



HAL
open science

Natural history of eukaryotic DNA viruses with double jelly-roll major capsid proteins

Mart Krupovic, Jens H Kuhn, Matthias G Fischer, Eugene V Koonin

► **To cite this version:**

Mart Krupovic, Jens H Kuhn, Matthias G Fischer, Eugene V Koonin. Natural history of eukaryotic DNA viruses with double jelly-roll major capsid proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 2024, 121 (23), pp.e2405771121. 10.1073/pnas.2405771121 .
pasteur-04597784

HAL Id: pasteur-04597784

<https://pasteur.hal.science/pasteur-04597784v1>

Submitted on 3 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



Natural history of eukaryotic DNA viruses with double jelly-roll major capsid proteins

Mart Krupovic^{a,1} , Jens H. Kuhn^b , Matthias G. Fischer^c , and Eugene V. Koonin^{d,1}

Contributed by Eugene V. Koonin; received March 20, 2024; accepted April 29, 2024; reviewed by Jonathan Filée and Modesto Redrejo Rodríguez

The phylum *Preplasmiviricota* (kingdom *Bamfordvirae*, realm *Varidnaviria*) is a broad assemblage of diverse viruses with comparatively short double-stranded DNA genomes (<50 kbp) that produce icosahedral capsids built from double jelly-roll major capsid proteins. Preplasmiviricots infect hosts from all cellular domains, testifying to their ancient origin, and, in particular, are associated with six of the seven supergroups of eukaryotes. Preplasmiviricots comprise four major groups of viruses, namely, polintons, polinton-like viruses (PLVs), virophages, and adenovirids. We used protein structure modeling and analysis to show that protein-primed DNA polymerases (pPolBs) of polintons, virophages, and cytoplasmic linear plasmids encompass an N-terminal domain homologous to the terminal proteins (TPs) of prokaryotic PRD1-like tectivirids and eukaryotic adenovirids that are involved in protein-primed replication initiation, followed by a viral ovarian tumor-like cysteine deubiquitinylase (vOTU) domain. The vOTU domain is likely responsible for the cleavage of the TP from the large pPolB polypeptide and is inactivated in adenovirids, in which TP is a separate protein. Many PLVs and transpovirons encode a distinct derivative of polinton-like pPolB that retains the TP, vOTU, and pPolB polymerization palm domains but lacks the exonuclease domain and instead contains a superfamily 1 helicase domain. Analysis of the presence/absence and inactivation of the vOTU domains and replacement of pPolB with other DNA polymerases in eukaryotic preplasmiviricots enabled us to outline a complete scenario for their origin and evolution.

polintons | virophages | adenovirus | protein-primed family B-DNA polymerase | deubiquitinylase

The phylum *Preplasmiviricota* (kingdom *Bamfordvirae*, realm *Varidnaviria*) is an expansive assemblage of viruses with moderate-length double-stranded DNA (dsDNA) genomes (<50 kbp) that produce virions with icosahedral capsids (40 to 80 nm) assembled from double jelly-roll major capsid proteins (DJR-MCPs) (1). Preplasmiviricots infect hosts from all cellular domains, testifying to their ancient origin and evolutionary success. Indeed, the prokaryotic preplasmiviricots appear to predate the last universal cellular ancestor (LUCA) (2), whereas association with eukaryotic hosts has been mapped to the onset of eukaryogenesis (3). Extant preplasmiviricots have been detected in six of the seven supergroups of eukaryotes (3) and include four prominent groups of viruses currently referred to as polintons (Mavericks), polinton-like viruses (PLVs), virophages, and adenovirids (1).

Polintons (currently class *Polintoviricetes*) were first described as large, self-synthesizing transposons encoding a predicted protein-primed family B-DNA polymerase (pPolB) and a retrovirid-like integrase (RVE-INT) as well as a cysteine protease of the Ulp1 deubiquitinylase family (vUlp1) and a genome packaging ATPase (Fig. 1) (4, 5). However, subsequent identification of the conserved genes for the major and minor capsid proteins (MCP and mCP, respectively) left little doubt that these elements are better described as endogenous viruses rather than transposons (6–9). Polintons are widespread across the eukaryotic evolutionary tree, being present in both diverse protists and animals, including vertebrates (7, 8, 10). Analysis of metagenomic databases using the polinton MCPs as queries led to the discovery of the distinct, highly diverse group of currently unclassified PLVs which, unlike bona fide polintons, mostly lack the pPolB and RVE-INT genes and instead encode other types of replication proteins and, occasionally, tyrosine superfamily integrases (11–13). However, a variant of PLVs, known as Tlr1 element (14), described in a ciliophore (*Tetrahymena thermophila* Nanney and McCoy, 1976), presents a notable exception. Tlr1 encodes RVE-INT but instead of pPolB it carries a gene for a large multidomain protein containing a C-terminal superfamily 1 helicase (S1H) domain (15), referred to as the “Tlr1-like helicase.” Other PLVs encode superfamily 3 helicases (S3H) fused to either DNA polymerase (DNAP) of family A (PolA) or archaeo-eukaryotic primase-polymerases (AEP) (11).

Significance

Structural modeling of protein domains using advanced AI-based methods such as AlphaFold2 may lead to insights into evolutionary relationships among proteins that are unreachable by sequence analysis. We applied this approach to elucidate the evolutionary relationships of four major groups of eukaryotic viruses: polintons, polinton-like viruses (PLVs), virophages, and adenovirids. We identified additional protein domains predicted to be essential for virus genome replication. Analysis of the presence/absence and inactivation of these domains suggests a complete scenario for the origin and evolution of this major part of the eukaryotic virosphere.

Author affiliations: ^aArchaeal Virology Unit, Institut Pasteur, Université Paris Cité, Paris 75015, France; ^bIntegrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, NIH, Fort Detrick, Frederick, MD 21702; ^cDepartment of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Heidelberg 69120, Germany; and ^dNational Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20894

Author contributions: M.K. and E.V.K. designed research; M.K. performed research; M.K., J.H.K., M.G.F., and E.V.K. analyzed data; and M.K. and E.V.K. wrote the paper.

Reviewers: J.F., Université Paris-Saclay; and M.R.R., Universidad Autónoma de Madrid.

The authors declare no competing interest.

Copyright © 2024 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: mart.krupovic@pasteur.fr or koonin@ncbi.nlm.nih.gov.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2405771121/-/DCSupplemental>.

Published May 28, 2024.

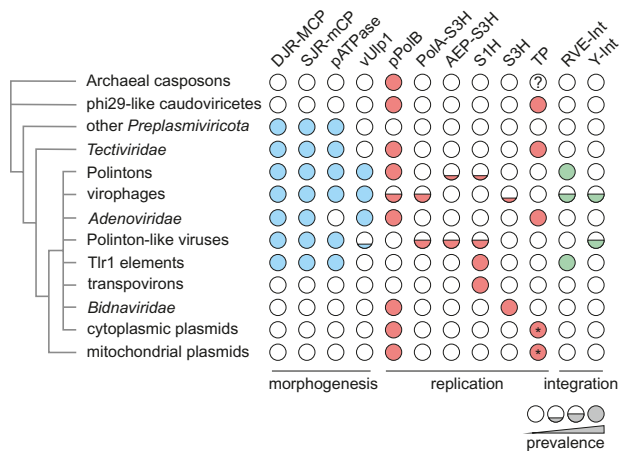


Fig. 1. Relationships among viruses and nonviral mobile genetic elements encoding pPolBs. The schematic tree on the left depicts relationships among the depicted elements and is loosely based on the pPolB phylogenies (9, 27). Note that family 1 archaeal casposons encode pPolBs that form a monophyletic group with homologs from archaeal viruses, distinct from other pPolBs (28, 29). Circles depict gene presence and absence in the corresponding viruses and plasmids. AEP, archaeo-eukaryotic primase-polymerase; DJR-MCP, double jelly-roll major capsid protein; pATPase, FtsK-family genome packaging ATPase; PolA, family A DNAP; RVE-Int, retrovirid-like integrase; S1H, superfamily 1 helicase; S3H, superfamily 3 helicase; SJR mCp, single jelly-roll minor capsid protein; TP, terminal protein; vUlp1, viral Ulp1-family cysteine protease; Y-Int, tyrosine superfamily integrase. Asterisks indicate that the TPs are encoded as part of pPolBs, whereas the question mark denotes uncertainty.

At least two PLVs have been cultured, including *Tetraselmis viridis* virus S1 (TvV-S1) and Gezel-14T, both infecting marine algae (16–18). The third group of eukaryotic preplasmiviricots consists of virophages (currently class *Maveriviricetes*) which have been discovered as hyperparasites of giant viruses assigned to bamfordviraen phylum *Nucleocytoviricota* (19–23). Virophages encode orthologous morphogenetic modules consisting of MCP and mCp, a packaging ATPase, and a vUlp1, but differ in terms of their genome replication proteins, with some virophages encoding pPolB (e.g., mavirus) and others encoding a PolA-S3H fusion (e.g., sputnikviruses) related to that found in some PLVs (24). Finally, the fourth group of eukaryotic preplasmiviricots consists of adenovirids (currently a distinct order in class *Tectiliviricetes*), a rather uniform group of viruses infecting diverse vertebrates, including humans (25). The gene content of adenovirids is similar to that of polintons, including pPolB, but they do not encode the typical genome packaging ATPase, which is replaced by a distinct ABC ATPase (26), nor RVE-INT and, indeed, do not integrate into host chromosomes.

