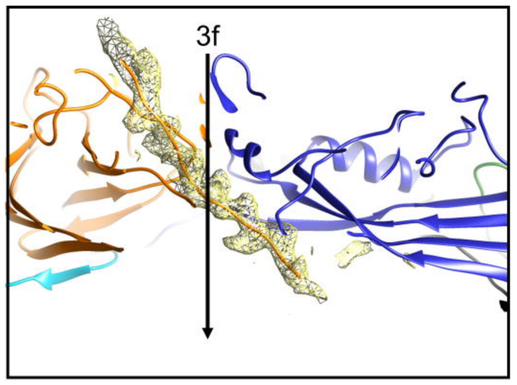
B



LB band

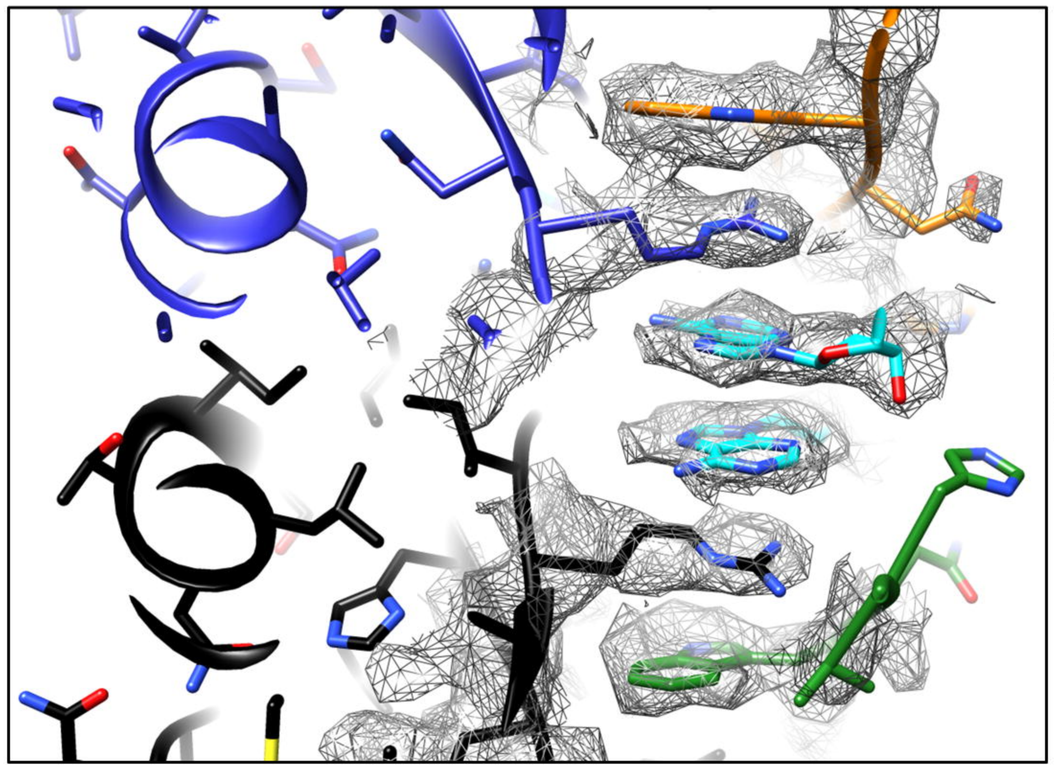


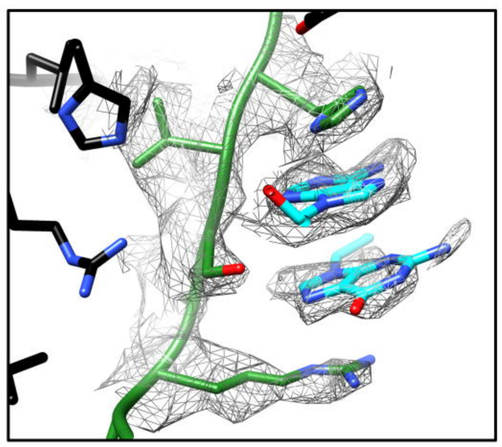
HB1 band



