Adomaviruses: an emerging virus family provides insights into DNA virus evolution

Nicole L. Welch[1], Natalya Yutin[2], Jennifer A. Dill[3], Alvin C. Camus[3], Yuk-Ying S. Pang[1], John T. Schiller[1], Ping An[4], Paul G. Cantalupo[4], James M. Pipas[4], Eric Delwart[5], Samantha Koda[6], Kuttichantran Subramaniam[6], Thomas B. Waltzek[6], Chao Bian[7], Qiong Shi[7], Zhiqiang Ruan[7], Eugene V. Koonin[2¶], Christopher B. Buck[1¶*], and Terry Fei Fan Ng[3, 5 § ¶**]

[1]Lab of Cellular Oncology, NCI, NIH, Bethesda, MD, 20892

[2]National Center for Biotechnology Information, NLM, NIH, Bethesda, MD 20894, USA

[3]Dept. of Pathology, Univ. of Georgia, Athens, GA 30602

[4]Dept. of Biological Sciences, Univ. of Pittsburgh, Pittsburgh, PA 15260

[5]Blood Systems Research Institute, San Francisco, CA 94118

[6]Dept. of Infectious Disease and Pathology, Univ. of Florida, Gainesville, FL, 32611

[7]Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, BGI, Shenzhen, China, 518083

§ Current address: Division of Viral Diseases, NCIRD, CDC, Atlanta, GA 30333

*Corresponding author: buckc@mail.nih.gov

**Corresponding author: ylz9@cdc.gov

¶ These authors contributed equally to this work.

Abstract

Diverse eukaryotic dsDNA viruses, including adenoviruses, are thought to have evolved from bacteriophages of the family *Tectiviridae*. The evolutionary relationship of the small circular dsDNA tumor viruses of the families *Papillomaviridae* and *Polyomaviridae* to other DNA virus families remains uncertain. Metagenomic surveys of fish reveal 5 previously unknown circular dsDNA viruses that could become founding members of a distinct viral family. These viruses encode predicted superfamily 3 helicases that are related to the replicative helicases of polyomaviruses and papillomaviruses. Additionally, the new viruses encode a gene cluster coding for homologs of adenovirus-like maturation proteases and putative homologs of adenovirus major and minor capsid proteins. We show that these predicted capsid protein are indeed incorporated into the adomavirus virions. Such combination of genes from unrelated virus families is unprecedented among known DNA viruses. We propose the name "Adomaviridae" for this emerging virus family.

Significance Statement

Although investigation of the origins of viruses is hampered by the poor conservation of the sequences of many viral genes, substantial progress has been made by applying powerful computational methods to compare key viral proteins. Another challenge for understanding viral evolution is the rampant gene exchange between different groups of viruses, and between viruses and their hosts. We describe a group of fish viruses that possess an unexpected combination of genes. The proteins implicated in virus genome replication have homologs in polyomaviruses and papillomaviruses, whereas the structural gene module is related to the corresponding genes of adenoviruses. The unusual genomic layout of these fish viruses suggests unanticipated evolutionary connections between multiple families of tumor-causing viruses.

Introduction

Polyomaviruses and papillomaviruses (collectively denoted papovaviruses) are non-enveloped small double-stranded DNA (dsDNA) viruses named for their ability to cause tumors (Gk. *oma*) in their animal hosts (1). The widespread, typically species-specific distribution of polyomaviruses among bilateral animals suggests that they have been infecting these hosts for at least half a billion years (2). The origins of papillomaviruses appear to be similarly ancient (3, 4). The slow rate at which papovaviruses appear to accumulate genetic change might be partly attributable to the fact that their genomes are replicated with high-fidelity by host-cell DNA polymerases. A key defining feature of papovaviruses is the presence of a gene encoding a replicase protein (known as large T antigen (LT) in polyomaviruses or E1 in papillomaviruses) containing a superfamily 3 (SF3) helicase domain that unwinds the viral origin of replication, allowing the initiation of theta-type bi-directional replication by host cell machinery (5, 6). Papovavirus helicase domains share the closest sequence similarity with one another and with the SF3 helicase domain of the replicase proteins of parvoviruses, a family of small single-stranded DNA (ssDNA) viruses. An additional domain of the papovavirus replicase proteins, known as the origin binding domain (OBD), shares structural similarity with the endonuclease or "nickase" domain of parvovirus replicase proteins (7, 8). The nickase/OBD is distantly related to the Rep proteins of a diverse group of circular ssDNA "CRESS" viruses. These observations suggest a possible evolutionary scenario in which the papovavirus and parvovirus replicase genes originate from a circular ssDNA ancestor or ancestors (6).

Papovavirus capsid protein genes are poorly conserved, even within each of the two families. For example, the VP1 major capsid proteins (MCPs) of recently discovered arthropod polyomaviruses do not share obvious sequence similarity with the VP1 proteins of vertebrate polyomaviruses (2). Likewise, the L1 MCP of a recently described fish papillomavirus shows <30% identity with L1 proteins of papillomaviruses from terrestrial vertebrates (4). Although their amino acid sequences are not recognizably similar, the core folds of VP1 and L1 are essentially identical, consisting of an 8-stranded ß-jellyroll (also known as a ß-barrel) (9). In the assembled virion, the ß-strands of the MCP are arranged perpendicularly to the surface of the virion in a distinctive "knobby" icosahedral architecture of 72 homo-pentamers. A vast assemblage of other dsDNA viruses also have icosahedral virions constructed of jellyroll MCPs in which the ß-strands are perpendicular to the capsid surface. This group includes adenoviruses, virophages, virus-like polinton transposons and nucleo-cytoplasmic large DNA viruses (NCLDVs), as well as bacterial tectiviruses and archaeal sphaerolipoviruses (9-11). The T=7 geometry of papovavirus virions is simpler than those of other dsDNA viruses with perpendicular-jellyroll MCPs. For example, adenoviruses encode a tandem double-jellyroll MCP, known as "hexon," that trimerizes into pseudo-hexameric capsomers that form the facets of a T=25 virion. A separate single-jellyroll capsid protein known as "penton" occupies the 5-fold vertices of the adenovirus capsid. Virophages and many NCLDVs also encode a combination of a double-jellyroll MCP and a single-jellyroll minor capsid protein, although the latter has not yet been experimentally characterized (9). Sphaerolipoviruses, a family of tail-less dsDNA phages of archaea, encode two separate single-jellyroll MCPs (VP4 and VP7) that homo- and hetero-oligomerize to form pseudo-hexameric capsomers that make up the facets of the T=28 icosahedral virion (12, 13).

It has been proposed that papovavirus MCP genes evolved from the single-jellyroll MCP genes of an ancestral single-stranded RNA (ssRNA) or ssDNA virus (6). Although this scenario accounts for the conserved suite of observed structural similarities among parvovirus and papovavirus MCP and replicase proteins, the model invokes a baroque transition in the orientation of the MCP ß-jellyroll from the parallel orientation observed in parvoviruses (and many other ssRNA and ssDNA viruses) to the perpendicular orientation seen in papovaviruses and most other dsDNA viruses. Another conceivable model is that the perpendicular single-jellyroll papovavirus MCPs arose from an ancestral dsDNA virus with a multi-jellyroll capsid morphology. A third possible scenario is that the papovavirus MCPs were independently derived from a cellular single-jellyroll protein (9).

In 2011, a novel circular dsDNA virus was discovered in a disease outbreak in endangered Japanese eels (*Anguilla japonica*)(14, 15). Although the Japanese eel virus encodes a replicase protein homologous to fish polyomavirus LTs, the predicted gene products in the remainder of the ~15 kb circular dsDNA viral genome showed no obvious sequence similarity to known proteins. More recently, a related virus has been isolated from Taiwanese marbled eels (*Anguilla marmorata*)(16). In the current report, we characterize three additional fish-associated members of this emerging group of viruses. Our analysis reveals homology between a cluster of conserved genes in this clade of fish viruses and the structural genes of adenoviruses. We denote this emerging group of viruses, which can be expected to become a separate virus family, "Adomaviridae," to emphasize the combined adenovirus- and polyomavirus/papillomavirus-like features that are observed in their genomes.

Results

Characterization of guitarfish, discus, and arowana adomavirus genomes

Sequencing of nuclease-resistant DNA extracted from skin lesions observed on a giant guitarfish (*Rhynchobatus djiddensis*) revealed a circular ~22 kb adomavirus-like sequence. Pathological characteristics of the fish are detailed in a separate manuscript (Dill et al., manuscript in press, mBio).

Post-mortem metagenomics analysis of an aquarium-bred Amazon red discus cichlid (*Symphysodon discus*) exhibiting lethargy and inflamed skin lesions revealed a circular ~20 kb adomavirus-like genome. In contrast to the guitarfish, histopathological analysis of skin lesions from the discus specimen showed no evidence of intranuclear inclusions or other obvious histopathology.

A virus from green arowana (*Scleropages formosus*) was originally identified from short contigs with similarity to multiple segments of the Japanese eel adomavirus genome using tBlastn searches of the NCBI Whole Genome Shotgun (WGS) database entry (Biosample SAMN03894453). Subsequently, the complete adomavirus genome was confirmed by sequencing of overlapping PCR products amplified from DNA extracted from a snip of healthy fin tissue used in the WGS project (17).