The genome replication proteins, i.e., pPolB and Tlr1-like helicase, connect preplasmiviricots to eukaryotic nonviral mobile genetic elements (Fig. 1) (30). In particular, two distinct groups of linear plasmids replicating, respectively, in the cytoplasm and mitochondria of fungi and in mitochondria of plants, encode pPolBs (9), whereas transpovirons, a group of linear dsDNA elements parasitizing mimivirids (31, 32), encode Tlr1-like helicases. Finally, the polinton pPolB gene has been horizontally transferred to insect parvovirids yielding a group of viruses with linear single-stranded DNA (ssDNA) genomes classified into family *Bidnaviridae* (33).

The evolutionary trajectory that led to the emergence and diversification of eukaryotic preplasmiviricots and their nonviral relatives remains unresolved. Phylogenetic analysis of the pPolB sequences suggested that bacterial viruses of the *Tectiviridae* family (currently *Preplasmiviricota*: *Tectiliviricetes*) gave rise to two lineages of eukaryotic elements, mitochondrial linear plasmids and polintons, respectively, with all other groups of eukaryotic viruses and plasmids evolving directly from polintons (9). Alternatively, it has been suggested that

polintons evolved from mavirus-like virophages (20). Phylogenetic reconstructions based on the structural proteins yielded contrasting results, with polintons being at the base of a clade which included all eukaryotic viruses with DJR-MCPs in one analysis (34), but emerging as a terminal branch in another study, with virophages occupying the basal position (35). Given the fast evolution of capsid proteins and nonuniform selection pressures elicited by different hosts (e.g., exposure to humoral immunity), interpretation of the results of phylogenetic analyses based on structural proteins might not be straightforward.

Here, we explore the evolution of eukaryotic preplasmiviricots and related nonviral elements through detailed structural analysis of their major replication proteins. Structural dissection showed that the pPolB of polintons carries a terminal protein (TP) domain implicated in protein-primed replication initiation, followed by a vOTU-like cysteine deubiquitinylase domain likely responsible for cleavage of the TP from the rest of the pPolB polypeptide. Analysis of the presence/absence as well as inactivation of the protease domains in the homologs encoded by eukaryotic preplasmiviricots and related plasmids enabled us to establish the most likely sequence of events in their evolution.

Results and Discussion

Complex Domain Architecture of Eukaryotic Preplasmiviricot pPolBs. The pPolBs encoded by preplasmiviricots vary widely in length, with the proteins of polintons being nearly twice as long as the pPolBs of tectivirids and mavirus-like virophages (Fig. 2A). To clarify the relationship among these pPolB homologs, we modeled and compared the pPolB structures from representatives of each virus group, namely, bacterial tectivirid *Enterobacteria phage PRD1*, *Caenorhabditis briggsae* polinton 1 (P1-CB), frog adenovirus 1 (FrAdV-1), mavirus virophage, and *Bombyx mori* bidensovirus (BmBDV; *Bidnaviridae*), as well as those of mitochondrial and cytoplasmic linear plasmids (Fig. 2B). Analysis of the structural models showed that all the polymerases contain the characteristic TPR1 and TPR2 subdomains within their palm domain, confirming that they belong to the protein-primed subgroup of family B DNAPs. In the pPolB of the caudoviricete *Bacillus subtilis* phage phi29, the only pPolB for which the structure has been determined experimentally, TPR1 mediates interaction with the TP that primes genome replication, whereas TPR2 endows pPolB with the processivity and strand-displacement activities (36, 37). Notably, mavirus pPolB contained only a remnant of the exonuclease domain (Fig. 2B), which is conserved in all other family B DNAPs, suggesting that the mavirus polymerase has lower fidelity than other pPolBs. Unlike most other known pPolB-encoding viruses, mavirus encodes an S3H (Fig. 1), which likely unwinds DNA during genome replication.

All eukaryotic pPolBs contain extended N-terminal regions which have no counterparts in pPolBs of prokaryotic viruses. Structural analysis of these regions revealed that they encompass two globular domains, provisionally labeled D1 and D2 in Fig. 2B. D1 is absent in the adenovirid pPolB, whereas pPolBs of mitochondrial plasmids lack D2. Notably, the pPolB of yeast (*Pichia kluyveri*) plasmid kalilo instead of D2 contains an unrelated domain, DX, which is predicted to consist of ten β -strands with an alternating up-and-down orientation and has no significant similarity to proteins with known structures (Fig. 2B). Given that the mavirus pPolB contains both D1 and D2, but lacks the exonuclease domain, it appears most likely that pPolB of mavirus-like virophages is a derivative of polinton pPolBs.

The D1 Domain Is Homologous to the TP of PRD1-Like Tectivirids and Adenovirids. As the pPolB name gives away, protein-primed polymerases work in tandem with TPs that are covalently linked

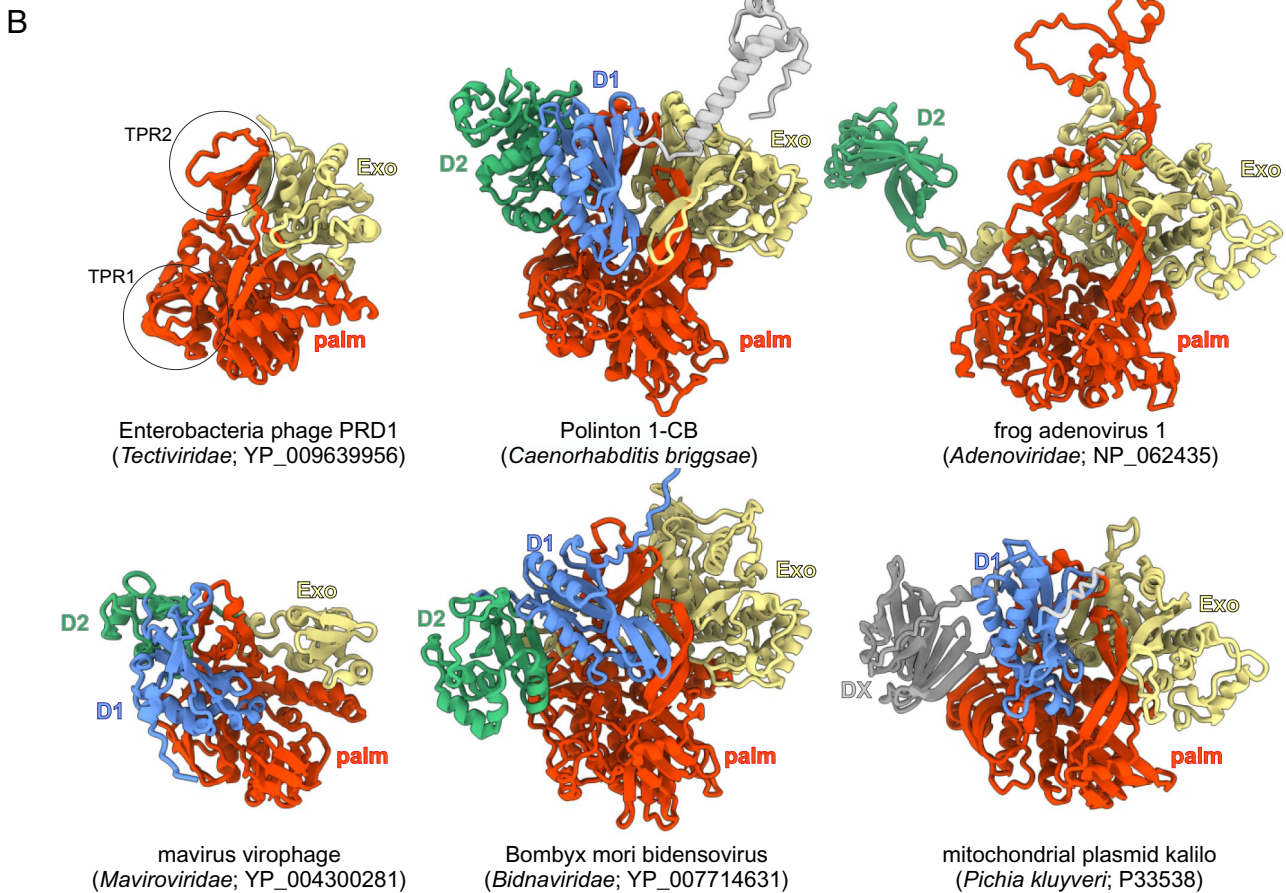
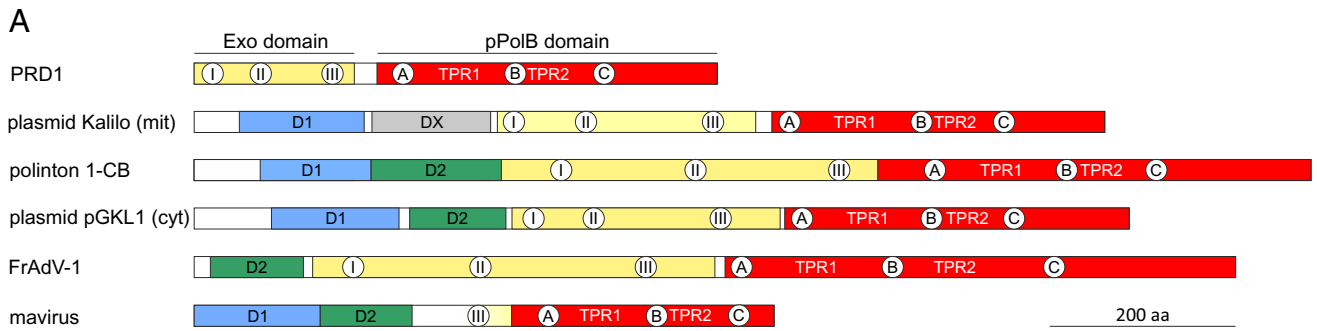


