Genomic analysis of hyperparasitic viruses associated with entomopoxviruses

Zachary K Barth, Ian Hicklin, Julien Thézé, Jun Takatsuka, Madoka Nakai, Elisabeth A Herniou, Anne M Brown, Frank O Aylward

VIRUS EVOLUTION

Virus Evolution, 2024, **10(1)**, veae051

DOI: https://doi.org/10.1093/ve/veae051
Advance Access Publication 12 July 2024
Research Article

Genomic analysis of hyperparasitic viruses associated with entomopoxviruses

Zachary K. Barth[1,*], Ian Hicklin[2], Julien Thézé[3], Jun Takatsuka[4], Madoka Nakai[5], Elisabeth A. Herniou[6], Anne M. Brown[2,7,8], Frank O. Aylward[1,7,*]

[1]Department of Biological Sciences, Virginia Tech, 926 West Campus Drive, Blacksburg, VA 24061, USA
[2]Department of Biochemistry, Virginia Tech, Blacksburg, VA 24061, USA
[3]Université Clermont Auvergne, INRAE, VetAgro Sup, UMR EPIA, Saint-Genès-Champanelle, France
[4]Forestry and Forest Products Research Institute, Matsunosato, Tsukuba, Ibaraki 305-8687, Japan
[5]Institute of Agriculture, Tokyo University of Agriculture and Technology, Saiwai, Fuchu, Tokyo 183-8509, Japan
[6]Institut de Recherche sur la Biologie de l'Insecte, UMR7261 CNRS-Université de Tours, 20 Avenue Monge, Parc de Grandmont, Tours 37200, France
[7]Center for Emerging, Zoonotic, and Arthropod-Borne Pathogens, Virginia Tech, 1981 Kraft Dr, Blacksburg, VA 24061, USA
[8]Research and Informatics, University Libraries, Virginia Tech, Blacksburg, VA 24061, USA

*Corresponding author. Department of Biological Sciences, Virginia Tech, 926 West Campus Drive, Blacksburg, VA 24061, USA. E-mail: zbarth@vt.edu; Center for Emerging, Zoonotic, and Arthropod-Borne Pathogens, Virginia Tech, 926 West Campus Drive, Blacksburg, VA 24061, USA. E-mail: faylward@vt.edu

Abstract

Polinton-like viruses (PLVs) are a diverse group of small integrative dsDNA viruses that infect diverse eukaryotic hosts. Many PLVs are hypothesized to parasitize viruses in the phylum *Nucleocytoviricota* for their own propagation and spread. Here, we analyze the genomes of novel PLVs associated with the occlusion bodies of entomopoxvirus (EPV) infections of two separate lepidopteran hosts. The presence of these elements within EPV occlusion bodies suggests that they are the first known hyperparasites of poxviruses. We find that these PLVs belong to two distinct lineages that are highly diverged from known PLVs. These PLVs possess mosaic genomes, and some essential genes share homology with mobile genes within EPVs. Based on this homology and observed PLV mosaicism, we propose a mechanism to explain the turnover of PLV replication and integration genes.

Keywords entomopoxvirus; polinton-like virus; viral satellite; virophage.

Introduction

Viral satellites are genetic hyperparasites, i.e. parasites of other parasites, that rely on host or "helper" viruses for their own propagation and spread (Gnanasekaran and Chakraborty 2018). Viral satellites are found throughout the tree of life, associating with viruses that infect each domain (Ren et al. 2013, Barreat and Katzourakis 2023, de Sousa et al. 2023). Within eukaryotes, these elements have been described in diverse hosts, including humans (Taylor 2012, Meier et al. 2020), bees (Olivier et al. 2008), yeast (Vepštaitė-Monstavičė et al. 2018), amoebae (La Scola et al. 2008), algae (Santini et al. 2013), and many plant species (Fritsch and Mayo 2018).

Historically, recognized eukaryotic satellites have been small entities, with RNA or single stranded DNA genomes less than 6kb and few to no coding sequences (Gnanasekaran and Chakraborty 2018). This trend has been challenged in the recent years with the discovery of hyperparasitic dsDNA viruses that integrate into eukaryotic cell chromosomes and have genomes in the range of ~15–25kb. These viruses are referred to as "virophages" because of their lifestyle where they "infect" and parasitize the viral factories of nucleocytoviruses (phylum *Nucleocytoviricota*) (Fischer 2021). To date, culture systems have been established for five of these virophages (La Scola et al. 2008, Fischer and Suttle 2011, Gaia et al. 2014, Rolf et al. 2018, Sheng et al. 2022, Roitman et al. 2023). These viruses are classified within the phylum *Preplasmiviricota* and class *Maveriviricetes* (Fischer 2021). A culture system has also been established for a satellite virus that has virophage-like life cycle, but is not a member of the *Maveriviricetes*. This element, PgVV Gezel 14T, parasitizes Phaeocystis globose virus (Roitman et al. 2023) and belongs to a large group of polinton-like viruses (PLVs) (Bellas et al. 2023).

PLVs are capsid-encoding elements related to the maverick-polinton class of self-synthesizing transposable elements (Koonin and Krupovic 2017). The distinction between virophages, polintons, and PLVs is sometimes blurry because of similarities in genome composition and the precise meaning of these terms continues to evolve as more diverse elements continue to be discovered. Initially classified as self-synthesizing transposons, polintons are typically defined by the presence of a *polB* gene, an *rve* integrase, and terminal inverted repeats, but the vast majority

of polinton elements also possess identifiable major capsid proteins (MCPs). The presence of MCPs suggests that polintons are in fact viruses, and a taxonomic class, *Polintoviricetes*, has been established for these putative viruses. Recently, it has been suggested that most, if not all PLVs, should be grouped within *Polintoviricetes* (Roux et al. 2023). Hereafter, we will use "virophage" to refer to members of the *Maveriviricetes*, and our use of "PLV" refers to accepted and prospective members of the *Polintoviricetes*, including those polintons that encode a MCP.

Aside from PgVV Gezel 14T, a single PLV, TsV-N1, has been experimentally characterized. TsV-N1 was found to cause autonomous lytic infection in the alga *Tetraselmis striata* (Pagarete et al. 2015). Three other PLVs were also co-isolated with the Chrysochromulina parva nucleocytovirus CpV-BQ2 (Stough et al. 2019), and there has been a report of mobilization and replication of two polintons in lepidopteran cell lines used to produce recombinant baculovirus virus-like particles (Starrett et al. 2021), suggesting that a virophage lifestyle may be a common adaptation for polintons and PLVs. Despite the paucity of experimental PLV systems, PLV genomes are abundant in metagenomic samples and as integrated proviruses within cellular genomes (Barreat and Katzourakis 2021, Bellas and Sommaruga 2021, Bellas et al. 2023, Roitman et al. 2023, Barreat and Katzourakis 2021, Bellas and Sommaruga 2021, Bellas et al. 2023); PLVs are found in diverse eukaryotic genomes, but are notably absent in those of mammals and land plants (Barreat and Katzourakis 2021). While some PLVs, including TsV-N1, have genomes in the 30–40kb range, overlapping with the autonomous adenoviruses, many are closer in size to virophages. In addition to the small size of many PLVs, similar gene content to virophages has been noted (Fischer and Suttle 2011), leading some to suggest that PLVs may retain antagonism toward nucleocytoviruses (Barreat and Katzourakis 2021).

Here, we have analyzed the genomes of three novel elements that we tentatively refer to as PLVs, which were sequenced from entomopoxvirus-derived occlusion bodies formed during infection of two lepidopteran host species (Thézé et al. 2013). The association of these PLVs with entomopoxvirus occlusion bodies provides compelling evidence for these elements possessing a hyperparasitic lifestyle, and is the first indication of a virophage-like element associated with poxviruses. Aligned with this hypothesis, we observe evidence of horizontal gene transfer between these PLVs and entomopoxviruses. Among these entomopoxvirus-associated PLVs, we see two distinct lineages based on phylogenies of structural gene modules. These lineages encode divergent replication modules and possess capsid genes highly diverged from other PLV lineages, demonstrating previously unrecognized diversity amongst eukaryotic small dsDNA viruses. We also observe replicon gene turnover amongst related elements and propose a mechanism to explain these patterns.

Results and discussion

Lepidopteran entomopoxvirus-associated elements resemble polinton-like viruses

A previous study provided sequencing and assembly of four entomopoxviruses (EPVs) of the Betaentomopoxvirus genus infecting the Lepidoptera: Adoxophyes honmai EPV (AHEV), Choristoneura biennis EPV (CBEV), Choristoneura rosaceana EPV (CREV), and Mythimna separata EPV (MySEV) (Thézé et al. 2013). Additional analyses of unplaced contigs have revealed three contigs ~12kb long that were co-sequenced from the same occlusion bodies as AHEV and CBEV (Supplementary Table S1). Here, we have analyzed these three contigs and found that they are likely complete genomes of viruses in the phylum Preplasmaviricota. We tentatively refer to them as PLVs, consistent with terminology that has recently been used (Bellas and Sommaruga 2021, Bellas et al. 2023). We have opted to rename them according to AHEV_PLV1, AHEV_PLV2, and CBEV_PLV1. We searched for additional related elements by performing basic local alignment search tool (BLAST) searches of the element's putative capsid genes against the National Center for Biotechnology Information (NCBI) database. We performed blastn, blastp, and tblastn searches of the putative capsid sequences, and only the tblastn search was able to produce an additional homolog using the AHEV_PLV2 sequence as a query. This homolog was located on the ninth chromosome of *Cetonia aurata*, and we were able to manually annotate the element this homolog belonged to. We named this additional element Ca_PLV1. All of the elements contained detectable terminal repeats, indicating that they are complete.

We were able to use HHpred (Gabler et al. 2020) to predict several gene functions for the novel PLVs. We found that the four novel PLVs encoded replication machinery characteristic of PLVs: AHEV_PLV1 and Ca_PLV1 each encode a family B DNA polymerase (PolB), while AHEV_PLV2 and CBEV_PLV1 encode a gene with a C-terminal D5-like SF3 helicase domain (Fig. 1, Supplementary Table S2). While HHpred was not able to detect a primase domain for these genes, we hypothesize that they are primase-helicases with a primase-polymerase domain, as this arrangement is common in PLV genomes (Yutin et al. 2015), and a primase-polymerase domain would presumably be necessary to support replication. In many PLVs (Yutin et al. 2013, 2015, Koonin and Krupovic 2017), the primase helicases are accompanied by tyrosine-recombinases, and this is the case for the AHEV_PLV2 and CBEV_PLV1 helicases.