Genome architecture and predicted proteins

A naming convention has previously been applied to Japanese eel adomavirus (NC_015123) based on the supposition that, similarly to the polyomaviruses, genes on the same strand as the LT ORF are expressed early in the infectious cycle whereas genes on the complementary strand are expressed late during infection. Although the five adomavirus genomes are highly divergent at the nucleotide and protein level, limited sequence identities (20%-45%) was observed between syntenic genes throughout the genome (Supplementary Table 1). Thus, we chose to adhere to the existing "early ORF" (EO)-"late ORF" (LO) gene naming convention for the hypothetically homologous ORFs in each of the adomaviruses (Figure 1).

Marbled eel adomavirus splicing pattern

RNA-seq data from a previous study (16) was used to analyze the splicing of marbled eel adomavirus transcripts (Supplemental Table 2). The paired-end reads were aligned to the marbled eel adomavirus genome with the splicing-aware aligner HISAT2. Three peaks of expression were observed at the 5' end of the genome upstream of the LO genes. This splicing pattern is reminiscent of the adenovirus tripartite leader (TPL), which enhances the translation of the viral late genes (18). Splice acceptors were found to be utilized immediately upstream of the inferred ATG codons of LO4, 5, 6, and 7 as well as EO2-4. An extensively utilized intron within the EO2-4 ORF appears to be consistent with the consensus splicing signals of minor-class (U12) introns.

Predicted structures and evolutionary relationships of inferred adomavirus replicase proteins

The replicase proteins (referred to as EO1) of each of the five adomaviruses encode a predicted helicase domain homologous to the SF3 helicases of papovaviruses and parvoviruses (Figure 1). In the three smaller adomaviruses, from eels and arowana, the helicase domains share high sequence similarity with polyomavirus LTs. In addition, the eel and arowana EO1 proteins share other signature LT features, including an N-terminal DnaJ domain, retinoblastoma (Rb) interaction motifs (LXCXE, LXXLFD, or related sequences), and a domain with clear similarity to polyomavirus OBDs (Supplemental Figure 1)(19)) Like other papovavirus OBDs, the EO1 genes lack the catalytic HUH motif required for nickase, which is an essential function in parvovirus and CRESS replicase proteins.

A recently identified polyomavirus protein designated ALTO is encoded by an overprinted ORF in the +1 frame of LT (20).. The EO1 genes of the marbled eel and arowana adomaviruses contain possible ALTO-like overprinted ORFs (Figure 1).

In contrast to the smaller adomaviruses, the syntenic EO1 genes of the larger adomaviruses, from guitarfish and discus, contain SF3 helicase domains with a much lower similarity to polyomavirus LT proteins. Nevertheless, the guitarfish and discus adomavirus helicase domains retain the three diagnostic helicase motifs (Walker A, Walker B and Switch 1)(19) (Supplemental Figures 1 and 2), indicating that these domains are functional helicases.

This conclusion is further supported by homology modeling of the structures of the putative adomavirus helicases, which reveal close structural similarity to the papovavirus helicase domains (Figure 2). HHPRED searches using an alignment of the guitarfish and discus adomavirus EO1 proteins revealed similarity (probability 99.6%) to the papillomavirus E1 OBD. The OBDs of these two larger adomaviruses also lack the nickase catalytic residues.

Phylogenetic analysis reflects the dichotomy of the helicase domains of the smaller and larger adomaviruses (Figure 3). The arowana and eel EO1 proteins occupy a basal position among the LT proteins of fish and arthropod polyomaviruses, whereas the helicase domains of the guitarfish and discus adomavirus EO1 sequences form a distinct phylogenetic branch that is roughly equidistant from polyomavirus LT proteins, papillomavirus E1 proteins, and parvovirus NS1 proteins. The basal phylogenetic position of the predicted helicases of the large adomaviruses implies an ancient common origin with the papovavirus and parvovirus helicases. In contrast, the predicted helicases of eel and arowana adomaviruses appears to have been acquired relatively recently from fish polyomaviruses, presumably via displacement of the ancestral adomavirus EO1 helicase gene.

LO4-8 gene arrays of adomaviruses are distantly related to the adenovirus structural and morphogenetic genes

Adenoviruses possess a tightly spaced cluster of genes that encode virion structural and maturation proteins. Pentameric capsomers of the adenovirus penton protein (encoded by gene III) form the base of the twelve icosahedral vertices of the virion. Trimers of the double-jelly-roll hexon proteins (encoded by gene II) form the facets of the virion. Capsid protein IV homo-trimerizes to form long fibers with C-terminal extensions pointing away from the vertices. The mature virion also contains minor structural components, including core protein X (pX, also known as L2 mu core protein), which is thought to help condense the viral chromatin (11). Another virion core protein, pVI, destabilizes cellular membranes during the infectious entry process (21). The adenovirus structural gene array also encodes a cysteine protease known as adenain, which cleaves seven virion-associated pre-proteins (including pX and pVI) during the virion maturation process (22).

Each of the five adomavirus LO8 proteins shows significant sequence similarity to human adenovirus 2 adenain in HHpred (23-25) searches, with probability values in the range of 91-97%. Examination of alignments of LO8 sequences with adenains reveals conservation of key structural elements as well as the catalytic cysteine and histidine residues (Supplemental Figure 3). Thus, LO8 genes can be confidently predicted to encode active thiol proteases homologous to adenain.

The LO6 ORFs of both eel adomaviruses show significant similarity to adenovirus pX (probabilities of 88-91%), suggesting that these proteins are homologous. Although pX is not a membrane protein, it contains a conserved hydrophobic segment that scores positive in algorithms designed to predict bacterial transmembrane domains http://tmdas.bioinfo.se/DAS/index.html (26). All 5 adomavirus LO6 proteins also encode

predicted transmembrane-like segments that are apparently homologous to those of pX. Papillomavirus L2 and polyomavirus VP2 virion interior proteins also contain conserved segments with transmembrane-like characteristics (27, 28). The role of these hydrophobic segments in virion morphogenesis remains to be elucidated.

Each of the adomavirus LO4 proteins contains a predicted C-terminal coiled-coil domain. Intriguingly, the top hit for Japanese eel adomavirus LO4 in HHPRED searches was the N-terminal coiled-coil domain of avian reovirus σC (also known as σ1), a homotrimeric coiled-coil fiber protein that extends from the vertices and mediates cell attachment (29, 30). Adenovirus fiber proteins also show HHPRED hits to mammalian reovirus type 3 σ1 fiber proteins. Notably, it has been shown that adenovirus fiber and reovirus σ1 are functionally interchangeable (31). If the predicted C-terminal coiled-coil of LO4 were structurally analogous to reovirus and adenovirus fiber proteins, the N-terminus of LO4 would be expected to be oriented toward the virion lumen, with little or none of its predicted C-terminal coiled-coil domain extending outside the virion. This would comport with the fact that adomaviruses, in contrast to reoviruses and adenoviruses, do not appear to have visible vertex fibers in negative stain electron micrographs (14, 16, 32).

The homology of the adomavirus LO6 and LO8 genes to two of the virion morphogenetic genes of adenoviruses (pX and adenain, respectively) led us to hypothesize that the LO5 and LO7 proteins might be syntenic homologs of adenovirus penton and hexon capsid proteins, respectively. Database searches failed to detect convincing similarity between the sequences of LO5 or LO7 and any other known proteins. Nevertheless, secondary structure predictions indicate that both proteins adopt predominantly β-sheet structures. Specifically, the structure-guided alignment analysis of LO5 appears to be compatible with a penton-like single-jellyroll fold (Supplemental Figure 4). The predicted structure of LO7 seems to be consistent with a hexon-like double-jellyroll fold as indicated by analysis that included double-jellyroll MCPs from multiple virus families (Supplemental Figure 5).

Adenovirus penton proteins can spontaneously assemble into 12-pentamer (T=1) subviral particles that might play a role of "decoy" pseudocapsids *in vivo* (33). The major capsid proteins of some dsDNA viruses, including polyomaviruses and papillomaviruses, can form virus-like particles (VLPs) when expressed in human embryonic kidney-derived 293TT cells (34). Codon-modified marbled eel LO1-LO8 expression plasmids were transfected individually into 293TT cells, and Optiprep ultracentrifugation was used to determine whether one or more of the proteins had self-assembled into particles. An HPV16 L1/L2 expression plasmid was used as a positive control for VLP formation (35-37). Cells transfected with either adomavirus LO4, LO5, or LO7 expression constructs showed robust production of particles that migrated down Optiprep gradients, whereas cells transfected with LO1, LO2-3, LO6, or LO8 alone did not show evidence of particle formation (Figure 4). Separate expression of LO5 and LO7 each produced irregular spherical particles, a subset of which were roughly 75 nm in diameter (Figure 5). LO4 particles were even more irregular, appearing as rough sheets bounded by smaller spherical structures.

In co-transfections of various combinations of LO genes, it was noted that inclusion of LO6 prevented the formation of LO5 and LO7 particles but did not impair the formation of LO4 particles (Supplemental Figure 6). Co-transfections of LO4-7 in which LO6 was included at $1/25^{th}$ of the transfection mixture produced slightly more regular spherical particles in some but

not all experiments. Inclusion of LO1, LO2-3, and/or LO8 in the co-transfection mixture did not show noticeable effects on particle regularity. These results are consistent with a model in which LO4, LO5, and LO7 are virion structural components. The ability of LO6 to antagonize LO5 and LO7 particle assembly when LO6 is overexpressed suggests that it is a minor virion component that interacts with LO5 and LO7.