Fig. 2. The domain organization of pPolBs encoded by preplasmiviricots and related elements. (A) Schematic domain organization of pPolBs with homologous domains shown with matching colors. The locations of conserved motifs of the exonuclease and polymerization domains are indicated within the corresponding circles. The locations of the TPR1 and TPR2 subdomains are also shown. (B) Structural models of pPolBs encoded by different groups of viruses and plasmids, with distinct domains colored using the same scheme as in panel A. The TPR1 and TPR2 subdomains characteristic of pPolBs are encircled in the case of pPolB of phage PRD1. The models colored using pLDDT quality scores are shown in *SI Appendix, Fig. S1*, whereas the structural model of pPolB of yeast (*K. lactis*) plasmid pGKL1 is shown in *SI Appendix, Fig. S2*.

through a phosphodiester bond to a nucleotide that functions as a primer for DNA synthesis. All three alcoholic amino acids (Ser, Thr, Tyr) were experimentally identified as linking or priming residues in TPs associated with different pPolBs (38, 39). Tectivirid and adenovirid TPs are encoded as separate proteins [adenovirids actually encode a precursor of the TP that is proteolytically processed postreplication, late in infection (40, 41)] whereas in cytoplasmic linear plasmids, the TP is fused to pPolB and is proteolytically cleaved off following replication initiation by an unknown protease (42). The N-terminal regions of the polinton and bidnavirid pPolBs were similarly predicted to function as TPs (9, 33). TPs encoded by viruses from different families generally lack sequence similarity (39) and a high-resolution TP structure is only available for the caudoviricete *Bacillus subtilis* phage phi29 that is unrelated to preplasmiviricots (43). The phi29 TP has

an elongated α -helical fold (*SI Appendix, Fig. S3*) (43) and is clearly unrelated to the D1 and D2 domains of the multidomain eukaryotic pPolBs that are predicted to adopt α/β folds with a central β -sheet (Fig. 2B, see below). To explore the structural relationships among the TPs of preplasmiviricots, we modeled the structures of TP of tectivirids Enterobacteria phage PRD1 and Bacillus phage Bam35c, and the preTP (pTP, for simplicity referred to as TP hereinafter) of adenovirids FrAdV-1 and human adenovirus C (HAdV-C).

Unexpectedly, the experimentally characterized TPs of PRD1 (44) and Bam35c (45) were predicted to have different folds (*SI Appendix, Fig. S3*). Although both proteins encompass a central β -sheet surrounded by α -helices, the topology of the β -sheet domains is different. Notably, the N-terminal elongated α -helical region of the Bam35c TP resembles the corresponding region of

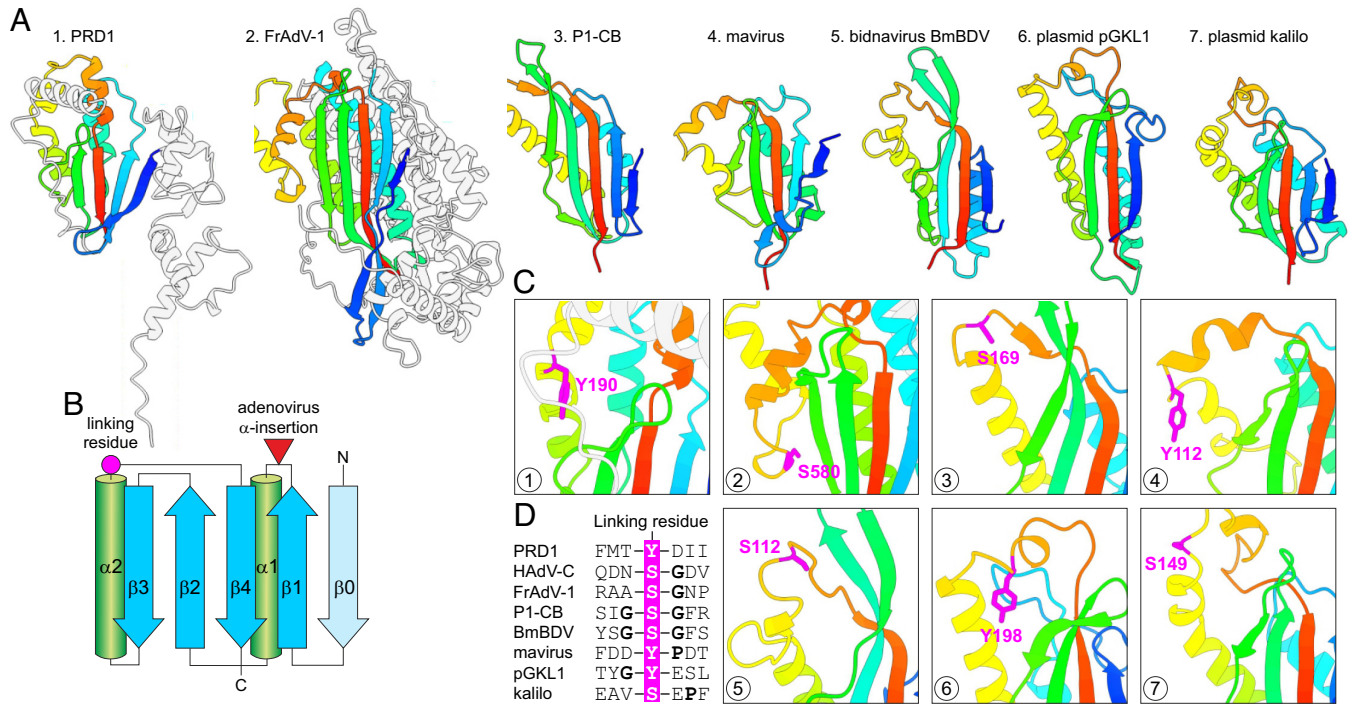


Fig. 3. Conservation of TPs. (A) Structural comparison of the TP of Enterobacteria phage PRD1 and pTP of frog adenovirus 1 (FrAdV-1) with D1s of pPolBs encoded by eukaryotic viruses and plasmids. Models are colored using the rainbow scheme from the N terminus (blue) to the C terminus (red). Nonconserved regions of the phage PRD1 and FrAdV-1 TPs are shown in gray. (B) Topology diagram of the PRD1-like TPs. The β -strand present in only some TPs is colored light blue and labeled β 0. The positions of the linking residue and large α -helical insertion in adenovirids are indicated. (C) Zoom-in on the (predicted) linking residues. The linking residues are shown in magenta using stick representation. Circled numbers in the corner of each box correspond to those in panel A, except for adenovirids (box 2), where a zoom-in on the experimentally characterized human adenovirus type C (HAdVC) pTP is shown. (D) Alignment of the regions flanking the predicted linking residues. Gly and Pro residues are shown in bold. Abbreviations: nematode [*C. briggsae* (Dougherty and Nigon, 1949)] polinton 1 (P1-CB); Bombyx mori bidensovirus (BmBDV).