The novel elements also encoded genes associated with morphogenesis. All four elements encoded an Adenovirus-type cysteine-protease, and the three lepidopteran elements encoded a protein similar to the phospholipase domain of parvovirus virion protein VP1 (Fig. 1). These gene classes have been described previously in PLVs and virophages (Yutin et al. 2013, 2015, Koonin and Krupovic 2017). In a few cases, HHpred returned conflicting or inconclusive results for proteins that appeared to be homologous between PLVs. We grouped these homologous protein sequences for multiple sequence alignment (MSA) generation (Supplementary Table S4) and input our MSAs into HHpred. This approach revealed the presence of an A32-like packaging ATPase (Supplementary Table S5) and an Adenovirus L2 mu core-like protein (Supplementary Table S6) in all four PLVs. While A32-like packaging ATPases are conserved components of PLV genomes, to our knowledge, Adenovirus L2 mu core-like proteins have not been previously associated with PLVs. While the precise role of this adenovirus protein is unknown, it may be involved in viral chromatin condensation for packaging (Anderson et al. 1989). It has been suggested that PLVs may comprise a sister taxa to adenoviruses due to the relatedness of their shared gene content (Barreat and Katzourakis 2023). The L2 mu core protein provides yet another example of gene classes shared between both groups.

Aside from PolB, one of the signature genes of the maverick-polinton group of PLVs is a retroviral-type integrase (rve) gene. An example of this can be seen with SfMaverick1 and 2, from the lepidopteran species *Spodoptera frugiperda* (Fig. 1). Curiously, we were unable to detect any known integration genes within AHEV_PLV1 and Ca_PLV1. As AHEV_PLV1 was found in occlusion bodies, it may be non-integrative; however, Ca_PLV1 was found integrated within the *C. aurata* genome. It is possible that these elements possess cryptic integration functions unknown to current gene annota-

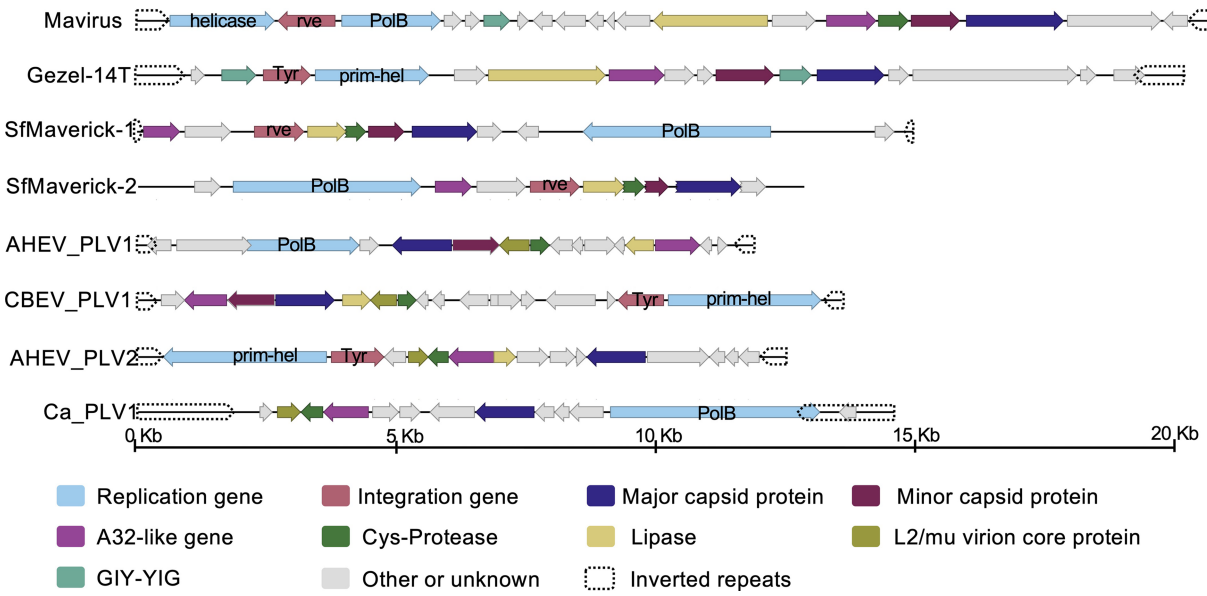

Figure 1. Entomopoxvirus-associated elements possess gene content characteristic of PLVs. Diagrams of several PLVs and the virophage Mavirus. For AHEV_PLV1, AHEV_PLV2, and CBEV_PLV1, sequences are depicted as beginning with one of their inverted repeat sequences to illustrate organizational similarities to other elements. Functional gene classes and inverted repeats are denoted by color. Replication and integration genes are labeled. "Tyr" is tyrosine recombinase. "Prim-hel" is primase-helicase.

tions, or they may rely on integration genes encoded *in trans*. Regardless, the lack of a detectable integrase does not appear to preclude the possibility of integration by these elements.

Entomopoxvirus-associated polinton-like viruses have highly divergent capsid proteins

Surprisingly, our initial HHpred did not detect capsid proteins within the four novel elements (Supplementary Table S2). On the basis of gene length and PLV organization, we formulated a list of candidate capsid genes and BLAST searched these against the IMG/VR database (Supplementary Table S3). We were able to identify homologs for two sets of homologous proteins shared between AHEV_PLV1 and CBEV_PLV1, and one set of homologous proteins shared between AHEV_PLV2 and Ca_PLV1. We aligned these protein sets into MSAs and predicted their function using HHpred (Supplementary Tables S7–S9). This confirmed the identity of major capsid and minor capsid lineages shared between AHEV_PLV1 and CBEV_PLV1, and a major capsid lineage shared between AHEV_PLV2 and Ca_PLV1. We generated predicted structures of our aligned sequences using Colabfold (Mirdita et al. 2022), which provided expected double jelly roll (DJR) fold and single jelly roll (SJR) fold structures characteristic of PLV capsid proteins (Fig. 2a–c, Supplementary Figure S1) (Krupovic et al. 2014). Searching our structures against the PDB 100 database using Foldseek further supported these proteins as PLV capsids (Supplementary Tables S10–S12).

We were unable to confidently identify minor capsid (mCP) proteins for AHEV_PLV2 and Ca_PLV1. The two PLVs have relatively large coding sequences of unknown function located next to their MCPs (that we suspect may be their mCPs). These genes possess 35% aa similarity across approximately180 aa but are otherwise distinct from each other. We were unable to find other homologs of these proteins or to generate high-quality structural predictions. From individual peptide sequences, Colabfold generated structures with SJR folds, but with low confidence in the SJR regions (Supplementary Figure S1). Although this suggests these proteins could be mCPs, it is too premature to designate them as such.

Having identified the MCPs of the four new PLV elements, we next sought to examine how they are related to other PLV capsid groups. We built a phylogeny that included the EPV-associated PLV MCPs as well as MCPs from Mavpol elements and MCPs belonging to a large cluster related to PgVV MCPs that had been previously identified (Yutin et al. 2015, Bellas et al. 2023). These clusters provide a considerable amount of breadth of cellular host species and cover nearly all PLVs that have been found in animal species. The AHEV_PLV1 and CBEV_PLV1 MCPs appear to be diverged members of the mavpol group, placing next to the mavpol 2 clade (Fig. 2d). The AHEV_PLV2 and Ca_PLV1 MCPs did not belong to any established clades, instead grouping next to the PgVV group, which contains Gezel-14T, and next to a small clade consisting of Polinton-3_NV and Polinton-N1A_NV. Despite their designation as polintons, Polinton-3_NV and Polinton-N1A_NV possess MCPs distinct from the mavpol MCP lineage (Yutin et al. 2015, Bellas and Sommaruga 2021, Bellas et al. 2023). The branching of AHEV_PLV2 and Ca_PLV1 with Polintons 3_NV and N1A_NV, and the PgVV was of intermediate confidence, having an ultrafast bootstrap support value of 89. The divergent placement of EPV-associated PLVs highlights the potential wealth of PLV diversity that remains undiscovered.

Entomopoxvirus-associated polinton-like viruses are mosaic elements

Outside of the MCP, the strongest phylogenetic markers for PLVs are their replication genes. Amongst the two distinct capsid lineages of EPV-associated PLVs, there are single members of each lineage (AHEV_PLV1 and Ca_PLV1, respectively) that encode a PolB gene. We inferred the phylogeny of PolB sequences from PLVs and virophages to evaluate if this gene would display similar interrelationships between the EPV PLVs as the MCPs. Overall, most PolB sequences grouped within four large clades, and these clades do not cluster discretely with separate capsid lineages (Fig. 3a). Polintons of the mavpol2 capsid lineage are restricted to one of the major PolB clades, while polintons of the mavpol1 lineage are found within two major clades, and virophages possess polBs of all