Most individual LO4 and LO7 particle preparations showed dsDNA signal in Quant-iT PicoGreen assays (Invitrogen), suggesting the presence of encapsidated DNA in the purified particles. Optiprep-purified particle preparations from cells co-transfected with LO4, 5, and 7 were subjected to an additional round of nuclease digestion with salt-tolerant Benzonase (Sigma) followed by agarose gel filtration to remove the nuclease and any residual digested DNA fragments. The gel filtered particles reproducibly contained roughly seven nanograms of DNA per microgram of total protein, confirming the presence of nuclease-resistant DNA within the particles.

To test the prediction, based on the inferred homology with adenovirus virion proteins, that LO4-8 are adomavirus virion components, we grew marbled eel adomavirus in EK-1 cells and subjected Optiprep-purified virions to mass spectrometry. Although the exterior morphology of the purified virions was irregular (Figure 6), qPCR-based experiments using EK-1 target cells showed that the purified virions were infectious. Analysis of bands excised from Coomassie-stained SDS-PAGE gels indicated the presence of substantial amounts of actin and myosin derived from the host eel cell culture. Other prominent bands were identified as LO4, 5, 7, and 8. The relative band intensities of LO5 and LO7 are consistent with the prediction that the two proteins constitute penton and hexon subunits, respectively. Smaller bands in the gel yielded lower-confidence hits for LO6, consistent with the possibility that this protein is present in virions in an LO8-cleaved form. We compiled a list of protein modifications of marbled eel adomavirus genes LO4, LO5, and LO7 observed in the mass spectrometric analysis (Figure 6).

Taken together, the results are compatible with the prediction that LO5 is the counterpart of the adenovirus penton, LO7 corresponds to the adenovirus hexon, and LO8 is the adomavirus maturation protease. LO4 and LO6 also appear to be present in the assembled virion but their structural role cannot be conclusively inferred.

Adomaviruses encode diverged archaeal-eukaryotic primases

Blastp and HHpred searches show highly significant sequence similarity (HHpred probability scores of 95-100%) between the EO2-4 proteins of marbled eel, arowana, and guitarfish adomaviruses and the small (catalytic) subunits of archaeo-eukaryotic primases (AEPs). The catalytic residues of the AEP family (8) are conserved in the EO2-4 proteins of these adomaviruses, suggesting that these genes encode active primases (Supplemental Figure 7). In the discus adomavirus, a small segment of the predicted EO2 protein shows weaker similarity to AEPs (HHpred score of 65%) and the catalytic motifs are disrupted, indicating that the primase is inactivated in this virus. Although the Japanese eel adomavirus proteins show no detectable similarity to AEP, it remains conceivable that an ancestral AEP sequence was present but has deteriorated beyond recognition in this virus.

Most large dsDNA viruses of eukaryotes, including NCLDVs, herpesviruses, and baculoviruses, encode AEPs (8). Although dsDNA viruses with moderate-size genomes, such as adenoviruses and virophages, typically lack primases, a handful of the more recently discovered moderate-sized polinton-related viruses encode AEPs (38, 39). The adomavirus AEP domains do not show phylogenetic affinity with any known viral AEPs. Instead, the adomavirus AEP homologs form branches deeply embedded in the eukaryotic part of the AEP tree (Figure 7). This phylogenetic position suggests that ancestral adomaviruses independently captured the AEP gene from a eukaryotic host.

Predicted histone lysine methyltransferases in adomaviruses

In the guitarfish adomavirus, a short ORF upstream of EO1 encodes a protein strongly similar (BLAST E-value < e-18) to cellular Su(var)3-9 Enhancer-of-Zeste and Trithorax (SET) domains, which are histone lysine methyltransferases involved in regulation of chromatin structure and transcription (40). A homologous gene in the discus adomavirus shows more limited similarity to SET domains and might be enzymatically inactive. Although many NCLDVs encode SET domain methyltransferases (41), the predicted adomavirus SET proteins are only weakly similar to these viral proteins (Supplemental Figure 8), suggesting independent acquisition from a eukaryotic host. Thus, the likely scenario for the evolution of this gene would mimic that of the adomavirus AEP, with apparent loss in some adomavirus lineages.

Discussion

Recent comparative analyses of viral genomes, particularly those extracted from metagenomics data, indicate that chimerization, typically in the form of gene module recombination, is a major route of virus evolution (6, 39). Among the papovaviruses, a notable example is an apparently chimeric lineage of viruses found in western barred bandicoots (an Australian marsupial species)(42). The bandicoot viruses combine an LT gene typical of avian polyomaviruses with the major and minor capsid protein genes (L1 and L2) of a marsupial papillomavirus (43). Another striking case of extensive chimerism has been identified in bidnaviruses, a family of insect ssDNA viruses that appear to have evolved via hybridization between a parvovirus and a polinton-related virus, with possible additional contributions from baculoviruses and reoviruses (44). Eel- and arowana-tropic adomaviruses offer another clear example of inter-familial chimerization, whereby polyomavirus-derived LT genes combine with structural and morphogenetic genes of adenoviruses. Although the simultaneous detection of an adomavirus and a polyomavirus in a single guitarfish tissue specimen highlights the plausibility of inter-familial horizontal gene transfer resulting from superinfection, the guitarfish- and discus-derived adomaviruses do not appear to be the result of chimerization with currently known polyomaviruses. Instead, the larger guitarfish and discus viruses encode a previously unknown variety of SF3 helicases that are phylogenetically equidistant from polyomavirus, papillomavirus, and parvovirus helicases. The presence of highly derived AEP genes in 4 of the 5 adomavirus species and divergent SET-domain genes in two of the 5 viruses is consistent with the possibility that adomaviruses emerged as a distinct lineage in an early animal host.

The majority of known viruses with perpendicular-jellyroll MCPs possess a distinct DNA-packaging ATPase of the FtsK family, whereas adenoviruses encode a distinct ATPase implicated in the same function (6). Adomaviruses, like papovaviruses, encode only one ATPase, the SF3 helicase. Although involvement of the papovavirus SF3 helicase in DNA packaging cannot be ruled out, it has been shown that polyomavirus and papillomavirus reporter "pseudovirions" can productively assemble around genome-sized dsDNA molecules in the absence of E1 or LT expression (34, 36, 45). Recent experiments using cell-free systems have confirmed that ATPase activity is not required for papillomavirus virion assembly or DNA packaging (46, 47). The assembly process could instead occur via a more passive, disulfide-driven Brownian ratchet mechanism (48). A passive packaging mechanism is consistent with the relatively low density of genomic DNA in papovavirus virions. Specifically, simple calculations based on virion diameter and genome size show that papovaviruses carry about 75 basepairs of DNA for every 10 nm^3 of virion volume. Adenoviruses, in contrast, carry about 95 basepairs of DNA per 10 nm^3, whereas bacteriophage PRD1, which has been directly shown to couple ATP hydrolysis to DNA insertion through a unique portal vertex(49), carries about 110 basepairs per 10 nm^3. By this measure, the smaller eel adomaviruses are more akin to the papovaviruses, with only about 70 basepairs per 10 nm^3. It thus seems possible that adomaviruses use the same, seemingly motor-independent assembly mechanisms that have previously been proposed for papovaviruses.

The discovery of adomaviruses in multiple, distantly related fish species raises the question of whether these viruses exist in terrestrial vertebrates. The identification of the arowana adomavirus as an accidental sequence contaminant in a shotgun genomics effort suggests that future WGS or transcriptomics datasets could be a profitable place to search for related viruses.

The overall similarity of the adomavirus LO4-8 gene block to the structural gene array of adenoviruses seems best compatible with LO7 being a double-jelly-roll MCP and LO5 being a single-jelly-roll minor capsid protein. Although unlikely, it remains possible, in principle, that LO7 is a single-jellyroll protein that homo- or hetero-oligomerizes to form the virion facets. The observation that both LO5 and LO7 can each independently form irregular ~75 nm particles, the assembly of which is antagonized by the expression of the putative pX-like core protein LO6, might be consistent with the possibility that adomavirus capsids are assembled from two separate single-jellyroll MCPs. A limitation of this study is that the virions and VLPs we observed were too few and irregular for detailed structural analysis with methods such as cryo-electron microscopy. Future work elucidating the fine-structure of these viruses will be needed to confirm that LO7 encodes a double-jellyroll MCP akin to adenovirus hexon rather than a single-jellyroll MCP like papovaviruses and sphaerolipoviruses. A better understanding of adomavirus capsid structure could shed important light on the evolutionary relationships between different types of perpendicular-jelly-roll MCPs.

In conclusion, adomavirus genomes combine a papovavirus-like replication module with a structural-morphogenetic module resembling those of adenoviruses. These distinct adomavirus genetic modules appear to be only distantly related to the known extant polyomaviruses, papillomaviruses, and especially adenoviruses, consistent with previous suggestions that these three lineages existed independently in early animals. More extensive genomic and metagenomic

studies of viruses of fish and invertebrates can be expected to provide more definitive information on the provenance of adomavirus genetic modules.

Materials and Methods

Sample Collection and cell culture

The pathology of a giant guitarfish (*Rhynchobatus djiddensis*) from which an adomavirus was isolated has been described previously (32). The same individual animal was also the source of a previously reported guitarfish polyomavirus (2, 50). Additional investigation of the animal's long-term disease course, and its associated pathological changes are detailed in separate manuscripts (Dill et al., 2018, in press).