the phi29 TP (*SI Appendix, Fig. S3A*) although no overall structural similarity between these domains was observed. By contrast, the TPs of phage PRD1 and FrAdV-1 are predicted to adopt the same fold, which is also found in the D1 domain conserved in the pPolBs of eukaryotic preplasmiviricots and plasmids (Fig. 3A). Indeed, pairwise comparison of the TP and D1 domain structures yielded significant DALI Z scores (*SI Appendix, Fig. S3B*). The TP/D1 fold consists of a 4-stranded β -sheet, which in some members is extended by an additional β -strand, and two α -helices (Fig. 3B). The TPs of both PRD1 and adenovirids contain N- and C-terminal α -helical extensions that are lacking in the D1 domains, suggesting that these regions mediate the interactions between the stand-alone TPs and the pPolB catalytic domain. In addition, adenovirid TPs contain an extensive α -helical insertion between β 1 and α 1, which is lacking in the PRD1 TP (Fig. 3A and B). The experimentally identified linking residues in the PRD1 (Tyr190) and adenovirid (Ser580) TPs occupy equivalent positions, between α 2 and the C-terminal β 4 strand, within non-structured loops. In all D1 domains, we identified either Ser or Tyr residues in positions equivalent to those of linking residues in the PRD1 and adenovirid TPs (Fig. 3C). Although the sequences surrounding the (predicted) linking residues were dissimilar, they all contained a helix-breaking Gly or Pro in the vicinity of the linking residue (Fig. 3D). By contrast, the negatively charged residues previously suggested to be a characteristic feature accompanying the linking residues based on the analysis of experimentally validated TPs (39) were not consistently present (Fig. 3D). Indeed, whereas the linking residue Ser580 in the human adenovirus (46) is flanked by Asp residues at the -2 and +2 positions, the homologous region in FrAdV-1 lacks both Asp residues, whereas Gly at the +1 position is conserved in both (Fig. 3D). Collectively, these observations suggest that D1 domains of

eukaryotic pPolBs are homologous to the TPs of tectivirids and adenovirids and accordingly are responsible for the priming of DNA synthesis, solving a long-standing puzzle in the field. Notably, the integrating viruses with pPolB, namely, polintons and mavirus-like virophages, should remove the TP prior to integration into the host chromosome; however, the mechanistic details of this process remain unclear.

PRD1-Like TPs Likely Evolved from the LF/PAD Domain of Family Y DNAPs. The lack of structural similarity between the PRD1, Bam35c, and phi29 TPs suggests that TPs have originated on at least three independent occasions. To gain insights into the provenance of the PRD1-like TPs, we performed searches queried with the structural models of TPs and pPolB D1 domains of preplasmiviricots and related elements against the PDB database using DALI. In all cases, the best hits with significant Z scores were to the little finger (LF) domain, known as the polymerase-associated domain (PAD) in eukaryotic proteins (*SI Appendix, Table S1*). LF/PAD is a unique noncatalytic domain exclusively found at the C terminus of family Y DNAPs, which are involved in mutagenic DNA repair in all domains of life (47–49). The LF/PAD is attached to the catalytic polymerase Y domain through an extended flexible linker and binds DNA across the major groove (47) (Fig. 4). It has been suggested that LF/PAD plays an important role in determining the enzymatic properties of the Y family DNAPs, modulating their processivity, ability to bypass template lesions, and capacity to generate base pair substitutions versus single-base deletions (50). Comparison of the PRD1-like TP models with the structures of the LF/PAD (Fig. 4) confirmed that the two groups of domains have the same fold consisting of a 4-stranded β -sheet and two α -helices (Fig. 3B). The conservation of the domain organization of the Y-family DNAPs across all

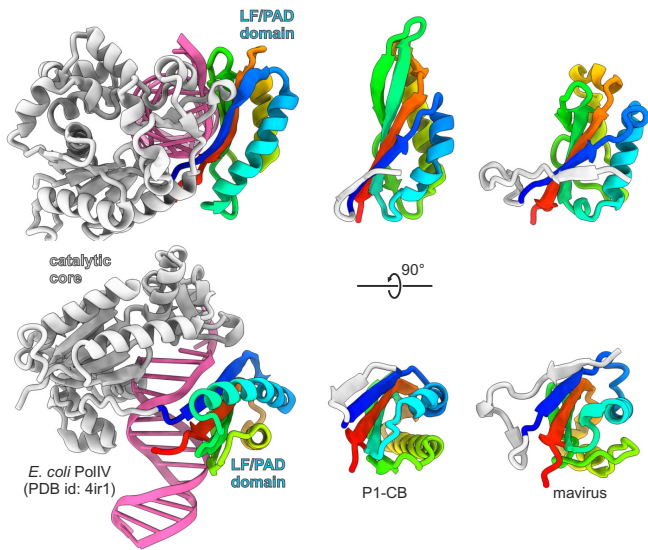


Fig. 4. Origin of PRD1-like TPs from the LF/PAD domain of family Y DNAPs. Structural comparison of the DNA-bound LF/PAD of the *Escherichia coli* DNAP IV (PolIV; PDB ID: 4IR1) with the TP domains of nematode [*C. briggsae* (Dougherty and Nigon, 1949)] polinton 1 (P1-CB) and mavirus virophage. The LF/PAD and TP domains are colored using the rainbow scheme from the N terminus (blue) to the C terminus (red), whereas the catalytic domain of PolIV and dsDNA are shown in gray and magenta, respectively.

domains of life testifies to the ancient association between the catalytic domain of Y-family DNAPs and LF/PAD, likely traceable to LUCA. Conversely, the sporadic distribution of LF/PAD-like TPs in bacterial viruses likely reflects a more recent recruitment. Thus, we hypothesize that the PRD1-like TP evolved from LF/PAD by gaining the ability to covalently bind the priming nucleotide, a facile adaptation given the simplicity of the sequence and structural contexts of the linking residue, which itself can vary (Fig. 3D). Notably, similar to Y family DNAPs, the TP domain is linked to the catalytic palm domain of pPolB in eukaryotic preplasmiviricots through an extended unstructured linker, which likely enables flexibility and movement. Conceivably, the preexisting capacity of the LF/PAD to interact with damaged DNA facilitated its recruitment as a TP to initiate replication of linear DNA genomes.

The D2 Domain Is a vOTU-Like Cysteine Protease. The fusion of the TP and catalytic pPolB domains raises the question whether and how the two domains are uncoupled during processive DNA synthesis, following the initiation of DNA replication. In the case of cytoplasmic linear plasmids, such as pGKL1 of *Kluyveromyces lactis* (yeast), it has been demonstrated that the 26 kDa N-terminal priming domain is proteolytically cleaved off by an unknown protease and stays covalently linked to the 5' of the nascent DNA strand (51). By contrast, in certain mitochondrial linear plasmids, such as the kalilo plasmid of *Pichia kluyveri* (yeast), the mass of the TP (120 kDa) corresponds to the entire DNAP, suggesting that the priming domain is not detached from the catalytic palm domain (52).

DALI searches queried with the structural models of the pPolB D2 domains from eukaryotic preplasmiviricots and related elements (Fig. 5A) yielded significant hits to diverse viral and cellular ovarian tumor domain (OTU) deubiquitinylases (SI Appendix, Table S1 and Fig. 5B), a superfamily of cysteine proteases with a distinct variant of the papain-like fold, also known as otubains (Fig. 5C) (53, 54). The viral OTU (vOTU) homologs process viral polyproteins and also function as deubiquitinylases counteracting antiviral defenses (55). We hypothesize that the vOTU domains

of pPolBs are responsible for the proteolytic detachment of the TP domains from the catalytic domains.

The invariant catalytic Cys and His residues located within the $\alpha 1$ and $\beta 4$ elements of the vOTUs, respectively (Fig. 5C), are conserved in polintons, mavirus-like virophages, and cytoplasmic plasmids (Fig. 5D), strongly suggesting that these are active proteases. To remove the TP domain from the palm domain, vOTU could cleave either immediately downstream of the TP domain or downstream of the vOTU itself. The experimentally determined size of the TP domain of the pPolB of *K. lactis* plasmid pGKL1 (51) is consistent only with the former scenario, i.e., cleavage between the TP and vOTU domains.

Although the pPolBs of adenovirids and bidnavirids contain vOTU-like domains, they encompass replacement of at least one of the two predicted active site residues across all known members of both these families of viruses (SI Appendix, Fig. S4). Conceivably, the split of the TP-encoding fragment into a stand-alone TP gene in adenovirids rendered the protease activity of vOTU domain obsolete, subsequently leading to its inactivation. By contrast, bidnavirids appear to employ a strategy akin to that of mitochondrial linear plasmids, whereby the entire pPolB remains covalently attached to the nascent DNA strand. Indeed, analysis of the structural proteins of bidnavirid particles showed the presence of a 118-kDa protein corresponding to pPolB (56), consistent with the inactivation of the vOTU domain in bidnavirids. Notably, the vOTU domain appears to be truncated in two mastadenoviruses (SI Appendix, Fig. S4A) and missing in a handful of GenBank entries. However, most of these truncations were corrected following resequencing of the virus genomes and more accurate prediction of the initiation codons (57). Thus, the inactivated vOTU domain was not lost in any known members of the *Adenoviridae* and *Bidnaviridae* families, suggesting that it performs a structural role in their genome replication. Notably, in adenovirids, the inactivated vOTU domain underwent further structural adaptations. In particular, the $\beta 4$ strand which harbors the catalytic histidine residue in active proteases is partially flipped out of the four-stranded β -sheet and contributes to formation of a second three-stranded β -sheet involving two N-terminal β -strands which are not part of the vOTU fold (SI Appendix, Fig. S5). The two β -sheets are roughly perpendicular to each other, forming a β -barrel not observed in other vOTU proteases.