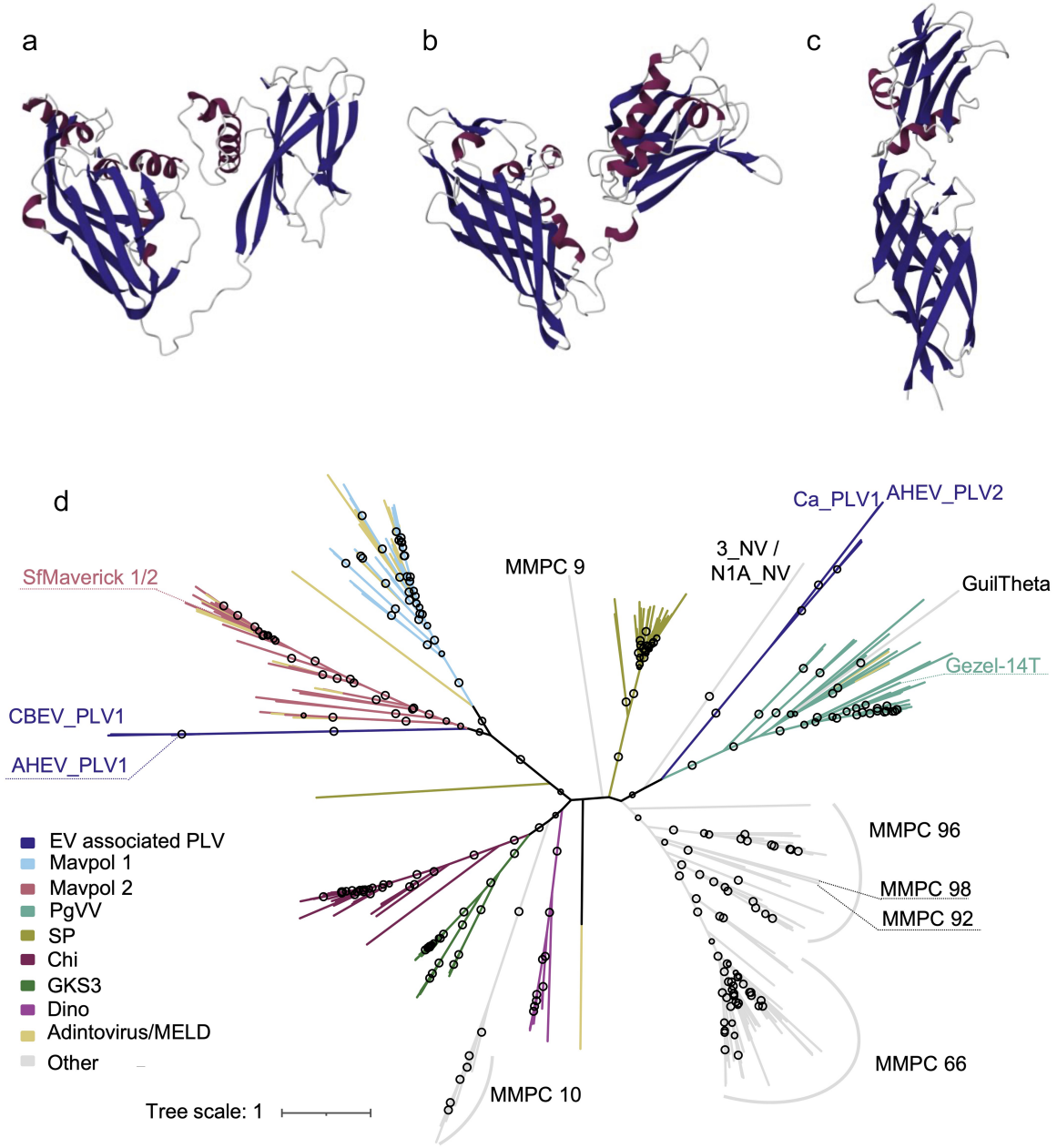

Figure 2. EPV-associated PLVs encode MCPs belonging to two separate lineages. (a) through (c) predicted protein structures for alignments of proteins including and related to the AHEV_PLV1 and CBEV_PLV1 putative MCP (a), the AHEV_PLV2 and Ca_PLV1 putative MCP (b), and the AHEV_PLV1 and CBEV_PLV1 putative mCP (c). Beta sheets are colored dark blue and helices are colored purple. For each model, alignments are available in Supplementary Data X, and a full list of aligned proteins for each model is available in Supplementary Table X. D. A phylogeny of major capsid protein genes amongst PLVs. Several major lineages of PLVs are denoted by color, and additional MCP clusters are labeled based on classification by Bellas and Sommaruga (2021) and Bellas et al. (2023). PLVs from Fig. 1 are also labeled.

four clades (Fig. 3a). The AHEV_PLV1 and Ca_PLV1 PolBs occurred in separate major clades. The closest relatives of the AHEV_PLV1 PolB were found in virophages, and also clustered with a clade of polintons from *Trichomonas vaginalis* for which we did not detect capsid proteins (Fig. 3a). The Ca_PLV1 PolB placed within a clade containing elements of the mavpol1 and mavpol2 capsid-lineage, both of which are known to occur within insects. In addition to our tree of curated PolBs from known polinton and virophage elements, we also performed a BLAST search against proteins in NCBI GenBank. In addition to the *T. vaginalis* polintons, the AHEV_PLV1 PolB also hit to proteins in *Entamoeba, Girardia, Oomycota*, and *Hexamita*, with 26–31% shared identity across >80% of the query sequence for these taxa (Supplementary Table S13), suggesting a taxonomic affinity with protist-encoded elements despite AHEV_PLV1 encoding an MCP related to animal polintons. Conversely, the Ca_PLV1 PolB was found to have a close homolog (61% identity across 82% of the query sequence) in an unplaced contig from *Tribolium castaneum* (Supplementary Table S13), a model coleopteran species known to harbor multiple mavpols (Haapa-Paananen et al. 2014). Other homologs were found in a coral, a bivalve, leafhoppers, and several spider species (Supplementary Table S13), despite the Ca_PLV1 MCP appearing unrelated to known animal-tropic elements other than AHEV_PLV2. These results demonstrate that the MCPs and PolBs of the EPV-associated

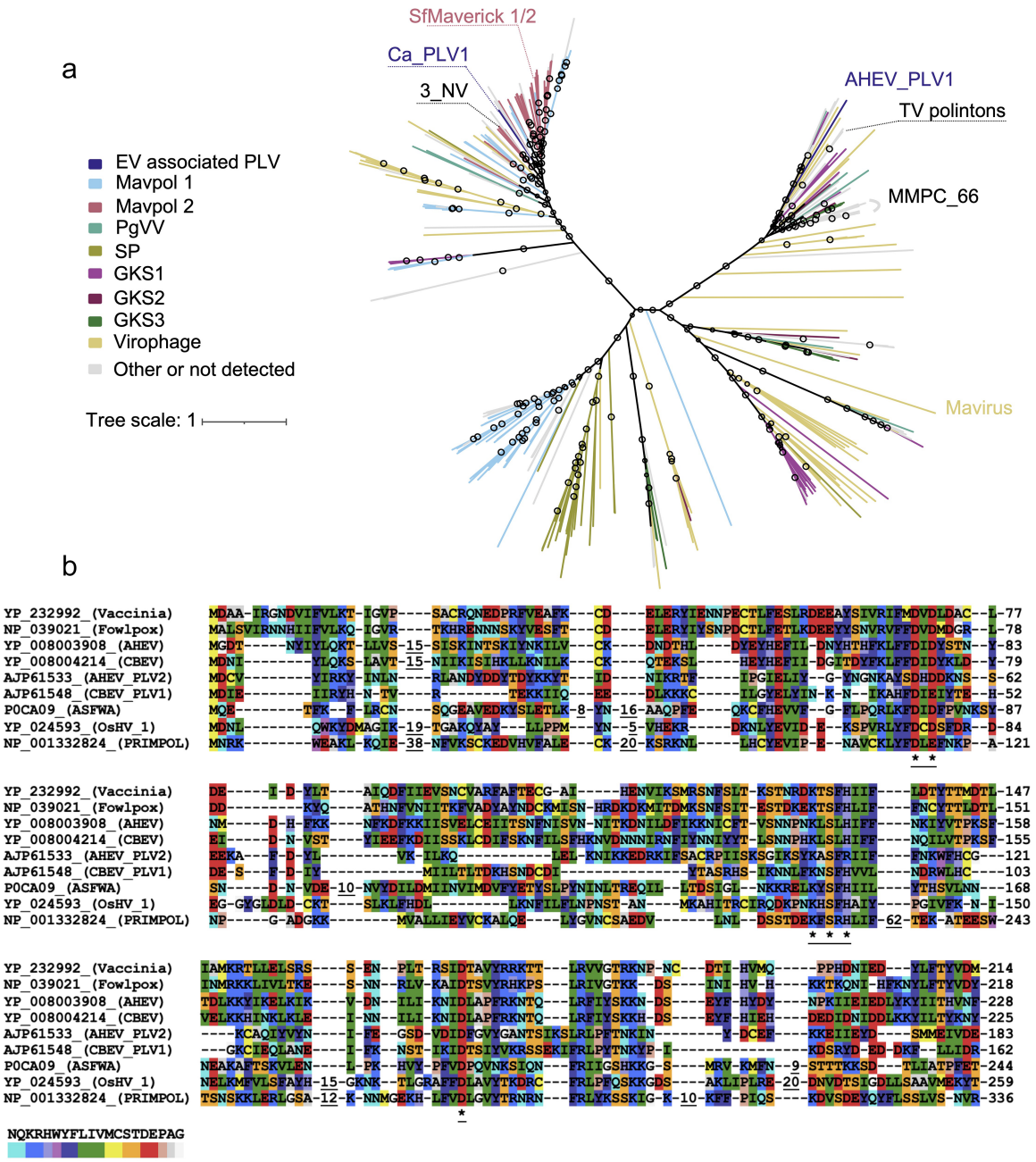

Figure 3. Entomopoxvirus-associated PLVs encode diverse replication machinery. (a) A phylogeny of PolB proteins encoded by polintons, PLVs, and virophages. Sequences are colored by their MCP lineage. For elements previously denoted as adintovirus or MELD virus, genomes are colored based on their MCP grouping in Fig. 2d. "TV polintons" refers to polintons 3, 4, and 5 from *Trichomonas vaginalis*. (b) A multiple sequence alignment of the primase domains from AHEV_PLV2 and CBEV_PLV1 primase-helicase proteins, the D5 proteins from Vaccinia virus, Fowlpox virus, *Adoxophyes honmai* entomopoxvirus, and *Choristoneura biennis* entomopoxvirus, putative primases from an African swine fever Warthog isolate (ASFWA) and Ostreid Herpesvirus-1 (OsHV_1), as well as the human primase-polymerase (PRIMPOL). Amino acids are highlighted in color based on their physical properties. Numbers at the end of lines denote the final amino acid position at the end of that line. Numbers within gaps represent the number of unaligned residues within that gap.

PLVs have contrasting evolutionary histories indicative of genomic mosaicism.

We also wanted to investigate the replication genes for AHEV_PLV2 and CBEV_PLV1. HHpred had detected large genes with c-terminal D5-like helicase domains in both of these genomes. A BLAST search detected the closest homologs of these genes in nucleocytoviruses. More distant hits (approximately 30–55% identity across 20–55% of the query sequence) to the AHEV_PLV2 helicase were present in green algae and chytrid fungi, as well as a few spider species (Supplementary Table S13). More distant homologs of the CBEV_PLV1 primase-helicase were abundant in *T. vaginalis* and *Rhizophagus irregularis* (Supplementary Table S13). This is notable given the high copy number of polintons in *T. vaginalis* (Bellas et al. 2023) and the presence of polintons in *R. irregularis* (formerly *Glomus intraradices*) (Kapitonov and Jurka 2006), and suggests that these species may encode many primase-helicase bearing polinton-like elements as well.

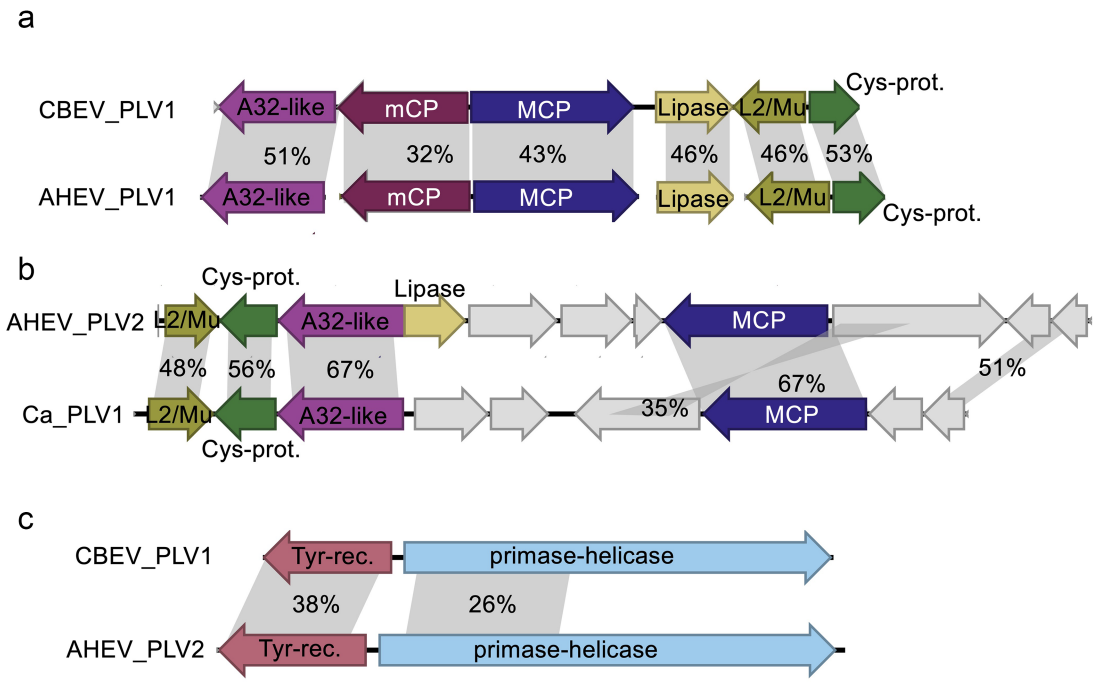

Figure 4. Entomopoxvirus-associated PLVs exhibit genomic mosaicism. Comparisons of entomopoxvirus-associated PLV loci with percent amino acid identity indicated across the shaded regions for CBEV_PLV1 and AHEV_PLV1 structural genes (a), AHEV_PLV2 and Ca_PLV1 structural genes (b), and CBEV_PLV1 and AHEV_PLV2 integration and replication genes (c).