A red discus cichlid (*Symphysodon discus*) was purchased a pet shop in Gainesville, Florida. The fish was moribund and showed erythematous skin lesions. Propagation of the discus adomavirus in cell culture was attempted by overlaying the skin tissue homogenates on Grunt Fin (GF) (Fish origin, ATCC) and Epithelioma Papulosum Cyprini (EPC) cell lines (Fathead minnow origin, ATCC). Neither cytopathic effects nor qPCR-based detection of viral replication was observed during two blind passages of 14 days a piece.

Dr. Chiu-Ming Wen generously provided EK-1 cells (a Japanese eel kidney line) infected with the Taiwanese marbled eel adomavirus (16). The viral culture was propagated by inoculation of supernatants from the infected culture into uninfected EK-1 cells cultured at room temperature in DMEM with 10 % fetal calf serum.

Human embryonic kidney-derived 293TT cells were cultured from in-house collection as previously described (34).

Viral genome sequencing

The guitarfish, marbled eel, discus, and arowana samples were prepared in four different laboratories (Supplemental Table 3). Unbiased metagenomics sequencing of nuclease-resistant nucleic acids was performed directly on guitarfish lesional material according to previously described protocols (51-54). In brief, homogenated skin lesions were filtered, followed by depletion of host nucleic acids in the filtrate using nucleases. Nuclease-resistant nucleic acids were extracted using the QIAamp Viral RNA Mini Kit (Qiagen) and sequence-independent amplification was performed using random priming. First strand synthesis was performed using a 28-base oligonucleotide whose 3′ end consisted of eight random nucleotides (primer N1_8N, CCTTGAAGGCGGACTGTGAGNNNNNNNN). The second strand was synthesized using Klenow fragment DNA polymerase (New England BioLabs). The resulting double-stranded DNA and cDNA were then PCR amplified using AmpliTaq Gold DNA polymerase and a 20-base primer (primer N1, CCTTGAAGGCGGACTGTGAG). This random amplicon was processed with Illumina library construction as described below.

Marbled eel adomavirus was cultured in EK-1, as previously described (16). Virions were purified from lysates of infected cells using Optiprep gradient ultracentrifugation (55). DNA extracted from Optiprep gradient fractions was subjected to rolling circle amplification (RCA,

TempliPhi, GE Health Sciences). For the discus adomavirus, total DNA was extracted directly from a skin lesion and subjected to deep sequencing. For the arowana adomavirus, overlapping PCR primers were designed based on GenBank accession numbers LGSE01029406, LGSE01031009, LGSE01028643, LGSE01028176, and LGSE01030049. PCR products were subjected to primer walking Sanger sequencing.

The guitarfish random amplicon, the marbled eel adomavirus RCA products, and the discus adomavirus total DNA were subjected to Nextera XT DNA Sample Prep kit and sequenced using the MiSeq (Illumina) sequencing system with 2 × 250 bp paired-end sequencing reagents. In addition, the marbled eel RCA product was digested with *AclI* and *EcoRI*. The resulting early and late halves of the viral genome were cloned separately into the *AclI* and *EcoRI* restriction sites of pAsylum+. The sequence of the cloned genome was verified by a combination MiSeq and Sanger sequencing. The clones will be made available upon request.

NGS data were analyzed using a bioinformatics pipeline, as previously described (51, 53). Adapter and primer sequences were trimmed using NCBI VecScreen and duplicate reads and low-quality tails were removed using a Phred quality score of 10 as the threshold. The cleaned reads were de novo assembled using an in-house sequence assembler employing an ensemble strategy consisting of SOAPdenovo2, AbySS, meta-Velvet, and CAP3, or solely by SPAdes (56, 57). The assembled sequences were compared with an in-house viral proteome database using BLASTx. Once the viral contigs were identified, they were further assembled iteratively to obtain the complete circular genome. Each viral genome was inspected manually base-by-base for sequence quality and coverage. The genome of the previously sequenced Japanese eel adomavirus (also known as Japanese eel endothelial cells-infecting virus, JEECV) was retrieved from GenBank using accession NC_015123. The genomes of adomaviruses described in this report have been assigned accession numbers MF946548, MF946548, MF946550, and MH282863.

Marbled eel adomavirus genomic mRNA analysis, late ORF expression, and viral purification

RNA-seq reads from Wen et al (16) were aligned to the marbled eel adomavirus genome (KX781210.1; isolate AMV-6) using HISAT2 version 2.0.5 with the following options: "--rna-strandness FR --dta --no-mixed --no-discordant". Integrated Genome Viewer (IGV) version 2.4.9 was used to determine splice junctions and their depth of coverage.

Codon-modified expression constructs encoding the marbled eel adomavirus LO1-LO8 proteins were designed according to a modified version of a previously reported algorithm (58). Detailed protocols and maps of expression plasmids used in this work are posted at https://home.ccr.cancer.gov/Lco/ . 293TT cells were transfected with combinations of LO expression constructs for roughly 48 hours. Cells were lysed in a small volume of PBS with 0.5% Triton X-100 or Brij-58 and Benzonase Dnase/Rnase (Sigma)(35). After one hour of maturation at neutral pH, the lysate was clarified at 5000 x g for 10 min. The clarified lysate was loaded onto a 15-27-33-39-46% Optiprep gradient in PBS with 0.8 M NaCl. Gradient fractions were collected by bottom puncture of the tube and screened by PicoGreen dsDNA stain (Invitrogen), BCA, or SDS-PAGE analysis. Electron microscopic analysis was performed by

spotting several microliters of Optiprep fraction material (or, in some instances, particles exchanged out of Optiprep through agarose gel filtration onto carbon support film on Cu/Ni/Au grids, followed by staining with 0.5% uranyl acetate.

EK-1 cells infected with the marbled eel adomavirus were purified following the same protocol as the late ORF expression in 293TT cells. Following agarose gel filtration, peak fractions were combined and precipitated with trichloroacetic acid (TCA). To the pooled 1 ml protein sample, 100 µl of 0.15% DOC was added and the resulting sample was incubated at room temperature for 10 minutes. 100 µl of 100% TCA was then added, the sample was vortexed and incubated on ice for 30 minutes. Following the incubation, the sample was centrifuged at 10,000 x g for 10 minutes at 4°C. The supernatant was carefully removed, and the remaining pellet was washed with ice-cold acetone to remove residual TCA.

Gene annotation, protein sequence analysis, structural modeling and surface conservation analysis

The protein-coding genes in the virus genomes were predicted using the GeneMarkS program with the settings adjusted for viruses. Protein sequence database searches were performed using PSI-BLAST, against the non-redundant protein sequence database at the NCBI (NIH, Bethesda), and using HHpred (25) against the PDB and Pfam-A, NCBI Conserved Domains, and PRK databases. Multiple sequence alignments were constructed using MUSCLE software.

The ProtSkin server (http://www.mcgnmr.mcgill.ca/ProtSkin/) (59) was used to assess the sequence-structure conservation among polyomavirus LTs and papillomavirus E1s. Using the protein sequence of the ATPase domain of SV40 LT (aa. 356 to 627) as the query, DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) identified 235 reference LT and E1 sequences as top hits (E-value $\leq$ 3e-17). The alignment of these sequences was uploaded to the ProtSkin server for conservation analysis. A conservation score was calculated for each alignment residue, and then mapped onto the solved structure of SV40 LT. The final conservation map was visualized using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

For structural modeling, the EO1 protein sequences of the five adomaviruses were analyzed using the PsiPred website http://bioinf.cs.ucl.ac.uk/psipred/ (60) to determine suitable templates. Both SV40 LT and HPV18 E1 were identified as templates with high confidence. The models were generated through PsiPred server and visualized and verified using PyMOL.

Phylogenetic analysis

Maximum likelihood (ML) and Bayesian inference analyses were performed for phylogenetic analysis of adomavirus genes. For comparison, representative papillomavirus E1, polyomavirus LT, parvovirus NS1/Rep, and AEP protein sequences were downloaded from GenBank. A subset of polyomavirus LT proteins were downloaded from PyVE https://home.ccr.cancer.gov/Lco/PyVE.asp. The compiled sequences were aligned using MUSCLE software (61) and trimmed to include only the OBD (or nickase) and SF3 helicase domains. ML phylogenetic trees were analyzed using the PhyML (62), in which the best protein substitution model (LG) was analyzed and obtained by Smart Model Selection (63) using the

Bayesian Information Criterion. Trees were further improved by Subtree-Pruning-Regrafting (SPR) and were displayed using FigTree http://tree.bio.ed.ac.uk/software/figtree/. Bayesian inference analysis was performed independently using MrBayes 3.1.2 (64), with the Markov chain run for a maximum of 1 million generations. Every 50 generations were sampled and the first 25% of mcmc samples were discarded as burn-in. The ML and Bayesian trees were compared for consistency in tree topography. Pairwise sequence identities were calculated using the translated amino acid sequences with the Sequence Demarcation Tool (65).