Tlr1-Like Helicases Are Multidomain Proteins Containing pPolB Domains. We have previously observed that Tlr1-like helicases encoded by PLVs and transpovirons (Fig. 1) contain sequence motifs similar to the catalytic motifs of pPolBs (Fig. 6A) (27). The availability of powerful structure-prediction tools, such as AlphaFold2 (58), now enables scrutinizing this observation in greater detail and further exploring the domain organization of these large proteins. We modeled the structures of the Tlr1-like helicases of Tlr1 element (15) and transpoviron Mama associated with mimivirid *Acanthamoeba castellanii* mamavirus (31) (Fig. 6B and SI Appendix, Fig. S1C). The models confirmed that Tlr1-like helicases, in addition to S1H domains, contain palm domains with the characteristic DNAP catalytic motifs (Fig. 6C). Comparison of the palm domains of RNA-primed and protein-primed PolBs with the palm domains of Tlr1-like helicases unequivocally shows that the latter belong to the protein-primed variety and contain the diagnostic TPR1 and TPR2 subdomains (Fig. 6D). Previously, we suggested that changing the first of the two Asp residues in the polymerase C motif (DTD) involved in metal coordination would render the palm domain inactive (27). However, the same Asp residue is missing in the PolBs of mavirus virophage (Fig. 6C), mitochondrial circular plasmids (59), pipolins (60), and certain

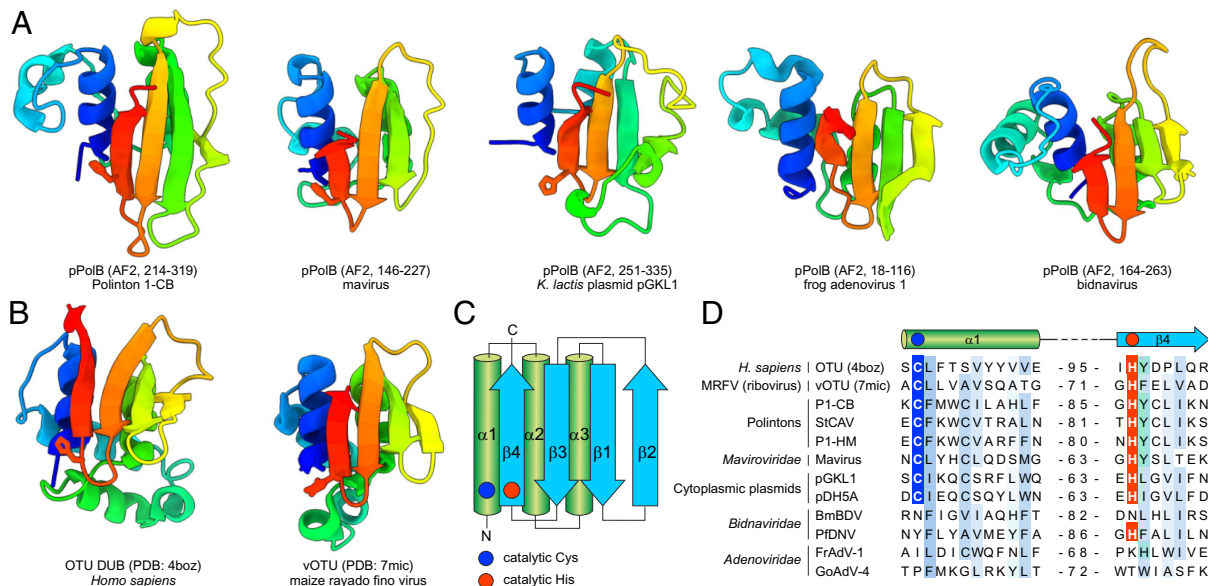


Fig. 5. Conservation of the vOTU-like cysteine protease domain in pPolBs. (A) Structural comparison of the vOTU domains present in pPolBs. (B) Crystal structures of OTU from humans (PDB: 4BOZ) and vOTU from an RNA virus, maize rayado fino virus (PDB: 7MIC). Models are colored using the rainbow scheme from the N terminus (blue) to the C terminus (red). When present, the catalytic Cys and His residues are shown using stick representation. (C) Topology diagram of the (v) OTU domains, with the catalytic Cys and His residues shown as blue and red circles, respectively. (D) Sequence alignment of the $\alpha 1$ and $\beta 4$ elements harboring the catalytic Cys (blue background) and His (red background) residues, respectively. The numbers in the middle denote the spacing between the two structural elements. BmBDV, *Bombyx mori* bidensovirus; FrAdV-1, frog adenovirus 1; GoAdV-4, goose adenovirus 4; P1-CB, nematode [*C. briggsae* (Dougherty and Nigon, 1949)] polinton 1; P1-HM, hydra (*Hydra vulgaris* Pallas, 1766) polinton 1; pDH5A, yeast (*Debaryomyces hansenii*) plasmid pDH5A; pGKL1, yeast (*K. lactis*) plasmid pGKL1; StCAV, Stylophora coral adintovirus; PfDNV, *Phylloscopus fuscatus* densovirus.

archaeal DNAPs (61), all of which are nevertheless catalytically active. By contrast, motifs A and B are conserved. Thus, we predict that Tlr1-like helicases possess DNAP activity and are functionally analogous to other polymerase-helicase fusion proteins, such as PolA-S3H and AEP-S3H.

Phylogenetic analysis of the palm domain of Tlr1-like helicases suggested that it evolved from polinton pPolBs (27). Similar to the latter, Tlr1-like helicases contain extended N-terminal regions (Fig. 6A) that thus far evaded functional annotation. Analysis of the structural models of the PLV-Tlr1 and transpoviron Mama helicases showed that both contain three globular domains upstream of the palm domain. The N-terminal domains are highly variable in different PLV and transpoviron helicases and none are similar to proteins with known structures in the PDB database. By contrast, the domains proximal to the palm domains were identified as vOTUs (SI Appendix, Fig. S6A), with both catalytic Cys and His residues conserved in all analyzed PLV and transpoviron sequences (SI Appendix, Fig. S6B), suggesting that these vOTUs are active proteases, likely involved in proteolytic processing of the Tlr1-like helicases, as proposed above for the pPolBs of eukaryotic preplasmiviricots. Unexpectedly, the topology of the middle domain, occupying a position equivalent to that of the TP in pPolBs was variable, with at least three different variants among PLVs, although all were predicted to adopt α/β folds, with central β -sheets (SI Appendix, Fig. S7). In particular, the corresponding domain in Tlr1 was unique, whereas the one in transpoviron Mama and some PLVs have a distinct fold resembling the Bam35c TP, although convergence in this case cannot be excluded. Notably, TvV-S1 contains a bona fide PRD1-like TP domain, suggesting that in this virus the middle domain functions as TP. Given that in S1H phylogenies, TvV-S1 occupied a basal position with respect to other PLVs and transpovirons (27), the PRD1-like TP domain is the likely ancestral state in the evolution of PLVs that was subsequently replaced or permuted into structurally distinct but functionally equivalent domains. Finally, at the C terminus, Tlr1-like helicases of different PLVs contain homing endonucleases

of two distinct families, GIY-YIG and HNH (Fig. 6A). Notably, polinton pPolBs also contain diverse homing endonuclease domains inserted at different positions of the protein (10), suggesting that these insertions in pPolB and Tlr1-like helicases result from intein invasions.

Evolutionary History of Preplasmiviricota. Dissection of the pPolB and Tlr1-like helicase structures described here prompted a scenario for the evolution of preplasmiviricots and related nonviral elements (Fig. 7). The presence of structurally distinct TPs in tectivirids and their uniformity in most of the eukaryotic pPolBs suggest that eukaryotic preplasmiviricots evolved from a particular subgroup of tectivirids with PRD1-like TPs. The fusion of the TP domain to the polymerase domain in all eukaryotic pPolBs except for adenovirids (see below), along with the association between preplasmiviricots and eukaryotes being traceable to the last eukaryotic common ancestors (LECA) (3), suggests that the fusion occurred early during eukaryogenesis or possibly in a yet undiscovered group of bacterial tectivirids that includes direct ancestors of eukaryotic preplasmiviricots. Phylogenetic analyses of eukaryotic pPolBs consistently suggest that the first split occurred between mitochondrial linear plasmids and the ancestor of the remaining eukaryotic pPolBs (Fig. 7) (9) and this conclusion is supported by our structural comparisons. Conceivably, this split coincided with the escape of the ancestor of polintons from the proto-mitochondrial bacterial endosymbiont into the cytoplasm of the emerging proto-eukaryote. In some mitochondrial linear plasmids [e.g., in rapeseed (*Brassica napus* L.) (62)], the TP domain is fused directly to pPolB, whereas in others (e.g., the kalilo plasmid of *Pichia kluyveri*) the two domains are separated by an additional domain (DX) of unknown function. The vOTU domain was recruited subsequently in only one of the two major branches of preplasmiviricots, the polintons, yielding a multidomain pPolB with the TP-vOTU-EXO-PALM domain organization that is also shared by cytoplasmic linear plasmids. Thus, polintons appear to have been the first group of eukaryotic preplasmiviricots to evolve