We hypothesized that the AHEV_PLV2 and CBEV_PLV1 helicases encoded n-terminal primase domains, as some primase or polymerase activity would be needed for replication. Alignment of the N-terminal regions of these proteins confirmed the presence of a D5-like archael-eukaryotic primase (AEP) domain (Supplementary Tables S4 and S14). The D5 primase-helicases are conserved amongst poxviruses and can be found throughout *Nucleocytoviricota* and include the primase-polymerase (PRIMPOL) DNA repair gene found in eukaryotes (Iyer et al. 2005). Key catalytic residues have previously been identified for both vaccinia virus D5 (De Silva et al. 2007) and human PRIMPOL (Boldinova et al. 2023). While the AHEV_PLV2 and CBEV_PLV1 D5 primase domains are more compact than other viral homologs, the catalytic residues are well-conserved (Fig. 3b). For the catalytic residues, the only difference in the PLV D5 primases is the substitution of a basic histidine for a more basic arginine in AHEV_PLV2 (Fig. 3b). Thus, despite belonging to separate MCP lineages, AHEV_PLV2 and CBEV_PLV1 have replication machinery that is related to each other and distinct from members within their own MCP lineage.

This pattern of mosaicism occurs across the genomes of the EPV-associated PLVs. The separate clustering of MCPs can be extended to most of the morphogenic genes we identified (Fig. 4a and b), revealing two lineages of morphogenic gene modules. When compared to proteins in NCBI GenBank, members of the AHEV_PLV1 and CBEV_PLV1 modules exhibited the highest affinity for each other, while the Ca_PLV1 module showed the highest affinity to that of AHEV_PLV2 (Supplementary Table S13). This is notable given that each morphogenic lineage includes two elements with dissimilar replication genes. Additionally, our BLAST results suggest that these morphogenic lineages have a broader presence in insects. While we were unable to find capsid homologs in GenBank, homologs of the other AHEV_PLV1/CBEV_PLV1 morphogenic genes were highly abundant within insect genomes. Homologs were present in diverse insect taxa and appeared especially abundant in *Hemiptera*, *Hymenoptera*, and *Coleoptera* (Supplementary Table S13). Amongst these hits to insect genomes, there were also several hits to annotated "adinotviruses" (Supplementary Table S13). Previous work (Bellas et al. 2023), as well as our own analysis (Fig. 2), has shown that adintoviruses are mostly synonymous to maverick-polintons (mavpols). Homology between the AHEV_PLV1 and CBEV_PLV1 morphogenic genes to adintoviruses aligns with our MCP tree that shows AHEV_PLV1 and CBEV_PLV1 to be divergent members of the mavpol2 MCP clade (Fig. 2). While the elements carrying these homologs are mostly unannotated on NCBI, their sequence similarity to adintovirus genes provides preliminary evidence that mavpol elements are highly abundant within sequenced insect genomes. Blastp results for the AHEV_PLV2 and Ca_PLV1 morphogenic genes also revealed homologs within insects, but such hits were generally weaker and fewer in number (Supplementary Table S13). This is consistent with our finding that the AHEV_PLV2 and Ca_PLV1 MCP sequences appear to be less closely related to those of known PLV groups (Fig. 2).

In addition to their related primase domains, AHEV_PLV2 and CBEV_PLV1 also share related tyrosine recombinases despite belonging to separate virion lineages (Fig. 4c). Notably, homology was variable on a sub-protein level. The AHEV_PLV2 and CBEV_PLV1 primase-helicases shared 26% amino acid identity across a 312aa n-terminal region (Fig. 4c), while the rest of the proteins, including the helicase domains, did not show detectable similarity, suggesting the prior occurrence of sub-genic domain swapping. Taken together, these observations show that EPV-associated PLVs are mosaic elements and have exchanged gene modules and functional domains over the course of their evolution.

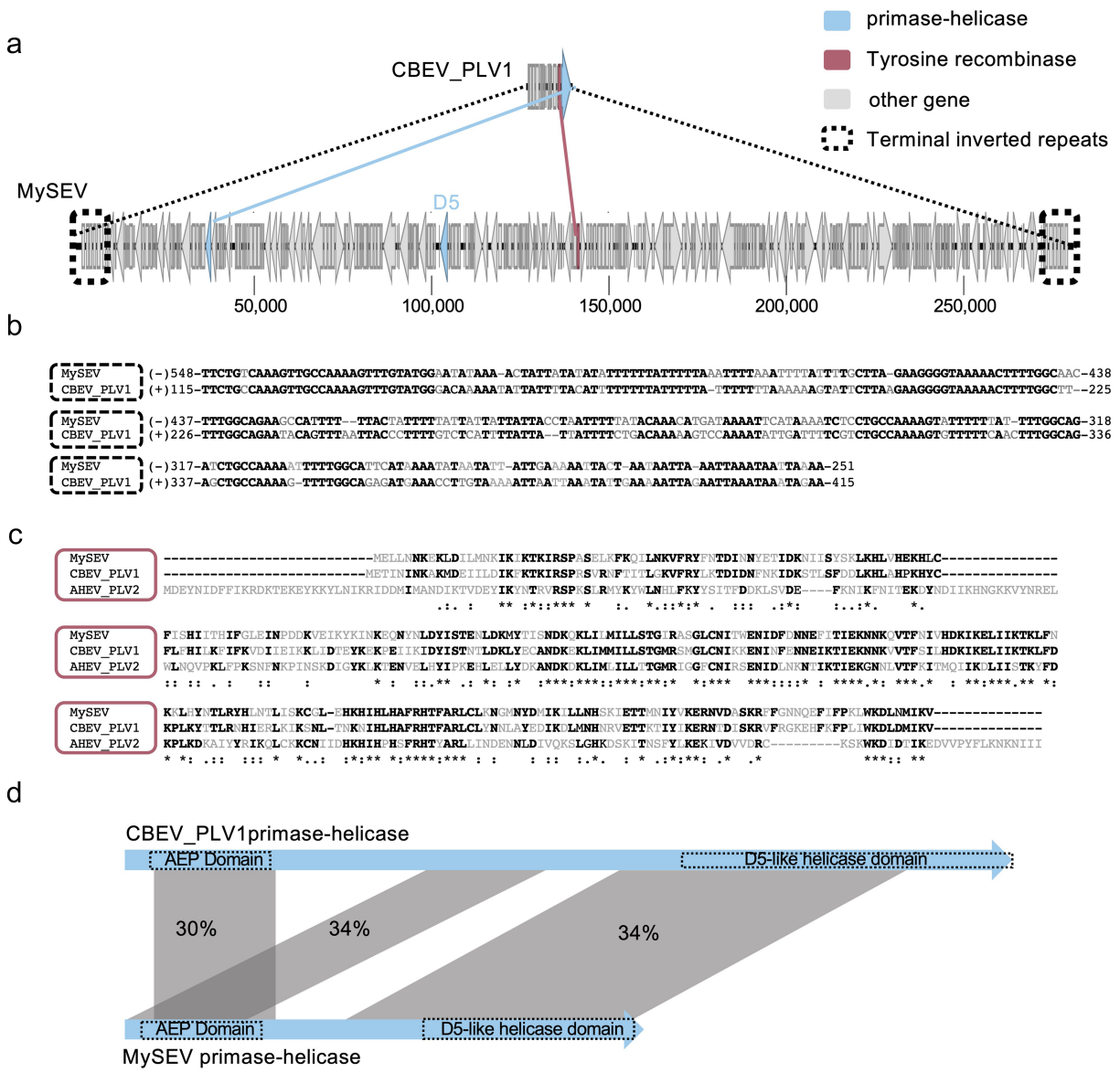

Figure 5. PLV end module shows homology to entomopoxvirus genes and non-coding regions. (a) Diagram of CBEV_PLV1 and MySEV genomes with homologous features labeled and indicated by colored lines. (b) An alignment of homologous nucleotide subsequences that occur within the TIRs of CBEV_PLV1 and MySEV. The strand on which the displayed sequences occur are denoted as (+) for the plus strand, and (-) for the minus strand. Mismatched bases are grey, matching bases are black and bolded. (c) An alignment of the tyrosine recombinase protein sequence from MySEV, CBEV_PLV1, and AHEV_PLV2. Positions with a consensus across at least two of the three proteins are bolded in black. "." indicates conservation across groups with weakly similar properties. ":" indicates conservation across groups with strongly similar properties. "*" indicates complete conservation of the residue. (d) Diagram showing the shared amino acid identity across shaded regions for MySEV and CBEV_PLV1 primase-helicase genes.

An entomopoxvirus-associated polinton-like virus end module is related to entomopoxvirus-encoded genes

It may be possible to infer aspects of PLV biology from their genomic organization. For both AHEV_PLV2 and CBEV_PLV1, the primase-helicase gene is encoded between a set of inverted repeats and the tyrosine recombinase. Tyrosine recombinases typically use inverted repeats as their substrates for recombination (Rajeev et al. 2009), and it is common within integrating elements for the integrase to be encoded next to the integration site (Smyshlyaev et al. 2021). The inverted repeats in EPV-associated PLVs are most likely cognate to the terminal inverted repeats (TIRs) that flank integrated polintons and PLVs, as integration of a mobile element via a tyrosine recombinase will generate left and right attachment sites that are inverted repeats of each other (Grainge and Sherratt 2007). The placement of the primase-helicase between the IRs and tyrosine recombinase suggests these three PLV components may interact, with the IRs functioning both as a substrate for integration by the tyrosine recombinase and as the origin of replication recognized by the primase-helicase. We hypothesize that such *cis*-acting functions in the IRs would select for co-clustering of their interacting proteins, giving rise to functional modules at the end of the genome. Consistent with this replicative end-module hypothesis, many previously published PLV genomes display replication and integration genes encoded either near each other at one end of the element or at opposite ends of the element (Yutin et al. 2013, 2015, Koonin and Krupovic 2017). Indeed, the replicases of all four of the novel PLVs described here are located near the IRs (Fig. 1).