Mass Spectrometry

After marbled eel adomavirus virions were precipitated by TCA, the pellet was mixed 1:3 with NuPAGE Sample Buffer + 15% BME (Sigma) and run on a 10-12% Bis-Tris SDS-PAGE gel (Thermo). The protein bands were visualized using InstantBlue (Expedeon). Thirteen gel bands were individually excised and placed into 1.5 ml Eppendorf tubes. The gel bands were sent to the National Cancer Institute in Fredrick, Maryland where they were de-stained, digested with Trypsin, and processed on the Thermo Fisher Q Exactive HF Mass Spectrophotometer. Thermo Proteome Discoverer 2.2 software was used for protein identification. The uninterpreted mass spectral data were searched against Anguilla proteins (Swiss-Prot and TrEMBL database containing 105,268 proteins), Bos taurus proteins (Swiss-Prot and TrEMBL database containing 48,288 proteins), a common contaminants database (cRAPome), and translated marbled eel adomavirus ORFs. Further analysis of the bands containing marbled eel adomavirus peptides was conducted using the Protein Metrics Biopharma Software to identify modifications missed from the first round of identification.

Ethics Statement

All animal tissue samples were received as diagnostic specimens for pathogen testing and disease investigation purpose.

Acknowledgments

The authors are grateful to Lisa Jenkins for her extensive technical guidance on analyzing the raw mass spectral data and to Diana Pastrana and Gabe Starrett for useful discussions.

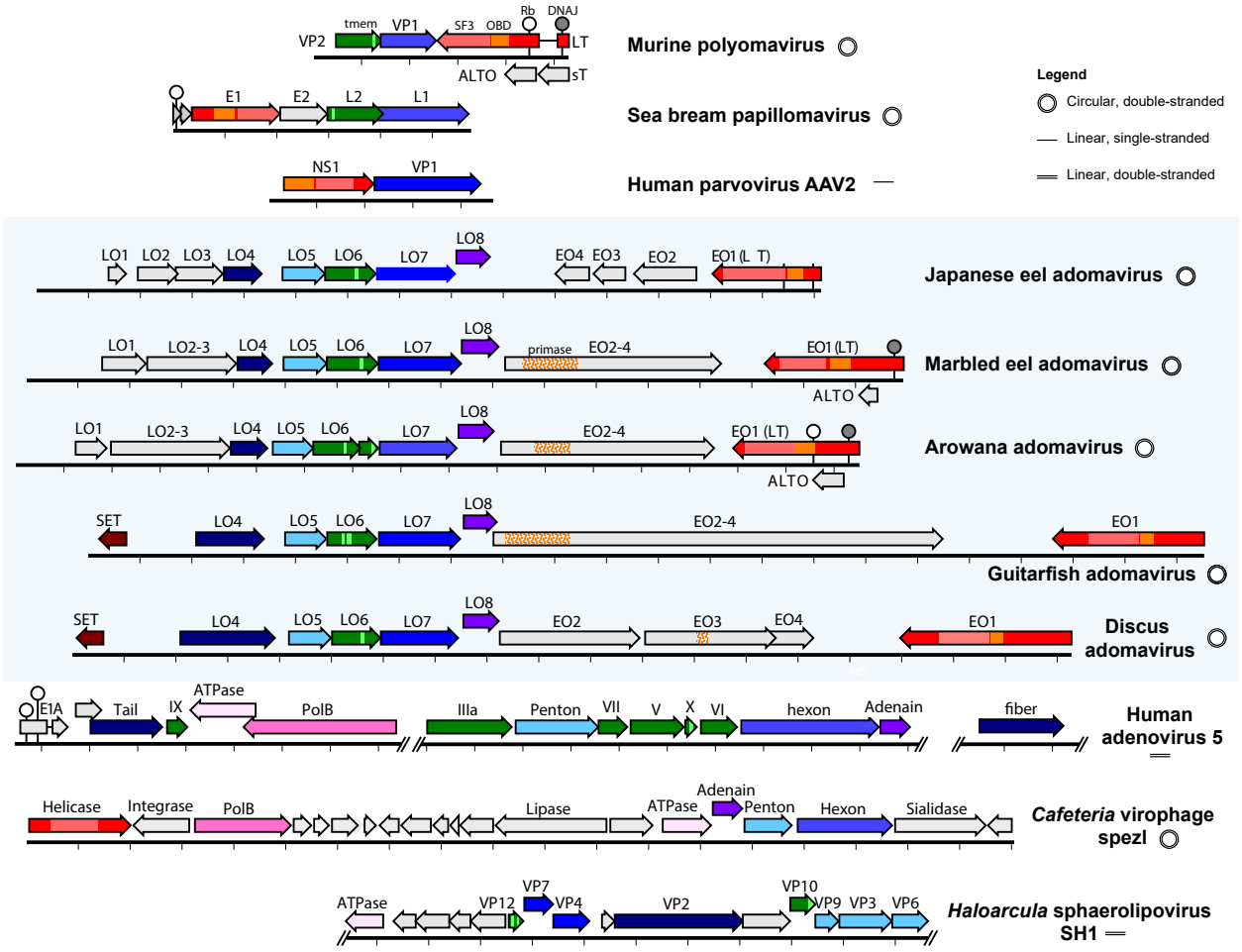

Figure 1: Comparative genomics of adomaviruses with other dsDNA viral families. Maps for the indicated viruses are drawn to a common scale, and circular genomes are linearized for display. Rulers indicate 1 kb intervals. Replicase genes (LT, E1, NS1, EO1) are colored red, with SF3 helicase domains shaded pink. Papovavirus origin-binding domains and the homologous parvovirus endonuclease domain are shaded orange. AEP domains are stippled orange. The common use of orange coloring denotes the similar folds of these domains (8). DNAJ domains are represented with gray lollipops and predicted retinoblastoma-interaction motifs are represented as white lollipops. Inferred major capsid proteins (MCPs) are colored medium blue. Genes thought to encode vertex-associated minor capsid proteins are colored sky blue. Predicted coiled-coil vertex fiber proteins are dark blue. Proteins thought to occupy the interior of the virion are colored green, with lighter green boxes indicating predicted transmembrane helices. Hashes represent sections of genomes removed for display purposes. GenBank accession numbers: murine polyomavirus, U27813; Sea bream papillomavirus, NC_030839; human parvovirus AAV2, NC_001401; human adenovirus 5, AC_000008; Cafeteria virophage spezl, NC_015230; Haloarcula sphaerolipovirus SH1, NC_007217.

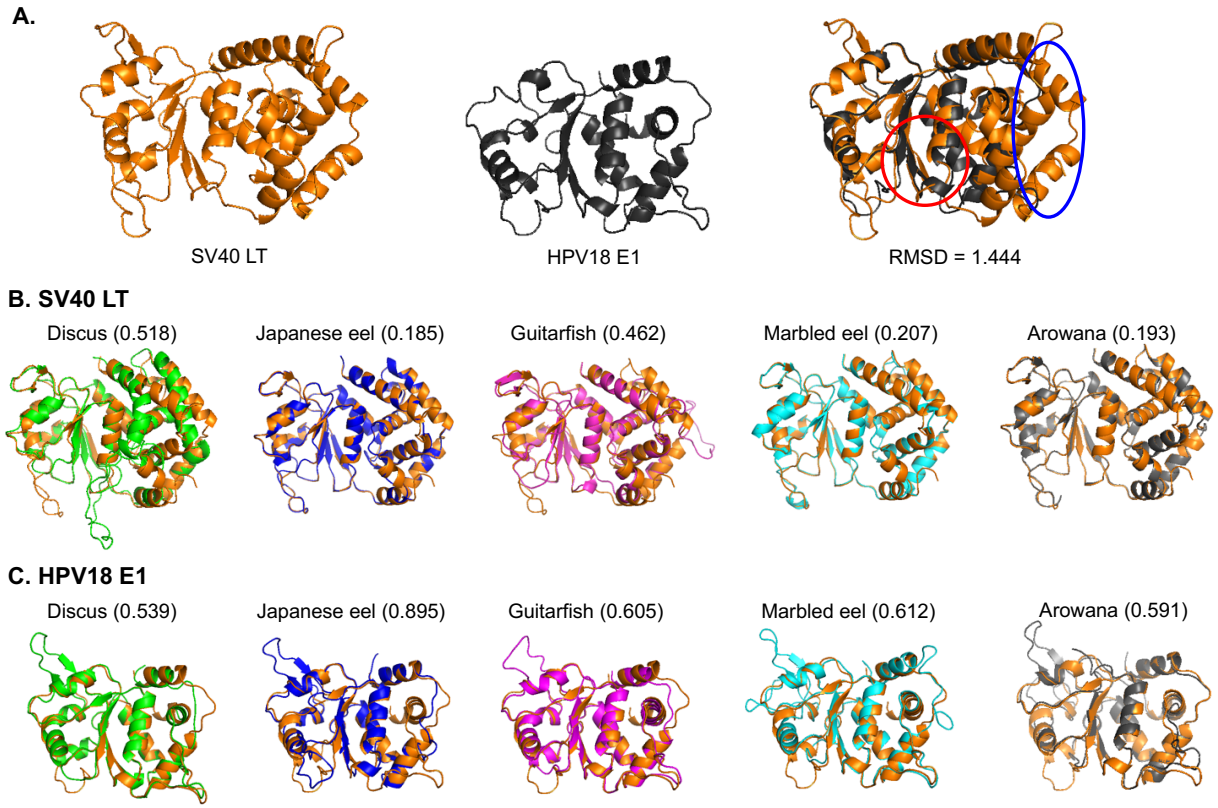

Figure 2. Structure modeling of the ATPase domains of adomaviruses. A) The template structures of SV40 LT and HPV18 E1, and the superposition of both structures. Red oval shows the well aligned Walker motifs. Blue oval highlights additional helices present only in SV40 LT. RMSD is listed at the bottom. B) Models of adomaviruses superposed with SV40 LT template. Template structure is shown in orange. The RMSD values for each comparison are listed in parentheses. C) Models of adomaviruses superposed with HPV18 E1 template. Template structure is shown in orange. The RMSD values for each comparison are listed in parentheses. (SV40, simian virus 40; HPV18, human papillomavirus type 18; Discus, Discus adomavirus; Japanese eel, Japanese eel adomavirus; Guitarfish, Guitarfish adomavirus; Marbled eel, Marbled eel adomavirus; Arowana, Arowana adomavirus; RMSD, root-mean-square deviation).