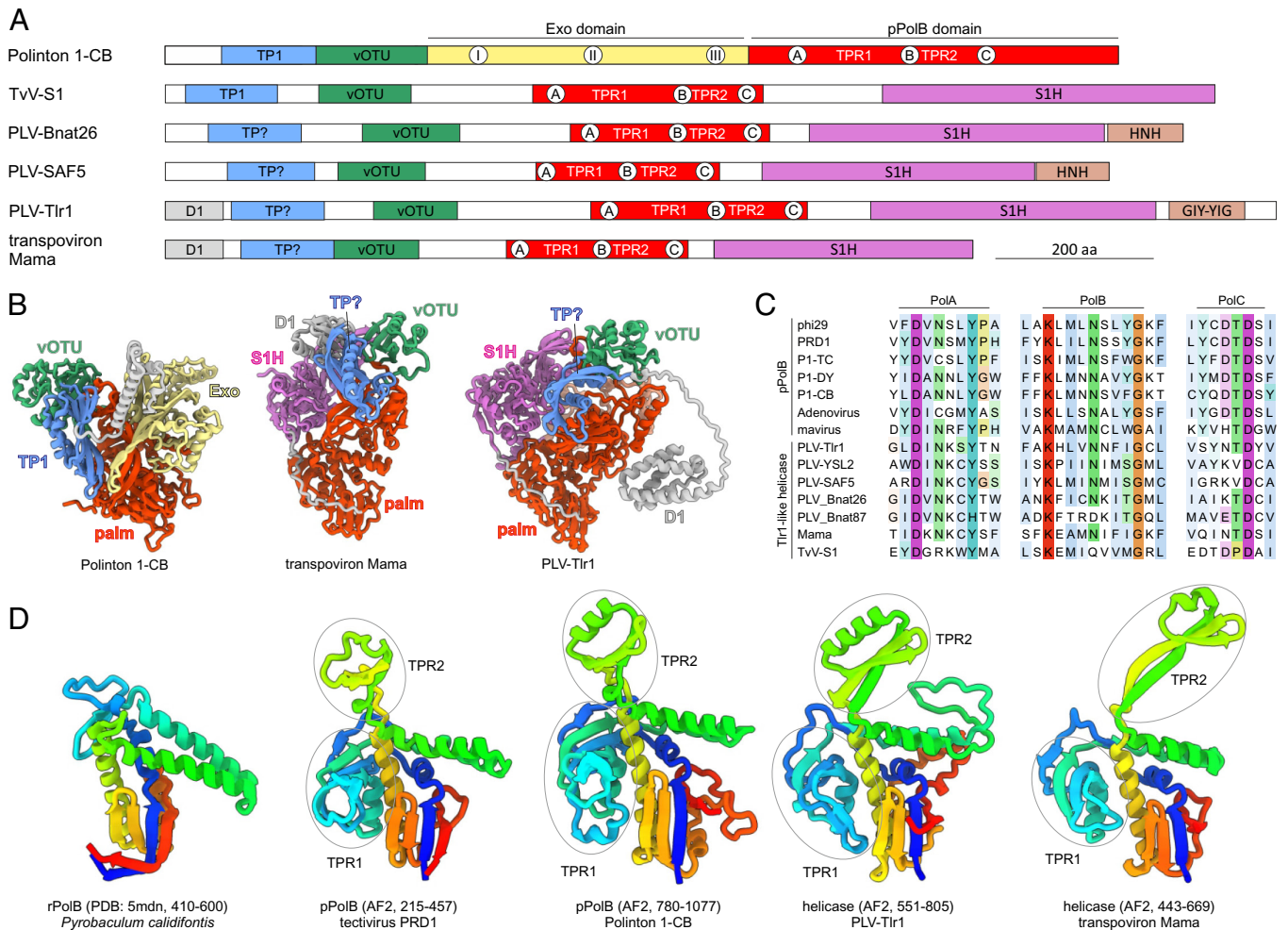


Fig. 6. The domain organization of Tlr1-like helicases. (A) Schematic domain organization of Tlr1-like helicases encoded by PLVs and transposons with homologous domains shown with matching colors. The annotated domain organization of the pPolB of nematode [*C. briggsae* (Dougherty and Nigon, 1949)] polinton 1 (P1-CB) is shown for comparison. The locations of conserved motifs of the exonuclease and polymerization domains are indicated within the corresponding circles. The locations of the TPR1 and TPR2 subdomains are also shown. Question marks denote uncertainty. (B) Structural models of Tlr1-like helicases encoded by transposon Mama and Tlr1 element, with distinct domains colored using the same scheme as in panel A. The model of pPolB of polinton 1-CB is shown for comparison. The models colored using the pLDDT quality scores are shown in *SI Appendix, Fig. S1*. (C) Sequence alignment of the motifs, PolA-C, conserved in family B DNAPs. (D) Structural comparison of the palm domains of RNA-primed and protein-primed family B DNAPs (rPolBs and pPolBs, respectively) with those of Tlr1-like helicases. Models are colored using the rainbow scheme from the N terminus (blue) to the C terminus (red). The TPR1 and TPR2 subdomains characteristic of protein-primed polymerases are circled. The coordinates of the depicted regions are indicated in parentheses.

from tectivirids, subsequently giving rise to the entire diversity of the eukaryotic preplasmiviricots. Phylogenetic analyses have confidently shown that cytoplasmic linear plasmids are derived from (proto)polintons through the loss of the morphogenetic module, switching to a nonviral mode of propagation (9).

Adenovirids encode a PRD1-like TP as a stand-alone protein, resembling the organization observed in bacterial viruses. However, the presence of inactivated vOTU domains at the N terminus of their pPolBs leaves no doubt that adenovirids evolved from polintons rather than inherited the split TP organization from bacterial viruses. The inactivated vOTU domain in adenovirids could have been repurposed for the interaction between the TP and pPolB or other replication factors. Notably, the split between the TP and vOTU domains seems to have occurred independently in different groups of polintons. For example, we identified a polinton in hood corals (*Stylophora pistillata* Esper, 1797) (8) (*SI Appendix, Fig. S8*) in which the TP is encoded by a separate gene, whereas the catalytically active vOTU domain is fused to pPolB, as in adenovirids (Fig. 5D). Thus, adenovirids might have evolved from a group of polintons in which the TP was already detached from pPolB. The vOTU domain was also inactivated in bidnavirids which, however,

retained the TP domain fused to the pPolB. Consistently, pPolB has been identified in the virions of bidnavirids (56), suggesting that, like in the case of some mitochondrial linear plasmids that lack the vOTU domain altogether, the pPolB is not processed and stays attached to the genomic DNA.

Polintons are also at the root of mavirus-like virophages and PLVs. The mavirus pPolB has a unique TP-vOTU-PALM domain organization, indicating that it is a highly derived DNAP that lost most of the exonuclease domain, ruling out the possibility that the mavirus pPolB is ancestral to that encoded by polintons. Nevertheless, given the high divergence of the virophage MCPs, it appears likely that virophages have diverged at an early stage of the polinton radiation (Fig. 7). The mavirus-like pPolB resembles the ancestor of the Tlr1-like helicases widespread in PLVs, which have the TP-vOTU-PALM-S1H domain organization. We predict that the palm domains of these proteins are active DNAPs, and accordingly, they should be more accurately denoted as pPolB-S1H fusion proteins, akin to PolA-S3H and AEP-S3H widespread in other viruses, rather than Tlr1-like helicases.

Notably, the loss of the exonuclease domain is accompanied by the gain of a helicase, S3H in maverisks (encoded by a