Further supporting the existence of a replicative end module, we found that all three components of CBEV_PLV1's end module had homologs within the genome of Mythimna separata entomopoxvirus (MySEV) (Fig. 5). Overlapping the CBEV_PLV1 IRs and occurring within MySEV TIRs, there are approximately 300bp regions that share 72% nucleotide identity. These sequences occur 250bp from MySEV's genomic termini and comprise the bulk of CBEV_PLV1's IR sequence, respectively (Fig. 5b, Supplementary Figure S2). Similarly, MySEV encoded a tyrosine recombinase (Fig. 5c) and a primase-helicase (Fig. 5d) that were closely related to those of CBEV_PLV1. While these MySEV components were encoded in separate regions of the MySEV genome, poxviruses have been hypothesized to initiate replication out of hairpin structures, referred to as telomeres, that occur within their TIRs (Shenouda et al. 2022). Tyrosine recombinase activity on these ends could be a mechanism for genome dimer resolution, as occurs in bacterial genomes, plasmids, and some bacteriophages (Haenebalcke and Haigh 2013, Castillo et al. 2017).

Like other poxviruses, MySEV encodes a D5 primase-helicase homolog that would be expected to serve as the main primase for genome replication (Fig. 5a). It is unclear why MySEV would encode two primase-helicase genes, but it is possible that the two primase-helicases have specialized functions. There is precedent for larger dsDNA viruses encoding multiple origins of replication that have idiosyncratic gene requirements for their replication (Brister 2008, Wu et al. 2014). The additional primase-helicase could be acting on an alternative replication origin or provide some auxiliary replicative function.

The dispersed loci of poxvirus replication genes contrast with the modular architecture of the PLVs. This difference may reflect different organizational principles of PLV and poxvirus genomes. Our analysis suggests a high level of mosaicism amongst EPV-associated PLVs. This requires that cooperating components of functional modules be encoded together, to reduce the chance that they be separated through recombination and rendered nonfunctional. While poxviruses are prone to recombination (Bobay and Ochman 2018, Sasani et al. 2018), they have much larger and more complex genomes, and do not exhibit the same mosaic modularity as the EPV-associated PLVs. Modularity may be a trait of small genomes, as greater complexity would likely require greater functional integration of genomic components.

The similarity of these MySEV components to the CBEV_PLV1 end module is remarkable in that they represent the strongest instances of both nucleotide and amino acid sequence similarity that we were able to detect between the EPV-associated PLVs and EPVs. This high level of homology indicates relatively recent horizontal gene transfer between EPVs and EPV-associated PLVs, and suggests that essential PLV replication and integration genes may have originated in viruses parasitized by PLVs.

Entomopoxvirus-associated polinton-like viruses encode replication genes related to mobile entomopoxvirus helicases

In addition to the MySEV-encoded primase-helicase, a number of nucleocytovirus-encoded helicases shared sequence identity with the primase-helicases of CBEV_PLV1 and AHEV_PLV2 (Fig. 6a and b). In particular, the primase-helicase from CBEV_PLV1 had sequence similarity to helicases from its co-isolated virus CBEV and CBEV's close relative CREV (Supplementary Figure S3), consistent with HGT between EPVs and associated PLVs. In contrast, the helicase domain of AHEV_PLV2 resembled those of primase-helicases from ascoviruses, namely the lepidopteran infecting Heliothis virescens ascoviruses (HVAV) and Diadromus pulchellus ascovirus (DPAV), isolated from a parasitoid wasp that preys on lepidopterans. The relationship between the AHEV_PLV2 helicase domain and those of the two ascoviruses suggests that EPV-associated PLVs are part of a broader gene flow network that includes non-poxvirus nucleocytovirus genomes, and may even extend within species interactions networks, as evidenced by other occurrences of horizontal gene transfers between lepidopterans, parasitoids, and viruses (Thézé et al. 2015, Gasmi et al. 2021).

All of the helicase domains with detectable similarity to the AHEV_PLV2 and CBEV_PLV1 domains have detectable similarity to the vaccinia virus primase-helicase protein D5 via HHpredor have close homologs where such similarity was readily detectable. Notably, these domains clustered independently from those of the core D5 homologs present within sequenced EPVs (Fig. 6b). In the case of the ascovirus specific cluster, the HVAV and DPAV lack a canonical D5 homolog and each have two copies of these divergent D5-like primase helicase. D5 is considered a core gene of nucleocytoviruses, so it is likely that one or both of the primase-helicases within the ascovirus genomes are essential, but further empirical evidence is required to infer their functions. On the other hand, all EPVs possess identifiable core D5 homologs, suggesting that their divergent helicases have a distinct function.

All of the CBEV, CREV, and Yalta virus-encoded helicases occur in the TIRs. Additionally, we hypothesized that the MySEV primase-helicase may act on MySEV's TIRs due to similarities with the CBEV_PLV1 end module, further tying this helicase cluster to EPV TIRs. Within nucleocytoviruses, the TIRs are often hotspots for diversity and repetitive sequences (Mönttinen et al. 2021). For example, CBEV and CREV have highly similar genomes with the bulk of their unrelated sequence being located within the TIRs (Thézé et al. 2013). A closer examination of the CBEV, CREV, and Yalta virus helicases suggested that these genes may themselves be mobile. In these genomes, the accessory helicases show an association with T5orf172 genes (Fig. 5c). The T5orf172 protein domain is a subfamily of the GIY-YIG endonuclease domain and was recently proposed to comprise a widespread family of homing endonuclease genes (HEGs) in bacteriophages (Barth et al. 2023). HEGs are mobile elements that mobilize through the cleavage of distinct, yet related, rival loci that lack the HEG. The broken DNA is repaired through homologous recombination, using the HEG coding locus as a repair template, which causes the HEG to be copied into the cognate genome, facilitating the HEG's spread (Burt and Koufopanou 2004, Stoddard 2011, Belfort and Bonocora 2014). HEGs are highly abundant in bacteriophages (Edgell et al. 2010, Barth et al. 2023) and have also been observed within nucleocytovirus genomes (Deeg et al. 2018, Mirzakhanyan and Gershon 2020).

The helicase and T5orf172 genes in CBEV and CREV have several features that suggest their mobility. The uneven level of sequence similarity and the presence of genetic rearrangements (Fig. 6c) suggest recombination at these loci, which is consistent with a homing mechanism. Similar to previously recognized HEGs (Barth et al. 2023), the EPV T5orf172 genes exist as multiple copies of homologous orfs with a modular and repetitive domain structure (Fig. 6d). While the helicases do not display the same modularity, they too exist as repeated homologous coding sequences (Fig. 6d). The helicases are flanked by repeat sequences, a hallmark of mobile elements (Fig. 6d, Supplementary Table S15, Supplementary Figure S4).

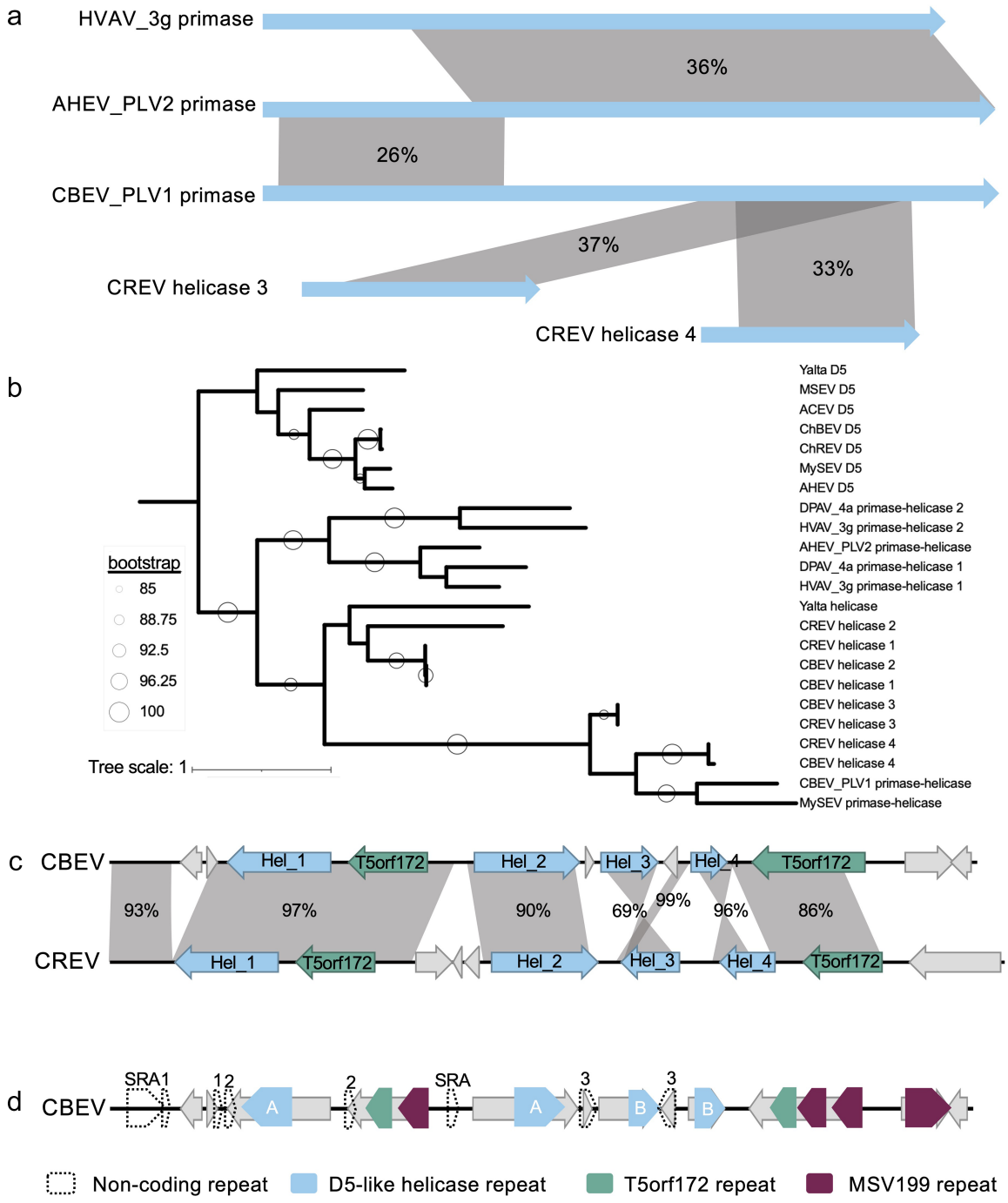

Figure 6. PLV replication genes are related to mobile primase-helicase genes from viruses. (a) Diagram showing regions of homology across primase-helicase and helicase genes. Genes are depicted as blue arrows. Percent amino acid identity across shaded regions is indicated. (b) Pairwise Identity matrix or tree of the D5-like helicase domains from proteins encoded by EPVs, EPV-associated PLVs, and ascoviruses. (c) Diagram of CBEV and CREV tandem inverted repeat (TIR) subregions. Gray shaded regions indicate areas of nucleotide similarity. "Hel" is helicase. (d) Repeat sequences within the CBEV TIR. Repeat types are denoted by color. For helicase-coding and non-coding repeats, homologous sequences are given the same label. "SRA" is short repeat array.