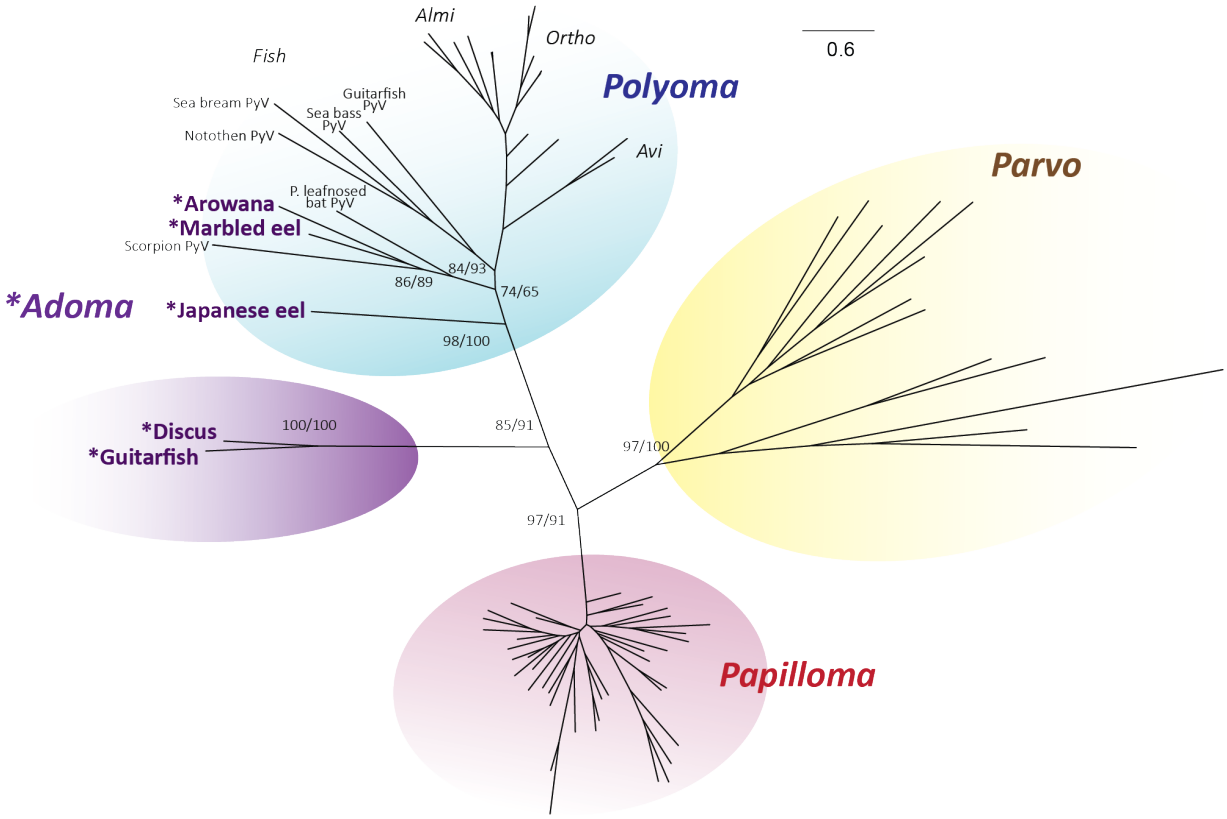

Figure 3. Phylogeny of adomavirus, papovavirus, and parvovirus SF3 helicase domains. Representatives of polyomavirus LT, papillomavirus E1, parovirus NS1/Rep, and the predicted helicase domains of the five adomavirus EO1 genes (asterisks) were analyzed using the maximum likelihood (ML) and Bayesian inference approach. The ML tree shown is midpoint rooted. Maximum likelihood bootstrap and Bayesian posterior probabilities support (ML/Bayesian) are indicated for major relevant nodes.

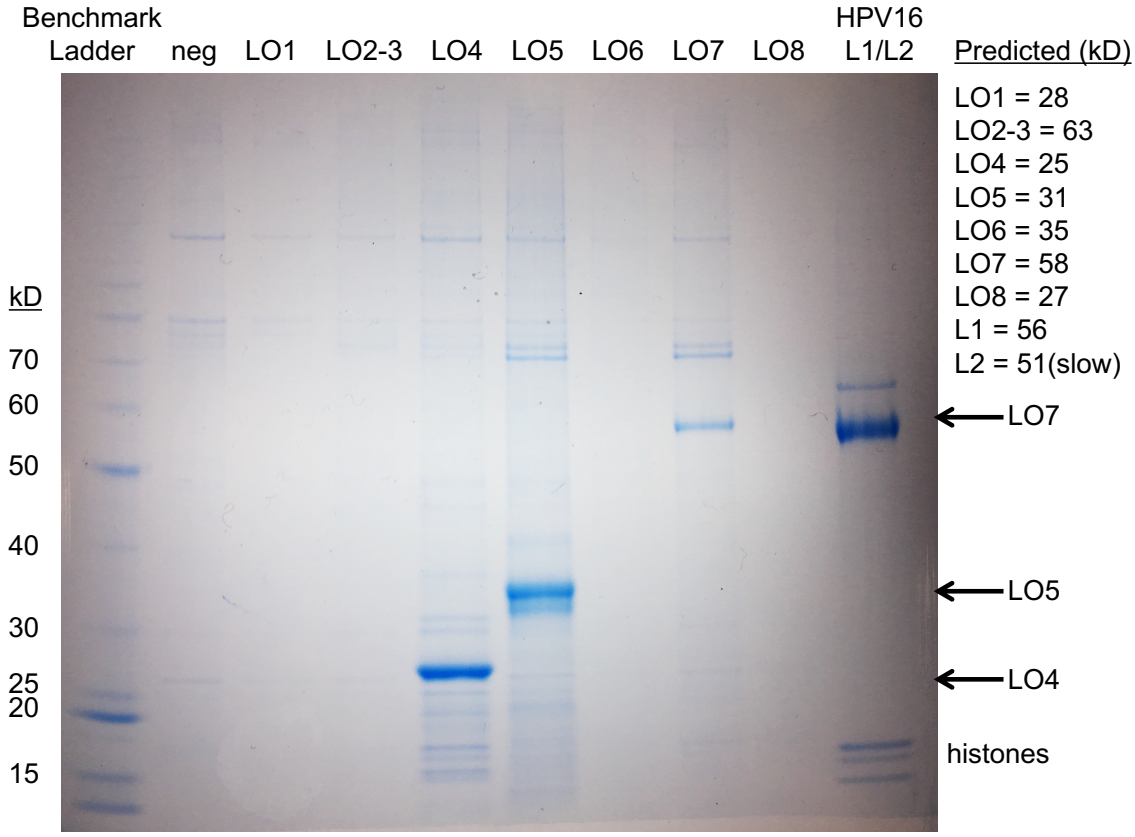

Figure 4. Particle assembly in mammalian cells transfected with individual LO expression constructs. 293TT cells were transfected with individual codon modified marbled eel LO expression plasmids indicated at the top of the image. The cells were lysed, subjected to nuclease digestion and a clarifying 5000 x g spin, and soluble material was ultracentrifuged through Optiprep gradients. Peak gradient fractions were subjected to SDS-PAGE analysis.