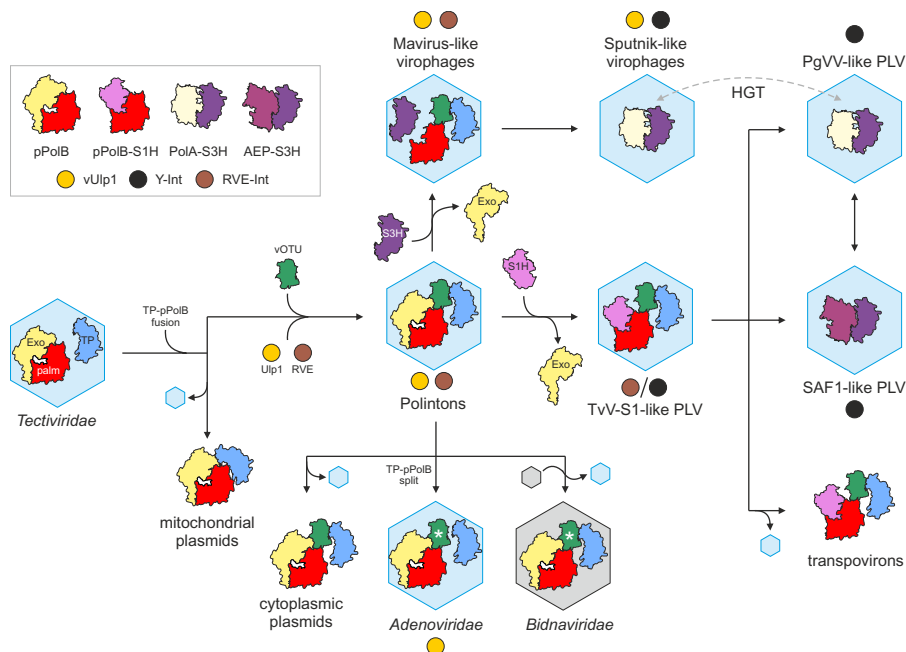


Fig. 7. Natural history of preplasmiviricots and derived eukaryotic elements. The schematic depicts a proposed scenario for the evolution of eukaryotic preplasmiviricots and related elements from a bacterial tectivirid ancestor. Blue hexagons represent icosahedral capsids built from DJR-MCPs. The gray hexagon indicates replacement of the DJR-MCP with an unrelated tectivirid protein in bidnavirids. Black and gray arrows indicate evolution and horizontal gene transfer (HGT), respectively. Asterisks indicate inactivation of vOTU domains. AEP, archaeo-eukaryotic primase-polymerase; Exo, exonuclease; PgVV, *Phaeocystis globosa* virus virophage; PLV, polinton-like virus; PolA, family A DNAP; pPolB, protein-primed family B DNAP; RVE-Int, retrovirid-like integrase; S1H, superfamily 1 helicase; S3H, superfamily 3 helicase; TP, terminal protein; TvV-S1, *Tetraselmis viridis* virus S1; (v)Ulp1, (viral) Ulp1-family cysteine protease; Y-Int, tyrosine superfamily integrase.

separate gene; *SI Appendix, Fig. S8*) and S1H fused to pPolB in PLVs. The two other polymerase-helicase fusions common in eukaryotic preplasmiviricots, AEP-S3H and PolA-S3H, also lack the exonuclease domain. Conversely, PolA encoded by diverse phages (e.g., T7) lack the helicase domain but contain the exonuclease domain (63). Thus, there seems to be a compensatory relationship between the exonuclease and helicase activities such that both increase the polymerase fidelity via different mechanisms. The presence of distinct helicases, S3H and S1H in mavirus and PLVs, respectively, and the pronounced divergence between the corresponding MCPs suggest that the exonuclease domains have been independently lost in the two virus groups. These findings clarify the position of PLVs and virophages in the evolution of *Preplasmiviricota*, suggesting that both virus groups diverged from a polinton-like ancestor with a bona fide pPolB.

The loss of the structural module yielded transpovirons, all of which encode pPolB-S1H replication proteins and thus apparently evolved from PLVs. The link between virophages, PLVs, and transpovirons also extends to their mode of propagation. All known virophages and transpovirons and at least some PLVs, such as *Phaeocystis globosa* virus virophage (PgVV) (17, 64), are parasites of giant viruses of phylum *Nucleocytoviricota*, suggesting that this interaction emerged early in the evolution of eukaryotic viruses. As noted above, besides the pPolB-S1H, some virophages and PLVs encode PolA-S3H or AEP-S3H replication proteins, respectively. In phylogenetic analyses based on the structural module, mavirus-like virophages, which encode pPolB lacking the exonuclease domain (see above), are consistently basal to virophages encoding PolA-S3H (34, 35), suggesting that pPolB is the ancestral replication protein which was replaced in some virophage lineages. Indeed, the unique PolA-S3H fusion protein first discovered in sputnikviruses (24) has now been identified also in diverse PLVs (12), suggesting horizontal exchange of the replication and structural modules between PLVs and virophages.

Taken together, these results clarify the evolutionary history of *Preplasmiviricota* and reassert the central position of polintons in the evolution of eukaryotic viruses, plasmids, and transpovirons. The revised scenario outlined here calls for refinement of the *Preplasmiviricota* taxonomy.

Materials and Methods

Protein Structure Prediction and Analysis. Structural models of pPolBs encoded by mitochondrial and cytoplasmic plasmids were downloaded from the AlphaFold database (65). The structural model of the pPolB of polinton 1 from a nematode [*Caenorhabditis briggsae* (Dougherty and Nigon, 1949)] was taken from a previous study (10). The remaining protein structures were modeled using AlphaFold2 (58) or DeepFold (66) through ColabFold v1.5.5 (67) "alphafold2_multimer_v3" model with six recycles. To improve the quality of the model, structural templates were provided when available. Structures were searched against the PDB database using DALI (68). Structural similarity between the TPs was assessed using the all-against-all comparison with DALI (68). Comparison of predicted and experimentally resolved structures from PDB was also performed using MatchMaker (69). Protein structures and structural models were visualized using ChimeraX v1.7.1 (70).

Data, Materials, and Software Availability. All protein structural models generated in this study are available for download from Zenodo (<https://zenodo.org/doi/10.5281/zenodo.10962812>) (71).

ACKNOWLEDGMENTS. This work was supported by a grant from l'Agence Nationale de la Recherche (ANR-23-CE02-0022) to M.K. E.V.K. is supported through the Intramural Research Program of the NIH (National Library of Medicine). This work was supported in part through a Laulima Government Solutions, Limited Liability Company (LLC), prime contract with the U.S. National Institute of Allergy and Infectious Diseases under Contract No. HHSN272201800013C. J.H.K. performed this work as an employee of Tunnell Government Services, a subcontractor of Laulima Government Solutions, LLC, under Contract No. HHSN272201800013C. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Health and Human Services or of the institutions and companies affiliated with the authors.