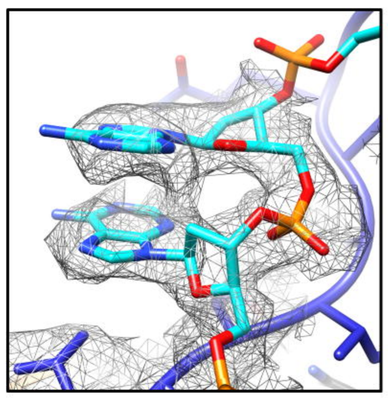
ORF1 VLPs

C

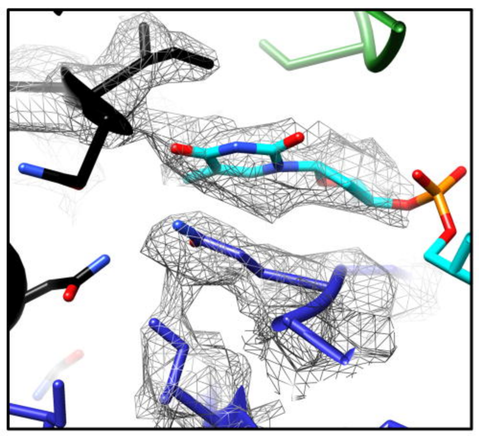




stack 1

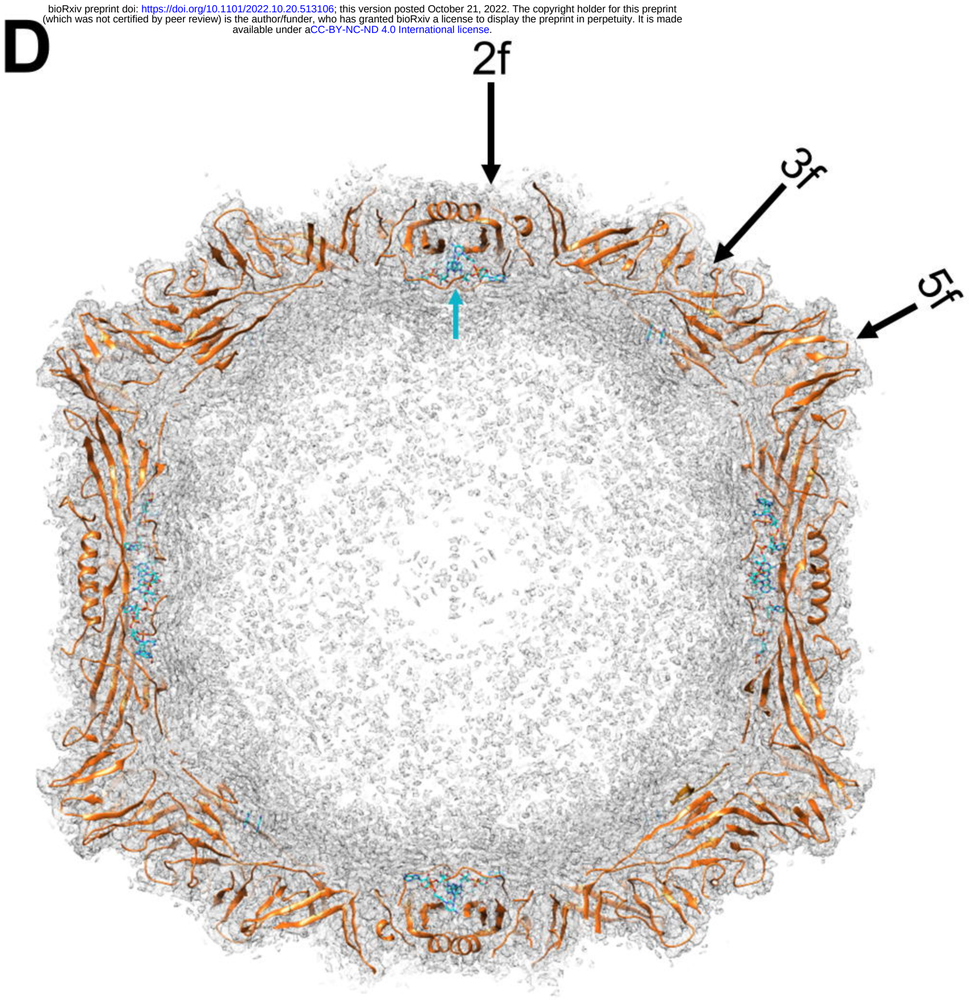


stack 3



stack 4

D



E