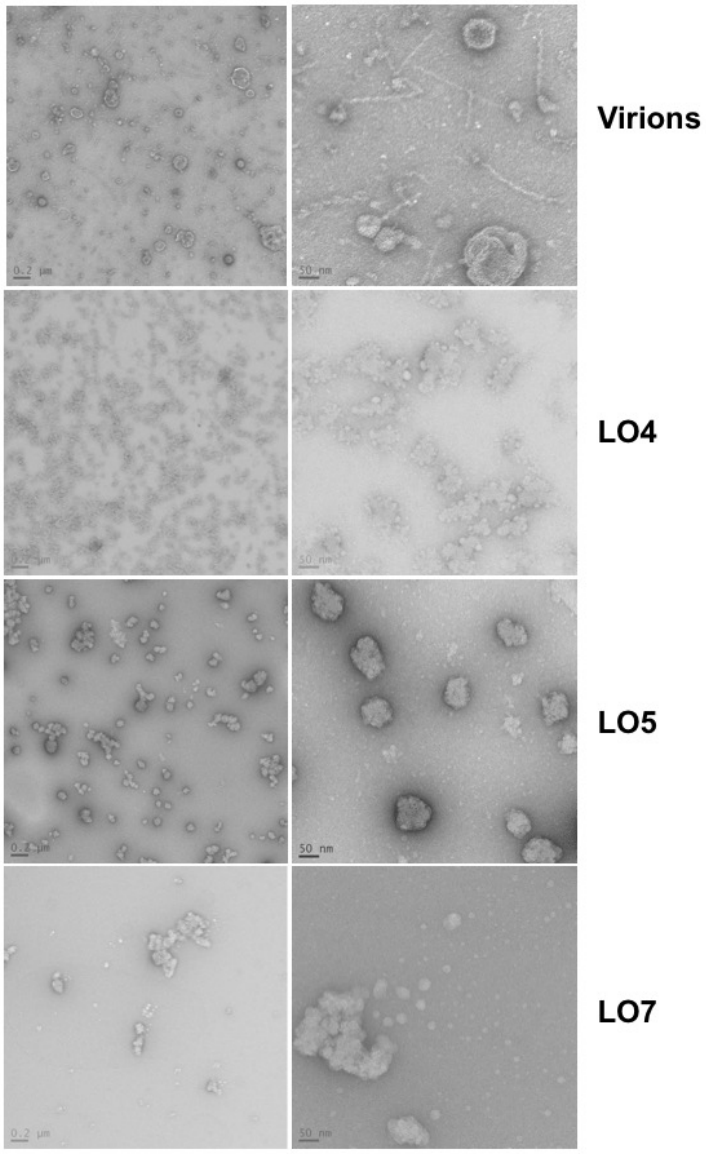

Figure 5. Negative stain EM of marbled eel adomavirus virions and particles formed after expression of single LO proteins in mammalian cells. Virions from EK-1 cells infected with marbled eel adomavirus were harvested by cell lysis followed by agarose gel filtration. 293TT cells were transfected with expression plasmids encoding the indicated marbled eel adomavirus proteins. Particles were extracted from cells by detergent lysis and purified by Optiprep. Scale bars are shown at the bottom left corner of each electron micrograph.

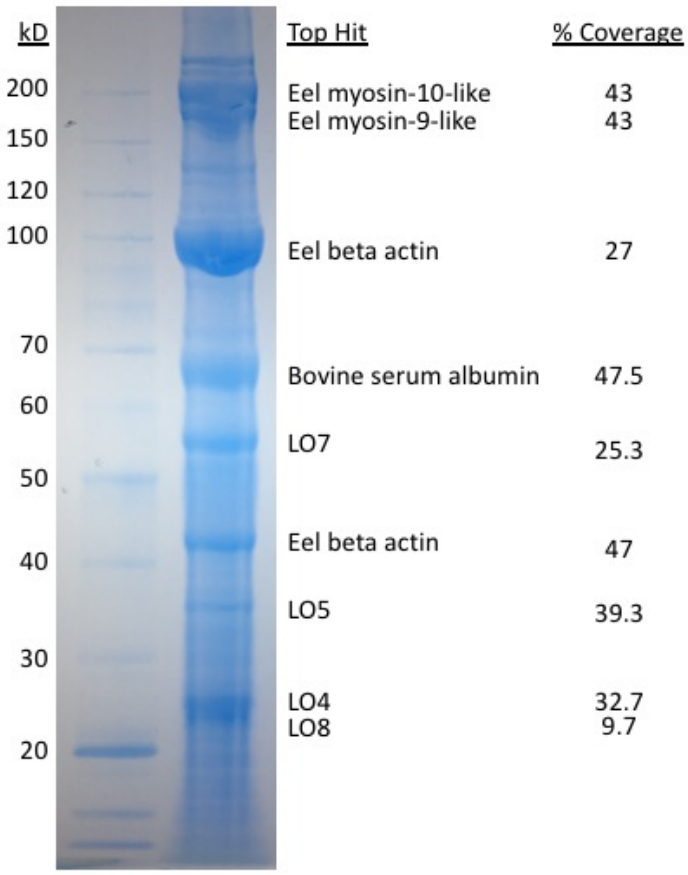

Figure 6. Mass spectrometric analysis of marbled eel adomavirus virions purified from infected EK-1 cells. Clarified lysates of infected cells were ultracentrifuged through an Optiprep gradient. Peak virion-containing fractions were selected then subjected to agarose gel filtration followed by TCA precipitation. The resulting precipitated sample was run on an SDS-PAGE gel. Thirteen gel bands were individually excised and processed on a Q Exactive HF Mass Spectrophotometer. Identification of peptides was conducted using Thermo Proteome Discoverer 2.2, Biopharma software.

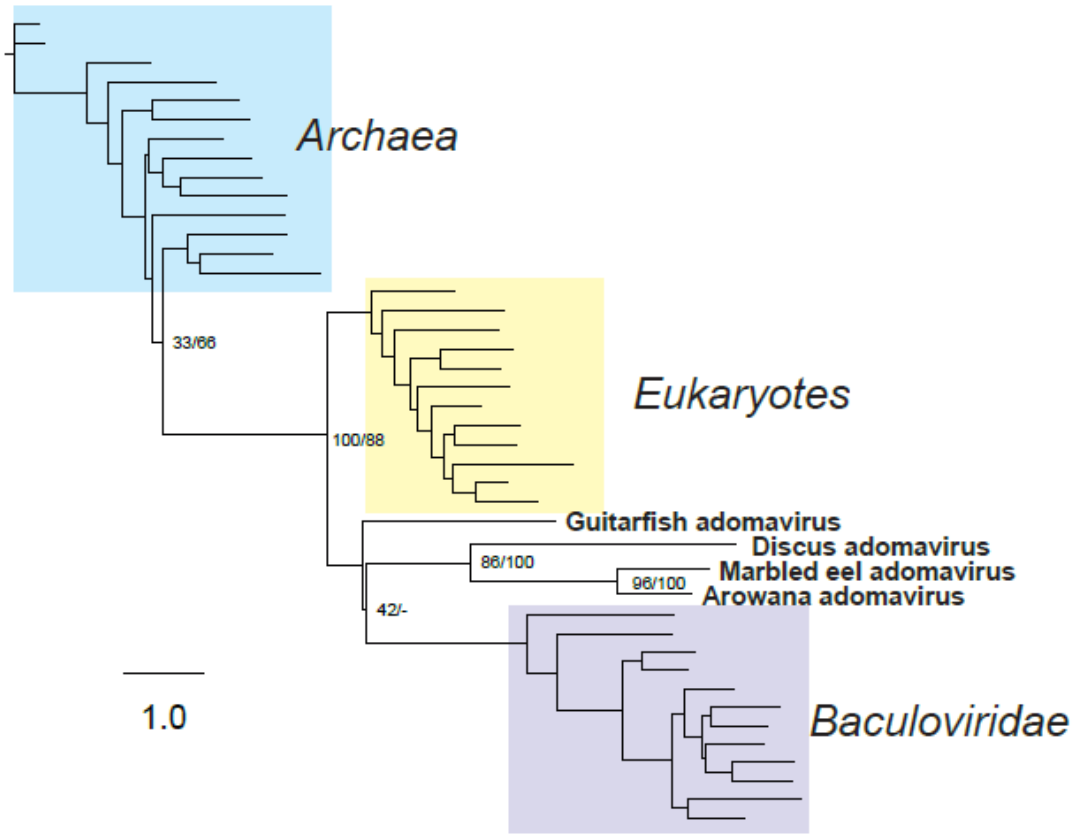

Figure 7. Phylogenetic analysis of the archaeo-eukaryotic primases in adomaviuses. The four adomavirus AEP domains were compared with representative AEPs from Eukaryotes, Archaea, and other viruses. Maximum likelihood bootstrap and Bayesian posterior probabilities (ML/Bayesian) are indicated for major relevant nodes. The ML tree shown is unrooted.