1. E. V. Koonin *et al.*, Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* **84**, e00061-19 (2020).
2. M. Krupovic, V. V. Dolja, E. V. Koonin, The LUCA and its complex virome. *Nat. Rev. Microbiol.* **18**, 661-670 (2020).
3. M. Krupovic, V. V. Dolja, E. V. Koonin, The virome of the last eukaryotic common ancestor and eukaryogenesis. *Nat. Microbiol.* **8**, 1008-1017 (2023).
4. V. V. Kapitonov, J. Jurka, Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4540-4545 (2006).
5. E. J. Pritham, T. Putliwala, C. Feschotte, Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **390**, 3-17 (2007).
6. M. Krupovic, D. H. Bamford, E. V. Koonin, Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. *Biol. Direct* **9**, 6 (2014).
7. M. Krupovic, E. V. Koonin, Self-synthesizing transposons: Unexpected key players in the evolution of viruses and defense systems. *Curr. Opin. Microbiol.* **31**, 25-33 (2016).
8. G. J. Starrett *et al.*, Adintoviruses: A proposed animal-tropic family of midsize eukaryotic linear dsDNA (MELD) viruses. *Virus Evol.* **7**, veaa055 (2021).
9. M. Krupovic, E. V. Koonin, Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat. Rev. Microbiol.* **13**, 105-115 (2015).
10. D. E. Jeong *et al.*, DNA polymerase diversity reveals multiple incursions of polintons during nematode evolution. *Mol. Biol. Evol.* **40**, msad274 (2023).
11. N. Yutin, S. Shevchenko, V. Kapitonov, M. Krupovic, E. V. Koonin, A novel group of diverse polinton-like viruses discovered by metagenome analysis. *BMC Biol.* **13**, 95 (2015).
12. C. M. Bellas, R. Sommaruga, Polinton-like viruses are abundant in aquatic ecosystems. *Microbiome* **9**, 13 (2021).
13. C. Bellas *et al.*, Large-scale invasion of unicellular eukaryotic genomes by integrating DNA viruses. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2300465120 (2023).
14. N. S. Patil, K. M. Karrer, A developmentally regulated deletion element with long terminal repeats has cis-acting sequences in the flanking DNA. *Nucleic Acids Res.* **28**, 1465-1472 (2000).
15. J. D. Wuitschick, J. A. Gershan, A. J. Lochowicz, S. Li, K. M. Karrer, A novel family of mobile genetic elements is limited to the germline genome in *Tetrahymena thermophila*. *Nucleic Acids Res.* **30**, 2524-2537 (2002).
16. A. Pagarete, T. Grebert, O. Stepanova, R. A. Sandaa, G. Bratbak, Tsv-N1: A novel DNA algal virus that infects *Tetraselmis striata*. *Viruses* **7**, 3937-3953 (2015).
17. S. Roitman *et al.*, Isolation and infection cycle of a polinton-like virus virophage in an abundant marine alga. *Nat. Microbiol.* **8**, 332-346 (2023).
18. D. V. Sizov, V. P. Polishchuk, Cultivation, purification and crystallization of virus of green algae *Tetraselmis viridis*. *Biopolym. Cell* **22**, 243-245 (2006).
19. B. La Scola *et al.*, The virophage as a unique parasite of the giant mimivirus. *Nature* **455**, 100-104 (2008).
20. M. G. Fischer, C. A. Suttle, A virophage at the origin of large DNA transposons. *Science* **332**, 231-234 (2011).
21. N. Yutin, D. Raoult, E. V. Koonin, Virophages, polintons, and transpovirons: A complex evolutionary network of diverse selfish genetic elements with different reproduction strategies. *Virus J.* **10**, 158 (2013).
22. M. G. Fischer, The virophage family *Lavidaviridae*. *Curr. Issues Mol. Biol.* **40**, 1-24 (2021).
23. M. G. Fischer, Giant viruses come of age. *Curr. Opin. Microbiol.* **31**, 50-57 (2016).
24. L. M. Iyer, S. Abhiman, L. Aravind, A new family of polymerases related to superfamily A DNA polymerases and T7-like DNA-dependent RNA polymerases. *Biol. Direct* **3**, 39 (2008).
25. B. Harrach, Z. L. Tarjan, M. Benko, Adenoviruses across the animal kingdom: A walk in the zoo. *FEBS Lett.* **593**, 3660-3673 (2019).
26. A. M. Burroughs, L. M. Iyer, L. Aravind, Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn.* **3**, 48-65 (2007).
27. M. Krupovic, N. Yutin, E. V. Koonin, Fusion of a superfamily 1 helicase and an inactivated DNA polymerase is a signature of common evolutionary history of polintons, polinton-like viruses, TIR1 transposons and transpovirons. *Virus Evol.* **2**, vew019 (2016).
28. M. Krupovic, K. S. Makarova, P. Forterre, D. Prangishvili, E. V. Koonin, Casposons: A new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol.* **12**, 36 (2014).
29. M. Krupovic, P. Béguin, E. V. Koonin, Casposons: Mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* **38**, 36-43 (2017).
30. E. V. Koonin, M. Krupovic, Polintons, virophages and transpovirons: A tangled web linking viruses, transposons and immunity. *Curr. Opin. Virol.* **25**, 7-15 (2017).
31. C. Desnues *et al.*, Virophages and transpovirons as the diverse mobilome of giant viruses. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18078-18083 (2012).
32. S. Jeudy *et al.*, Exploration of the propagation of transpovirons within *Mimiviridae* reveals a unique example of commensalism in the viral world. *ISME J.* **14**, 727-739 (2020).
33. M. Krupovic, E. V. Koonin, Evolution of eukaryotic single-stranded DNA viruses of the *Bidnaviridae* family from genes of four other groups of widely different viruses. *Sci. Rep.* **4**, 5347 (2014).
34. A. C. Woo, M. Gaia, J. Guglielmini, V. Da Cunha, P. Forterre, Phylogeny of the *Varidnaviria* morphogenesis module: Congruence and incongruence with the tree of life and viral taxonomy. *Front. Microbiol.* **12**, 704052 (2021).
35. J. G. N. Barreat, A. Katzourakis, A billion years arms-race between viruses, virophages, and eukaryotes. *Life* **12**, RP86617 (2023).
36. E. Dufour *et al.*, An aspartic acid residue in TPR-1, a specific region of protein-priming DNA polymerases, is required for the functional interaction with primer terminal protein. *J. Mol. Biol.* **304**, 289-300 (2000).
37. M. Salas, M. de Vega, Protein-primed replication of bacteriophage Phi29 DNA. *Enzymes* **39**, 137-167 (2016).
38. M. Salas, Protein-priming of DNA replication. *Annu. Rev. Biochem.* **60**, 39-71 (1991).
39. M. Redrejo-Rodríguez, M. Salas, Multiple roles of genome-attached bacteriophage terminal proteins. *Virology* **468-470**, 322-329 (2014).
40. D. M. Rekosh, W. C. Russell, A. J. Bellet, A. J. Robinson, Identification of a protein linked to the ends of adenovirus DNA. *Cell* **11**, 283-295 (1977).
41. A. Webster, I. R. Leith, J. Nicholson, J. Hounsell, R. T. Hay, Role of preterminal protein processing in adenovirus replication. *J. Virol.* **71**, 6381-6389 (1997).
42. R. Klassen, F. Meinhardt, "Linear protein-primed replicating plasmids in eukaryotic microbes" in *Microbial Linear Plasmids. Microbiology Monographs*, F. Meinhardt, R. Klassen, Eds. (Springer, Berlin, Heidelberg, 2007), vol. 7, pp. 187-226.
43. S. Kamtekar *et al.*, The phi29 DNA polymerase: protein-primer structure suggests a model for the initiation to elongation transition. *EMBO J.* **25**, 1335-1343 (2006).
44. S. Y. Shue, J. C. Hsieh, J. Ito, Mapping of the DNA linking tyrosine residue of the PRD1 terminal protein. *Nucleic Acids Res.* **19**, 3805-3810 (1991).
45. M. Berjón-Otero, L. Villar, M. Salas, M. Redrejo-Rodríguez, Disclosing early steps of protein-primed genome replication of the Gram-positive tectivirus Bam35. *Nucleic Acids Res.* **44**, 9733-9744 (2016).
46. J. E. Smart, B. W. Stillman, Adenovirus terminal protein precursor. Partial amino acid sequence and the site of covalent linkage to virus DNA. *J. Biol. Chem.* **257**, 13499-13506 (1982).
47. J. D. Pata, Structural diversity of the Y-family DNA polymerases. *Biochim. Biophys. Acta* **1804**, 1124-1135 (2010).
48. W. Yang, An overview of Y-family DNA polymerases and a case study of human DNA polymerase eta. *Biochemistry* **53**, 2793-2803 (2014).
49. K. Timinskas, C. Venclovas, New insights into the structures and interactions of bacterial Y-family DNA polymerases. *Nucleic Acids Res.* **47**, 4393-4405 (2019).
50. F. Boudsocq *et al.*, Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. *J. Biol. Chem.* **279**, 32932-32940 (2004).
51. M. Takeda, H. Hiraishi, T. Takesako, S. Tanase, N. Gunge, The terminal protein of the linear DNA plasmid pGKL2 shares an N-terminal domain of the plasmid-encoded DNA polymerase. *Yeast* **12**, 241-246 (1996).
52. P. J. Vierula, H. Bertrand, A deletion derivative of the kalilo senescence plasmid forms hairpin and duplex DNA structures in the mitochondria of *Neurospora*. *Mol. Gen. Genet.* **234**, 361-368 (1992).
53. K. S. Makarova, L. Aravind, E. V. Koonin, A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem. Sci.* **25**, 50-52 (2000).
54. M. Y. Balakirev, S. O. Tcherniuk, M. Jaquinod, J. Chroboczek, Otubains: A new family of cysteine proteases in the ubiquitin pathway. *EMBO Rep.* **4**, 517-522 (2003).
55. C. Lombardi *et al.*, A compact viral processing proteinase/ubiquitin hydrolase from the OTU family. *PLoS Pathog.* **9**, e1003560 (2013).
56. T. Hayakawa *et al.*, Analysis of proteins encoded in the bipartite genome of a new type of parvo-like virus isolated from silkworm-Structural protein with DNA polymerase motif. *Virus Res.* **66**, 101-108 (2000).
57. A. J. Davison, M. Benko, B. Harrach, Genetic content and evolution of adenoviruses. *J. Gen. Virol.* **84**, 2895-2908 (2003).
58. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
59. Q. Li, F. E. Nargang, Two *Neurospora* mitochondrial plasmids encode DNA polymerases containing motifs characteristic of family B DNA polymerases but lack the sequence Asp-Thr-Asp. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4299-4303 (1993).
60. M. Redrejo-Rodríguez *et al.*, Primer-independent DNA synthesis by a family B DNA polymerase from self-replicating mobile genetic elements. *Cell Rep.* **21**, 1574-1587 (2017).
61. D. Kazlauskas, M. Krupovic, J. Guglielmini, P. Forterre, C. Venclovas, Diversity and evolution of B-family DNA polymerases. *Nucleic Acids Res.* **48**, 10142-10156 (2020).
62. H. Handa, K. Itani, H. Sato, Structural features and expression analysis of a linear mitochondrial plasmid in rapeseed (*Brassica napus* L.). *Mol. Genet. Genomics* **267**, 797-805 (2002).
63. D. Czerneki, A. Nourisson, P. Legrand, M. Delarue, Reclassification of family A DNA polymerases reveals novel functional subfamilies and distinctive structural features. *Nucleic Acids Res.* **51**, 4488-4507 (2023).
64. S. Santini *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. U.S.A.* **110**, 10800-10805 (2013).
65. M. Varadi *et al.*, AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439-D444 (2022).
66. J. W. Lee *et al.*, DeepFold: Enhancing protein structure prediction through optimized loss functions, improved template features, and re-optimized energy function. *Bioinformatics* **39**, btad712 (2023).
67. M. Mirdita *et al.*, ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**, 679-682 (2022).
68. L. Holm, DALI and the persistence of protein shape. *Protein Sci.* **29**, 128-140 (2020).
69. E. C. Meng, E. F. Pettersen, G. S. Couch, C. C. Huang, T. E. Ferrin, Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* **7**, 339 (2006).
70. E. C. Meng *et al.*, UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
71. M. Krupovic, Dataset "Structural models of viral proteins described in Krupovic M, *et al.*, Proc Natl Acad Sci U S A " Zenodo. <https://doi.org/10.5281/zenodo.10962813>. Deposited 11 April 2024.