Beyond our own observations in EPVs, an association between D5-like helicases and T5orf172 genes has been previously observed throughout nucleocytoviruses (Mönttinen et al. 2021). The association of these gene groups likely has a mechanistic basis. SF3 helicases domains are known to assist in rolling circle replication for several eukaryotic viruses and bacterial plasmids. The helicase domain is thought to assist in unpairing DNA at the replication origin, while an HUH nuclease domain generates a single strand nick that allows for the initiation of rolling circle replication (Zhao et al. 2019, Tarasova and Khayat 2021). A family of eukaryotic transposons, the helitrons, have repurposed this process for transposition while swapping out their SF3 helicase domain for an SF1B helicase domain of the Pif1 family (Chandler et al. 2013, Craig 2023). As helitrons appear to have replaced the ancestral helicase domain for one of analogous function, it is conceivable that a different nicking nuclease, fused to the helicase or

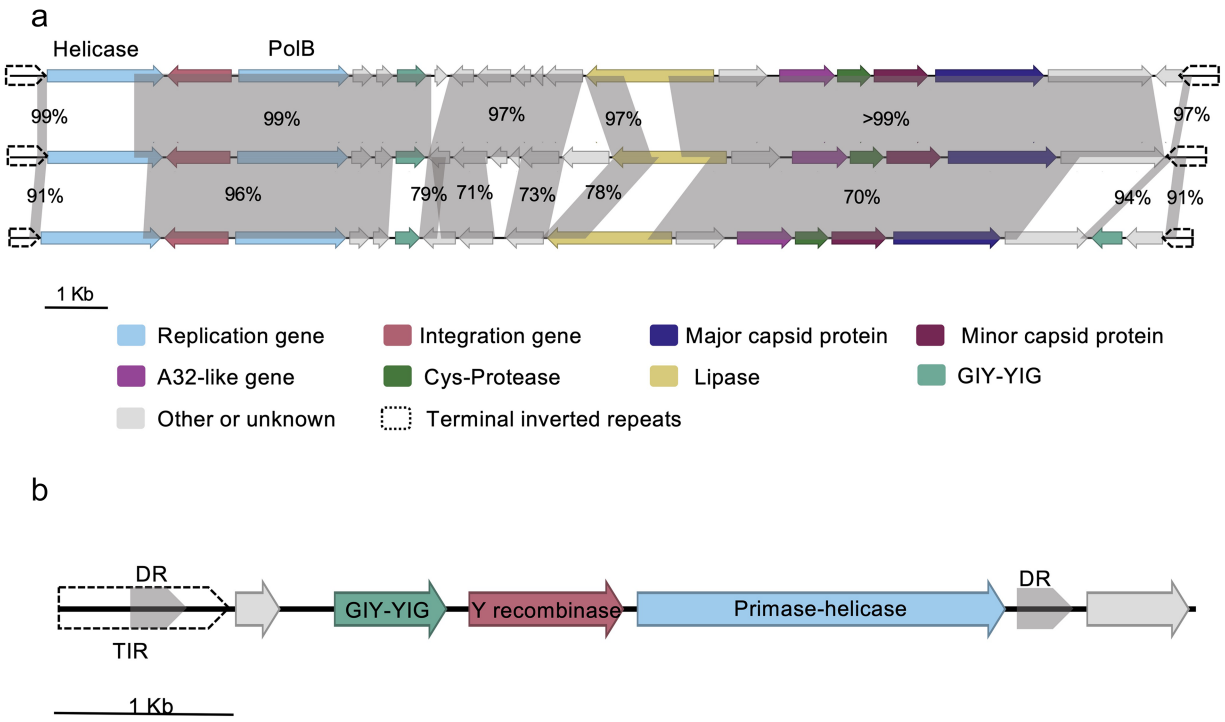

Figure 7. Virophage and PLV end modules bear signatures of mobility. (a) Diagram showing regions of shared nucleotide identity across Mavirus isolates. TIRs and genes of interest are indicated by color. (b) A subregion of the Gezel-14T genome with genes and genomic features labeled. "Y recombinase" is tyrosine recombinase. "DR" is direct repeat.

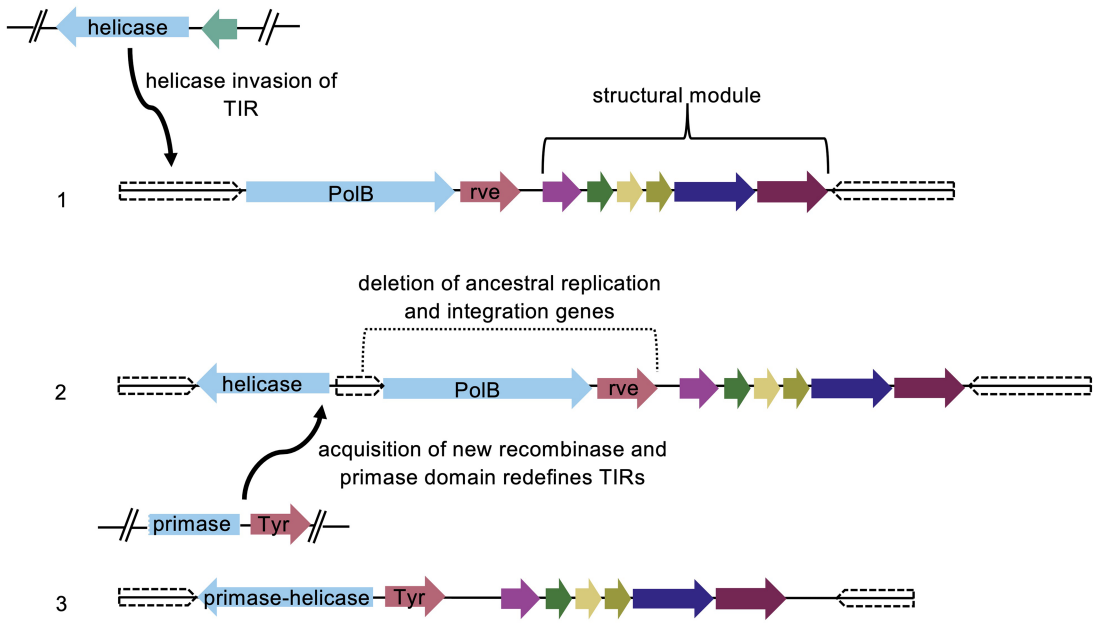

Figure 8. A model for PLV replicon replacement. A hypothetical model of replicon replacement in PLVs and related elements. First, mobile helicase with affinity for the terminal inverted repeats (TIRs) invades the element TIR. Second, the mobile helicase acquires replication and recombination functions while the original genes providing these functions for the element are lost. The end result is a chimeric element where the morphogenic module has been maintained and the TIRs and replication and integration genes have been replaced.

as its own separate gene, could fulfill the role of the HUH domain.

The repeat sequences flanking the helicase do not encompass the T5orf172 endonuclease genes, suggesting that the genes do not transpose as a single unit, but this organization does not preclude the possibility of these genes functioning together. HEGs sometimes mobilize adjacently encoded self-splicing introns through a process called "collaborative homing" (Bonocora and Shub 2009, Zeng et al. 2009). In these instances, HEGs target the intronless allele, facilitating the co-mobilization of the intron and HEG through homologous recombination (Zeng et al. 2009). This arrangement has been proposed to be a predecessor of the classically described intronic HEGs found in many diverse organisms

Table 1. Virus genomes analyzed.

Name	Accession number	Notes
Mavirus (1)	KU052222	
Mavirus (2)	OX559518	
Mavirus (3)	OX560865	
Gezel-14T	NC_021333	
SFMaverick1	BK011025	
SFMaverick2	BK011026	
AHEV_PLV1	KJ683044	
AHEV_PLV2	KJ683045	
CBEV_PLV1	KJ683046	
Ca_PLV1	OX421890	*Cetonia aurata* chromosome 9 positions 23463350 to 23477421
Adoxophyes honmai entomopoxvirus (AHEV)	HF679131	
Choristoneura biennis entomopoxvirus (CBEV)	NC_021248	
Choristoneura rosaceana entomopoxvirus (CREV)	NC_021249	
Mythimna separata entomopoxvirus (MySEV)	NC_021246	
Anomala cuprea entomopoxvirus (ACEV)	NC_023426	
Amsacta moorei entomopoxvirus (AMEV)	NC_002520	
Diachasmimorpha longicaudata entomopoxvirus (DlEPV)	KR095315	
Melanoplus sanguinipes entomopoxvirus (MSEV)	NC_001993	
Yalta virus isolate UA_Yal_14_16 (Yalta)	MT364305	
Heliothis virescens ascovirus 3g (HVAV_3g)	NC_044939	
Diadromus pulchellus ascovirus 4a (DPAV_4a)	NC_011335	

(Bonocora and Shub 2009). In bacteriophages, HEG mobilization is facilitated largely through strand invasion (Stoddard 2014), a process that, like rolling circle replication and helitron transposition, involves extrusion of ssDNA. Collaborative homing could provide an alternative to transposition for the co-mobilization of helicase and nuclease gene functions.

The presence of these helicases in EPV TIRs and their similarity to components of the CBEV_PLV1 replication end module led us to take a closer look at PLV and virophage replication machinery. Apart from their proximity to TIRs, we were able to find other similarities between some PLV and virophage primase-helicases and mobile helicases in EPVs. In multiple cases, primase-helicases or PolB genes are encoded proximally to GIY-YIG endonucleases, a protein superfamily that includes T5orf172. The putative mavirus end module consists of a *rve* integrase and a PolB gene. Less than 800bp from the PolB gene is a GIY-YIG endonuclease. On the other side of the mavirus end module, a helicase occurs between the TIR and the *rve* integrase. This end module appears in divergent virophage genomes (Fig. 7a), and there are also variant alleles of the helicase, and GIY-YIG genes, suggesting that the end module and flanking genes are mobile. Along similar lines, a GIY-YIG endonuclease occurs between the Gezel-14T end module and its TIR (Fig. 7b). Reminiscent of repeat sequences flanking EPV mobile helicases, a >300bp direct repeats occurs within the Gezel-14T TIR as well as immediately downstream of the end module (Fig. 6b), suggesting that the GIY-YIG endonuclease, tyrosine recombinase, and primase-helicase may all be part of a mobile module.

From these observations, we have developed an evolutionary model for replicative and integrative end module turnover in PLVs and virophages (Fig. 8). First, an element bearing a PolB end module has one of its TIRs invaded by a mobile helicase, disrupting the TIR sequence. Acquisition of a primase domain to the helicase and tyrosine recombinase restores replication and recombination functions to the altered element end. The ancestral PolB and *rve* integrase no longer confer a benefit to the element and will eventually be lost through mutation. Alternatively, the association of some replicon loci with nucleases, such as the GIY-YIG endonuclease in Gezel-14T, may implicate nuclease conflicts, such as endonuclease homing or host–parasite conflicts, as being drivers of replicon turnover, as has been suggested for some bacteriophages and bacteriophage satellites (Goodrich-Blair and Shub 1996, Barth et al. 2021).