References

1. Buck CB & Ratner L (2015) Chapter 5: Oncogenic Viruses. *Cancer: Principles & Practice of Oncology*, eds DeVita VT, Lawrence TS, & Rosenberg SA (Lippincott Williams & Wilkins, Philadelphia), 10th Ddition Ed Vol 1, pp 69-82.
2. Buck CB*, et al.* (2016) The Ancient Evolutionary History of Polyomaviruses. *PLoS Pathog* 12(4):e1005574.
3. Van Doorslaer K (2013) Evolution of the papillomaviridae. *Virology* 445(1-2):11-20.
4. Lopez-Bueno A*, et al.* (2016) Concurrence of Iridovirus, Polyomavirus, and a Unique Member of a New Group of Fish Papillomaviruses in Lymphocystis Disease-Affected Gilthead Sea Bream. *J Virol* 90(19):8768-8779.
5. Sowd GA & Fanning E (2012) A wolf in sheep's clothing: SV40 co-opts host genome maintenance proteins to replicate viral DNA. *PLoS pathogens* 8(11):e1002994.
6. Koonin EV, Dolja VV, & Krupovic M (2015) Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* 479-480:2-25.
7. Hickman AB, Ronning DR, Kotin RM, & Dyda F (2002) Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol Cell* 10(2):327-337.
8. Iyer LM, Koonin EV, Leipe DD, & Aravind L (2005) Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res* 33(12):3875-3896.
9. Krupovic M & Koonin EV (2017) Multiple origins of viral capsid proteins from cellular ancestors. *Proc Natl Acad Sci U S A* 114(12):E2401-E2410.
10. Abrescia NG*, et al.* (2004) Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* 432(7013):68-74.
11. Nemerow GR, Stewart PL, & Reddy VS (2012) Structure of human adenovirus. *Curr Opin Virol* 2(2):115-121.
12. Jaalinoja HT*, et al.* (2008) Structure and host-cell interaction of SH1, a membrane-containing, halophilic euryarchaeal virus. *Proc Natl Acad Sci U S A* 105(23):8008-8013.
13. Demina TA*, et al.* (2016) Archaeal Haloarcula californiae Icosahedral Virus 1 Highlights Conserved Elements in Icosahedral Membrane-Containing DNA Viruses from Extreme Environments. *MBio* 7(4).
14. Mizutani T*, et al.* (2011) Novel DNA virus isolated from samples showing endothelial cell necrosis in the Japanese eel, Anguilla japonica. *Virology* 412(1):179-187.
15. Okazaki S*, et al.* (2016) Detection of Japanese eel endothelial cells-infecting virus in Anguilla japonica elvers. *J Vet Med Sci* 78(4):705-707.
16. Wen CM, Chen MM, Wang CS, Liu PC, & Nan FH (2015) Isolation of a novel polyomavirus, related to Japanese eel endothelial cell-infecting virus, from marbled eels, Anguilla marmorata (Quoy & Gaimard). *J Fish Dis*.
17. Bian C*, et al.* (2016) The Asian arowana (Scleropages formosus) genome provides new insights into the evolution of an early lineage of teleosts. *Sci Rep* 6:24501.
18. Logan J & Shenk T (1984) Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. *Proceedings of the National Academy of Sciences of the United States of America* 81(12):3655-3659.
19. An P, Saenz Robles MT, & Pipas JM (2012) Large T antigens of polyomaviruses: amazing molecular machines. *Annu Rev Microbiol* 66:213-236.

20. Carter JJ, *et al.* (2013) Identification of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes. *Proc Natl Acad Sci U S A* 110(31):12744-12749.
21. Moyer CL, Besser ES, & Nemerow GR (2015) A Single Maturation Cleavage Site in Adenovirus Impacts Cell Entry and Capsid Assembly. *J Virol* 90(1):521-532.
22. Mangel WF & San Martin C (2014) Structure, function and dynamics in adenovirus maturation. *Viruses* 6(11):4536-4570.
23. Alva V, Nam SZ, Soding J, & Lupas AN (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res* 44(W1):W410-415.
24. McGrath WJ, Ding J, Didwania A, Sweet RM, & Mangel WF (2003) Crystallographic structure at 1.6-A resolution of the human adenovirus proteinase in a covalent complex with its 11-amino-acid peptide cofactor: insights on a new fold. *Biochim Biophys Acta* 1648(1-2):1-11.
25. Zimmermann L, *et al.* (2017) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* S0022-2836(17):30587-30589.
26. Cserzo M, Wallin E, Simon I, von Heijne G, & Elofsson A (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* 10(6):673-676.
27. DiGiuseppe S, *et al.* (2015) Topography of the Human Papillomavirus Minor Capsid Protein L2 during Vesicular Trafficking of Infectious Entry. *J Virol* 89(20):10442-10452.
28. Aydin I, *et al.* (2017) A central region in the minor capsid protein of papillomaviruses facilitates viral genome tethering and membrane penetration for mitotic nuclear entry. *PLoS pathogens* 13(5):e1006308.
29. Grande A, Costas C, & Benavente J (2002) Subunit composition and conformational stability of the oligomeric form of the avian reovirus cell-attachment protein sigmaC. *J Gen Virol* 83(Pt 1):131-139.
30. Goldenberg D, *et al.* (2010) Genetic and antigenic characterization of sigma C protein from avian reovirus. *Avian Pathol* 39(3):189-199.
31. Tsuruta Y, *et al.* (2007) A mosaic fiber adenovirus serotype 5 vector containing reovirus sigma 1 and adenovirus serotype 3 knob fibers increases transduction in an ovarian cancer ex vivo system via a coxsackie and adenovirus receptor-independent pathway. *Clin Cancer Res* 13(9):2777-2783.
32. Camus A, Dill J, McDermott A, Camus M, & Fan Ng TF (2016) Virus-associated papillomatous skin lesions in a giant guitarfish Rhynchobatus djiddensis: a case report. *Diseases of aquatic organisms* 117(3):253-258.
33. Vragniau C, *et al.* (2017) Studies on the Interaction of Tumor-Derived HD5 Alpha Defensins with Adenoviruses and Implications for Oncolytic Adenovirus Therapy. *J Virol* 91(6).
34. Buck CB, Pastrana DV, Lowy DR, & Schiller JT (2004) Efficient intracellular assembly of papillomaviral vectors. *J Virol* 78(2):751-757.
35. Buck CB & Thompson CD (2007) Production of papillomavirus-based gene transfer vectors. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al* Chapter 26:Unit 26.21.
36. Pastrana DV, *et al.* (2009) Quantitation of human seroresponsiveness to Merkel cell polyomavirus. *PLoS pathogens* 5(9):e1000578.

37. Buck CB, *et al.* (2008) Arrangement of L2 within the papillomavirus capsid. *J Virol* 82(11):5190-5197.
38. Krupovic M & Koonin EV (2015) Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat Rev Microbiol* 13(2):105-115.
39. Yutin N, Shevchenko S, Kapitonov V, Krupovic M, & Koonin EV (2015) A novel group of diverse Polinton-like viruses discovered by metagenome analysis. *BMC Biol* 13:95.
40. Nwasike C, Ewert S, Jovanovic S, Haider S, & Mujtaba S (2016) SET domain-mediated lysine methylation in lower organisms regulates growth and transcription in hosts. *Ann N Y Acad Sci* 1376(1):18-28.
41. Koonin EV & Yutin N (2010) Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. *Intervirology* 53(5):284-292.
42. Woolford L, *et al.* (2007) A novel virus detected in papillomas and carcinomas of the endangered western barred bandicoot (Perameles bougainville) exhibits genomic features of both the Papillomaviridae and Polyomaviridae. *J Virol* 81(24):13280-13290.
43. Bennett MD, *et al.* (2010) The first complete papillomavirus genome characterized from a marsupial host: a novel isolate from Bettongia penicillata. *J Virol* 84(10):5448-5453.
44. Krupovic M & Koonin EV (2014) Evolution of eukaryotic single-stranded DNA viruses of the Bidnaviridae family from genes of four other groups of widely different viruses. *Sci Rep* 4:5347.
45. Buck CB, Thompson CD, Pang YY, Lowy DR, & Schiller JT (2005) Maturation of papillomavirus capsids. *J Virol* 79(5):2839-2846.
46. Cerqueira C, *et al.* (2015) A Cell-Free Assembly System for Generating Infectious Human Papillomavirus 16 Capsids Implicates a Size Discrimination Mechanism for Preferential Viral Genome Packaging. *J Virol* 90(2):1096-1107.
47. Cerqueira C & Schiller JT (2017) Papillomavirus assembly: An overview and perspectives. *Virus Res* 231:103-107.
48. Cardone G, *et al.* (2014) Maturation of the human papillomavirus 16 capsid. *MBio* 5(4):e01104-01114.
49. Mattila S, Oksanen HM, & Bamford JKH (2015) Probing protein interactions in the membrane- containing virus PRD1. *J Gen Virol* 96(Pt 2):453-462.
50. Dill JA, Ng TF, & Camus AC (2016) Complete Sequence of the Smallest Polyomavirus Genome, Giant Guitarfish (Rhynchobatus djiddensis) Polyomavirus 1. *Genome Announc* 4(3).
51. Ng TF, *et al.* (2013) Distinct lineage of vesiculovirus from big brown bats, United States. *Emerg Infect Dis* 19(12):1978-1980.
52. Ng TF, *et al.* (2015) A metagenomics and case-control study to identify viruses associated with bovine respiratory disease. *J Virol* 89(10):5340-5349.
53. Ng TF, *et al.* (2012) High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol* 86(22):12161-12175.
54. Victoria JG, Kapoor A, Dupuis K, Schnurr DP, & Delwart EL (2008) Rapid identification of known and new RNA viruses from animal tissues. *PLoS pathogens* 4(9):e1000163.
55. Peretti A, FitzGerald PC, Bliskovsky V, Buck CB, & Pastrana DV (2015) Hamburger polyomaviruses. *J Gen Virol* 96(Pt 4):833-839.

56. Deng X, *et al.* (2015) An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. *Nucleic Acids Res* 43(7):e46.
57. Bankevich A, *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology* 19(5):455-477.
58. Pastrana DV, *et al.* (2004) Reactivity of human sera in a sensitive, high-throughput pseudovirus-based papillomavirus neutralization assay for HPV16 and HPV18. *Virology* 321(2):205-216.
59. Ritter B, *et al.* (2004) Two WXXF-based motifs in NECAPs define the specificity of accessory protein binding to AP-1 and AP-2. *EMBO J* 23(19):3701-3710.
60. Buchan DW, Minneci F, Nugent TC, Bryson K, & Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 41(Web Server issue):W349-357.
61. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797.
62. Guindon S, *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* 59(3):307-321.
63. Lefort V, Longueville JE, & Gascuel O (2017) SMS: Smart Model Selection in PhyML. *Molecular biology and evolution* 34(9):2422-2424.
64. Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* 17(8):754-755.
65. Muhire BM, Varsani A, & Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9(9):e108277.