The validity of our end module turnover model can only become clear with experimental follow-up, but it highlights an intriguing relationship between mobile genes in nucleocytovirus TIRs, and the end modules of PLVs and virophages. Connections between chromosomal ends and mobile elements have been described in diverse systems. Some herpesviruses can integrate within the telomeres at the end of chromosomes in animals (Gennart et al. 2015, Wood et al. 2021), and more recently, some herpesvirus-related viruses were shown to integrate in the telomeres of thraustochytrid protists (Collier et al. 2023). In circular chromosomes of bacteria, many mobile elements integrate into the chromosome dimer resolution site located near the replication terminus (Midonet and Barre 2014, 2016). In line with our model, there have been multiple instances of mobile element domestication for chromosome end maintenance. In *Drosophila*, telomeres are maintained by retrotransposons that integrate specifically at the ends of chromosomes (Mason et al. 2008, Pardue and DeBaryshe 2008). Especially analogous to our model of TIR invasion and end module turnover are transposons of the Tn5053 family found in bacteria. These transposons are "res-site hunters" that integrate into the resolution sites of other mobile elements, abolishing the functions of those resolution sites (Minakhina et al. 1999). Loss of resolution site function would be an evolutionary dead end for these mobile elements, but the Tn5053 family transposons encode their own resolution sites and recombinases, and most likely become an essential component of the elements they invade (Minakhina et al. 1999). The importance of replication and recombination for the maintenance of chromosomal ends likely makes these sites attractive targets for mobile element invasion. We suspect that element acquisition at chromosome ends and the domestication of such elements for end maintenance is likely a recurrent feature in evolution. In PLVs specifically, it may explain the high levels of genomic mosaicism.

Conclusions

Here, we describe the genomes of four novel PLVs, the first of such elements to be linked to poxviruses. The presence of three of these elements in EPV occlusion bodies strongly suggests that they are hyperparasites of EPVs. Symbiosis between these elements and

EPVs, along with their highly diverged capsid genes, suggests that our knowledge of PLV biology and diversity still remains fairly limited. Although several PLVs are linked with other viruses in the phylum *Nucleocytoviricota*, none have ever been found for the class *Pokkesviricetes*, which includes the families *Poxviridae* and *Asfarviridae* (Koonin et al. 2020, Aylward et al. 2021). Hence, our results suggest that hyperparasitic viruses may be a common feature of large DNA viruses in the *Nucleocytoviricota*. Given that many viruses in this phylum infect humans or agriculturally important livestock, our results suggest that it may even be possible to discover hyperparasitic viruses that could be useful for therapeutic or prophylaxis applications.

Analysis of the four EPV-associated PLV genomes shows PLVs to be mosaic elements that have variable associations between different morphogenesis modules and the modules that facilitate both replication and integration. The four genomes we analyzed can be split into two separate lineages based on MCP phylogeny. One lineage appears to be a novel branch of the maverick-polintons, while the other lineage did not fall within any recognized groups but appeared most closely related to the PgVV group of PLVs (Fig. 2d). For the elements' replication genes, two elements encoded unrelated PolB proteins (Fig. 3a), while the other two elements encoded primase-helicases that had detectable homology in their primase domains but not their helicase domains (Fig. 3b). These observations suggest genomic exchanges of both entire gene modules and sub protein domains have occurred in the course of the evolution of these elements.

In addition to genetic exchanges between EPV-associated PLVs, it is likely that there have also been exchanges between EPVs and PLVs. Recombination and integration modules appear linked to the PLV IRs, suggesting that they might use them as an origin of replication and recombination substrate for integration. We observe homology between these end modules and EPV genes, including a number of helicases that show signs of mobility and appear to be linked to homing endonucleases. Given that homing endonucleases, recombination, and homology to replication-associated domains have been weaponized in virus–satellite conflicts (Barth et al. 2021, 2023, Nguyen et al. 2022), we expect the gene flow between EPVs and PLVs to be a major driver of their evolution. We have also put forward a model that some PLV end modules are mobile, or at least have evolutionary roots as mobile elements.

Materials and methods

Sequences analyzed

Genomes of EPVs and their associated PLVs were assembled from 454 pyrosequencing reads generated from occlusion bodies (Thézé et al. 2013). We mapped reads from the original 454 data and confirmed that the PLVs were present in approximately 1× to 1.5× sequencing depth as the EPVs (Supplementary Table S1). Raw sequencing reads were deposited in NCBI SRA under the Bioproject accession PRJNA1078858, and all genome sequences used in this work are publicly available. Major capsid protein sequences for previously recognized PLVs, apart from adintoviruses and MELD viruses, were obtained from Bellas et al. (2023). Major capsid sequences for adintoviruses and MELD viruses were obtained from Starrett et al. (2021), Starrett et al. (2021), and Wallace et al. (2021). All other sequences analyzed appear in Table 1.

Gene function and structural prediction

Gene functional predictions were used using the HHpred (Steinegger et al. 2019) tool hosted by the Max Planck Institute Bioinformatics toolkit (Zimmermann et al. 2018, Gabler et al. 2020).

Protein sequence alignments were generated using Muscle5 (Edgar 2022) using default parameters.

Protein models were generated using the ColabFold webserver (Mirdita et al. 2022) using our custom MSA or single sequence templates. The alphafold2_ptm (Jumper et al. 2021) model type was used for monomeric model prediction with 48 recycles and a recycle early stop tolerance of 0.0. We moved forward with the highest ranked models for further analysis. Structural homology was searched using the Foldseek webserver (van Kempen et al. 2023) using default parameters. Data Bank (PDB) files for the predicted structures are included in the supplementary data.

Phylogenetic tree construction

Protein sequence alignments were generated using Muscle5 (Edgar 2022) using default parameters, and IQTree (Nguyen et al. 2015) was used to generate trees from the alignments, using ModelFinder Plus (Kalyaanamoorthy et al. 2017) for model selection and ultrafast bootstraps (Hoang et al. 2018). The best tree of 10 runs was selected. For the major capsid alignment, the alignment was trimmed using trimAl (Capella-Gutiérrez et al. 2009) prior to tree generation removing positions with a gap in 90% or more of sequences. For the D5-like helicase domains, full length proteins were manually trimmed prior to alignment based on locations of predicted domains and homologous regions. Trees were visualized using the iTOL webserver (Letunic and Bork 2021). Newick files for all trees are included in the supplementary data.

Sequence similarity detection

Homologs of entomopoxvirus-associated PLV proteins in GenBank were identified through BLAST using the NCBI web portal with default search parameters.

Sequence similarity presented in Figs 4, 5d, 6a and c, and 7a were detected through BLAST (BLAST+ 2.7.1) (Camacho et al. 2009) using the parameters word size 11, Match 4, Mismatch −5, gap costs existence 12 extension 8 for nucleotide sequences and word size 3 matrix BLOSUM62, and gap cost existence 11 extension 1 for peptide sequences. Alignments in Fig. 5b and c, Supplementary Figures S2 and S4 were generated using ClustalOmega (Sievers et al. 2011, Sievers and Higgins 2018) using default parameters and were manually curated.

Acknowledgements

We thank Basil Arif for assistance collecting the original samples. We would also like to thank members of the Alward lab for useful discussion. We thank Abdeali Jivali for assistance with PyMol. We acknowledge the use of the Virginia Tech Advanced Research Computing cluster for research performed in this study.

Supplementary data

Supplementary data is available at *VEVOLU Journal* online.

Conflict of interest: None declared.

Funding

This work was supported by grants from the National Science Foundation (CAREER-2141862 to FOA) and National Institutes of Health (1R35GM147290-01 to FOA).

Data availability

All sequences analyzed in this work are publicly available. Accession numbers for PLVs and entomopoxviruses are provided in Table 1 under materials and methods. Newick files and pdb files used to produce figures can be found online at DOI: 10.5281/zenodo.12585766.

References

Anderson CW, Young ME, Flint SJ. Characterization of the adenovirus 2 virion protein, mu. *Virology* 1989;**172**:506–12.

Aylward FO, Moniruzzaman M, Ha AD *et al*. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol* 2021;**19**:e3001430.

Barreat JGN, Katzourakis A. Phylogenomics of the maverick virus-like mobile genetic elements of vertebrates. *Mol Biol Evol* 2021;**38**:1731–43.

Barreat JGN, Katzourakis A. A billion years arms-race between viruses, virophages, and eukaryotes. *eLife* 2023;**12**: RP86617.

Barth ZK, Dunham DT, Seed KD. Nuclease genes occupy boundaries of genetic exchange between bacteriophages. *NAR Genom Bioinform* 2023;**5**.

Barth ZK, Nguyen MH, Seed KD. A chimeric nuclease substitutes a phage CRISPR-Cas system to provide sequence-specific immunity against subviral parasites. *eLife* 2021;**10**:e68339.

Belfort M, Bonocora RP. Homing endonucleases: from genetic anomalies to programmable genomic clippers. *Methods Mol Biol* 2014;**1123**:1–26.

Bellas C, Hackl T, Plakolb M-S *et al*. Large-scale invasion of unicellular eukaryotic genomes by integrating DNA viruses. *Proc Natl Acad Sci USA* 2023;**120**:e2300465120.

Bellas CM, Sommaruga R. Polinton-like viruses are abundant in aquatic ecosystems. *Microbiome* 2021;**9**:13.

Bobay L-M, Ochman H. Biological species in the viral world. *Proc Natl Acad Sci USA* 2018;**115**:6040–45.

Boldinova EO, Baranovskiy AG, Gagarinskaya DI *et al*. The role of catalytic and regulatory domains of human PrimPol in DNA binding and synthesis. *Nucleic Acids Res* 2023;**51**:7541–51.

Bonocora RP, Shub DA. A likely pathway for formation of mobile group I introns. *Curr Biol* 2009;**19**:223–28.

Brister JR. Origin activation requires both replicative and accessory helicases during T4 infection. *J Mol Biol* 2008;**377**:1304–13.

Burt A, Koufopanou V. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr Opin Genet Dev* 2004;**14**:609–15.

Camacho C, Coulouris G, Avagyan V *et al*. BLAST+: architecture and applications. *BMC Bioinf* 2009;**10**:421.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972–73.

Castillo F, Benmohamed A, Szatmari G. Xer site specific recombination: double and single recombinase systems. *Front Microbiol* 2017;**8**:453.

Chandler M, de la Cruz F, Dyda F *et al*. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol* 2013;**11**:525–38.

Collier JL, Rest JS, Gallot-Lavallée L *et al*. The protist Aurantiochytrium has universal subtelomeric rDNAs and is a host for mirusviruses. *Curr Biol* 2023;**33**:5199-207.

Craig RJ. Replitrons: A major group of eukaryotic transposons encoding HUH endonuclease. *Proc Natl Acad Sci USA* 2023;**120**:e2301424120.

Deeg CM, Chow C-ET, Suttle CA. The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. *eLife* 2018;**7**:e33014.

De Silva FS, Lewis W, Berglund P *et al*. Poxvirus DNA primase. *Proc Natl Acad Sci USA* 2007;**104**:18724–29.

de Sousa JAM, Fillol-Salom A, Penadés JR *et al*. Identification and characterization of thousands of bacteriophage satellites across bacteria. *Nucleic Acids Res* 2023;**51**:2759–77.

Edgar RC. Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun* 2022;**13**:6968.

Edgell DR, Gibb EA, Belfort M. Mobile DNA elements in T4 and related phages. *Virol J* 2010;**7**:290.

Fischer MG. The virophage family lavidaviridae. *Curr Issues Mol Bio* 2021;**40**:1–24.

Fischer MG, Suttle CA. A virophage at the origin of large DNA transposons. *Science* 2011;**332**:231–34.

Fritsch C, and Mayo MA Satellites of plant viruses. In: Mandahar CL (ed.), *Plant Viruses*.Boca Raton, FL, United States: CRC Press, 2018, 289–321.

Gabler F, Nam S-Z, Till S *et al*. Protein sequence analysis using the MPI bioinformatics toolkit. *Curr Protoc Bioinf* 2020;**72**:e108.

Gaia M, Benamar S, Boughalmi M *et al*. Zamilon, a novel virophage with Mimiviridae host specificity. *Plos One* 2014;**9**:e94923.

Gasmi L, Sieminska E, Okuno S *et al*. Horizontally transmitted parasitoid killing factor shapes insect defense to parasitoids. *Science* 2021;**373**:535–41.

Gennart I, Coupeau D, Pejaković S *et al*. Marek's disease: genetic regulation of gallid herpesvirus 2 infection and latency. *Vet J* 2015;**205**:339–48.

Gnanasekaran P, Chakraborty S. Biology of viral satellites and their role in pathogenesis. *Curr Opin Virol* 2018;**33**:96–105.

Goodrich-Blair H, Shub DA. Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell* 1996;**84**:211–21.

Grainge I, Sherratt DJ. Site-specific recombination. In: Aguilera A, Rothstein R (eds.), *Molecular Genetics of Recombination. Topics in Current Genetics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, 443–67.

Haapa-Paananen S, Wahlberg N, Savilahti H. Phylogenetic analysis of maverick/polinton giant transposons across organisms. *Mol Phylogen Evol* 2014;**78**:271–74.

Haenebalcke L, and Haigh JJ. Cre/loxP – transgenics. In: Malloy S and Hughes K (eds.), *Brenner's Encyclopedia of Genetics*, 2nd edn. Cambridge, MA, United States: Academic Press 2, 2013, 212–17.

Hoang DT, Chernomor O, von Haeseler A *et al*. Ufboot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;**35**:518–22.

Iyer LM, Koonin EV, Leipe DD *et al*. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res* 2005;**33**:3875–96.

Jumper J, Evans R, Pritzel A *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–89.

Kalyaanamoorthy S, Minh BQ, Wong TKF *et al*. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017;**14**:587–89.

Kapitonov VV, Jurka J. Self-synthesizing DNA transposons in eukaryotes. *Proc Natl Acad Sci USA* 2006;**103**:4540–45.

Koonin EV, Dolja VV, Krupovic M *et al*. Global organization and proposed megataxonomy of the virus world. *Microbiol Mol Biol Rev* 2020;**84**:10–128.

Koonin EV, Krupovic M. Polintons, virophages and transpovirons: a tangled web linking viruses, transposons and immunity. *Curr Opin Virol* 2017;**25**:7–15.

Krupovic M, Bamford DH, Koonin EV. Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. *Biol Direct* 2014;**9**:6.

La Scola B, Desnues C, Pagnier I *et al*. The virophage as a unique parasite of the giant mimivirus. *Nature* 2008;**455**: 100–04.

Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;**49**:W293–W296.

Mason JM, Frydrychova RC, Biessmann H. Drosophila telomeres: an exception providing new insights. *BioEssays* 2008;**30**: 25–37.

Meier AF, Fraefel C, Seyffert M. The interplay between adeno-associated virus and its helper viruses. *Viruses* 2020; **12**:662.

Midonet C, and Barre F-X. Xer site-specific recombination: promoting vertical and horizontal transmission of genetic information. *Microbiol Spectr* 2014;**2**:163–82.

Midonet C, Barre F-X. How Xer-exploiting mobile elements overcome cellular control. *Proc Natl Acad Sci USA* 2016;**113**:8343–45.

Minakhina S, Kholodii G, Mindlin S *et al*. Tn5053 family transposons are res site hunters sensing plasmidal res sites occupied by cognate resolvases. *Mol Microbiol* 1999;**33**:1059–68.

Mirdita M, Schütze K, Moriwaki Y *et al*. ColabFold: making protein folding accessible to all. *Nature Methods* 2022;**19**: 679–82.

Mirzakhanyan Y, Gershon PD. Structure-based deep mining reveals first-time annotations for 46 percent of the dark annotation space of the 9,671-member superproteome of the nucleocytoplasmic large DNA viruses. *J Virol* 2020;**94**:10–128.

Mönttinen HAM, Bicep C, Williams TA *et al*. The genomes of nucleocytoplasmic large DNA viruses: viral evolution writ large. *Microb Genom* 2021;**7**:000649.

Nguyen L-T, Schmidt HA, von Haeseler A *et al*. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268–74.

Nguyen MHT, Netter Z, Angermeyer A *et al*. A phage weaponizes a satellite recombinase to subvert viral restriction. *Nucleic Acids Res* 2022;**50**:11138–53.

Olivier V, Blanchard P, Chaouch S *et al*. Molecular characterisation and phylogenetic analysis of chronic bee paralysis virus, a honey bee virus. *Virus Res* 2008;**132**:59–68.

Pagarete A, Grébert T, Stepanova O *et al*. Tsv-N1: a novel DNA algal virus that infects tetraselmis striata. *Viruses* 2015;**7**:3937–53.

Pardue M-L, DeBaryshe PG. Drosophila telomeres: a variation on the telomerase theme. *Fly* 2008;**2**:101–10.

Rajeev L, Malanowska K, Gardner JF. Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiol Mol Biol Rev* 2009;**73**:300–09.

Ren Y, She Q, Huang L. Transcriptomic analysis of the SSV2 infection of Sulfolobus solfataricus with and without the integrative plasmid pSSVi. *Virology* 2013;**441**:126–34.

Roitman S, Rozenberg A, Lavy T *et al*. Isolation and infection cycle of a polinton-like virus virophage in an abundant marine alga. *Nat Microbiol* 2023;**8**:332–46.

Rolf M, Junglas L, Loch S *et al*. Experimental co-infection of saccamoeba lacustris with mimivirus-like giant virus and a small satellite virus. *Endocyt Cell Res* 2018;**29**:1-6.

Roux S, Fischer MG, Hackl T *et al*. Updated virophage taxonomy and distinction from polinton-like viruses. *Biomolecules* 2023;**13**:204.

Santini S, Jeudy S, Bartoli J *et al*. Genome of phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proc Natl Acad Sci USA* 2013;**110**:10800–05.

Sasani TA, Cone KR, Quinlan AR *et al*. Long read sequencing reveals poxvirus evolution through rapid homogenization of gene arrays. *eLife* 2018;**7**:e35453.

Sheng Y, Wu Z, Xu S *et al*. Isolation and identification of a large green alga virus (chlorella virus XW01) of mimiviridae and its virophage (chlorella virus virophage SW01) by using unicellular green algal cultures. *J Virol* 2022;**96**:e0211421.

Shenouda MM, Noyce RS, Lee SZ *et al*. The mismatched nucleotides encoded in vaccinia virus flip-and-flop hairpin telomeres serve an essential role in virion maturation. *PLoS Pathogens* 2022;**18**:e1010392.

Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* 2018;**27**:135–45.

Sievers F, Wilm A, Dineen D *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;**7**:539.

Smyshlyaev G, Bateman A, Barabas O. Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. *Mol Syst Biol* 2021;**17**: e9880.

Starrett GJ, Tisza MJ, Welch NL *et al*. Adintoviruses: a proposed animal-tropic family of midsize eukaryotic linear dsDNA (MELD) viruses. *Virus Evol* 2021;**7**:veaa055.

Steinegger M, Meier M, Mirdita M *et al*. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinf* 2019;**20**:473.

Stoddard BL. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* 2011;**19**:7–15.

Stoddard BL. Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mobile DNA* 2014;**5**:7.

Stough JMA, Yutin N, Chaban YV *et al*. Genome and environmental activity of a chrysochromulina parva virus and its virophages. *Front Microbiol* 2019;**10**:703.

Tarasova E, Khayat R. A structural perspective of reps from CRESS-DNA viruses and their bacterial plasmid homologues. *Viruses* 2021;**14**:37.

Taylor JM. Virology of hepatitis D virus. *Semin Liver Dis* 2012;**32**:195–200.

Thézé J, Takatsuka J, Li Z *et al*. New insights into the evolution of entomopoxvirinae from the complete genome sequences of four entomopoxviruses infecting adoxophyes honmai, choristoneura biennis, choristoneura rosaceana, and mythimna separata. *J Virol* 2013;**87**:7992–8003.

Thézé J, Takatsuka J, Nakai M *et al*. Gene acquisition convergence between entomopoxviruses and baculoviruses. *Viruses* 2015;**7**:1960–74.

van Kempen M, Kim SS, Tumescheit C *et al*. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* 2023;**42**: 243–46.

Vepštaitė-Monstavičė I, Lukša J, Konovalovas A *et al*. Saccharomyces paradoxus K66 killer system evidences expanded assortment of helper and satellite viruses. *Viruses* 2018; **10**:564.

Wallace MA, Coffman KA, Gilbert C *et al*. The discovery, distribution and diversity of DNA viruses associated with *Drosophila melanogaster* in Europe. *Virus Evol* 2021;**7**: veab031.

Wood ML, Veal CD, Neumann R *et al*. Variation in human herpesvirus 6B telomeric integration, excision, and transmission between tissues and individuals. *eLife* 2021;**10**:e70452.

Wu Y-L, Wu C-P, Huang Y-H *et al*. Identification of a high-efficiency baculovirus DNA replication origin that functions in insect and mammalian cells. *J Virol* 2014;**88**:13073–85.

Yutin N, Raoult D, Koonin EV. Virophages, polintons, and transpovirons: a complex evolutionary network of diverse selfish genetic elements with different reproduction strategies. *Virol J* 2013;**10**:158.

Yutin N, Shevchenko S, Kapitonov V *et al*. A novel group of diverse Polinton-like viruses discovered by metagenome analysis. *BMC Biol* 2015;**13**:95.

Zeng Q, Bonocora RP, Shub DA. A free-standing homing endonuclease targets an intron insertion site in the psbA gene of cyanophages. *Curr Biol* 2009;**19**:218–22.

Zhao L, Rosario K, Breitbart M *et al*. Eukaryotic circular rep-encoding single-stranded DNA (CRESS DNA) viruses: ubiquitous viruses with small genomes and a diverse host range. *Adv Virus Res* 2019;**103**:71–133.

Zimmermann L, Stephens A, Nam S-Z *et al*. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 2018;**430**:2237–43.

Virus Evolution, 2024, 10(1), veae051 , DOI: https://doi.org/10.1093/ve/veae051, Advance Access Publication 12 July 2024Research Article