Trends in Microbiology (Review)

Paleovirology of the DNA viruses of eukaryotes

Jose Gabriel Nino Barreat[1], Aris Katzourakis[1]

[1]Department of Zoology, University of Oxford, Oxford, OX1 3SY, United Kingdom.

*Correspondence: aris.katzourakis@zoo.ox.ac.uk (A. Katzourakis)

Key words

DNA viruses, endogenous viral elements (EVEs), eukaryotes, paleovirology, transposons.

Abstract

Paleovirology is the study of ancient viruses and how they have coevolved with their hosts. An increasingly detailed understanding of the diversity, origins and evolution of the DNA viruses of eukaryotes has been obtained through the lens of paleovirology in recent years. Members of multiple viral families have been found integrated in the genomes of eukaryotes, providing a rich fossil record to study. These elements have extended our knowledge of exogenous viral diversity, host ranges, the timing of viral evolution and are revealing the existence of entire new families of eukaryotic integrating dsDNA viruses and transposons. Future work in paleovirology will continue to provide insights into antiviral immunity, viral diversity, potential applications and reveal other secrets of the viral world.

Main text

Viruses (see Glossary) are one of the dominant biological entities on Earth, containing a significant portion of the genetic diversity in the biosphere [1] and outnumbering cells by a factor of 2.6-160 in many ecosystems [2]. They are an ancient component of the Earth's biomes and may have driven major evolutionary transitions by the interaction with their hosts [3,4]. DNA viruses comprise an enormous diversity of viruses infecting eukaryotes. They range in size from the porcine circoviruses (some of the smallest viruses known), to the **giant viruses** that infect microbial eukaryotes [5,6]. Their genomes are also diverse, they can be single or double-stranded, circular or linear DNA molecules, and two families undergo **reverse transcription** [7–9]. Their close relationship as obligate intracellular parasites of eukaryotes has translated into frequent genetic exchanges throughout evolutionary time. In particular, the large genome sizes of eukaryotes, have meant that ancestral viral integrations can accumulate in the genomes of their hosts [10], forming a rich genomic fossil record and providing an opportunity to study the deep viral past. In this review, we will highlight key achievements and exciting discoveries in the paleovirology of the DNA viruses of eukaryotes, which are leading to novel insights and increasing our understanding of their diversity, origins and evolution. We will focus on DNA viruses since they have received less attention than their RNA counterparts in terms of paleovirology, even though classic work on DNA viruses laid the foundations of this science more than 70 years ago [11,12].

Paleovirology: probing virus origins and evolution

Paleovirology is the study of ancient viruses over macroevolutionary timescales. It is concerned with how major virus lineages have originated, diversified and coevolved with their hosts. Paleovirology relies mostly on comparative genomics and phylogenetics of extant viral sequences to draw inferences about the viral past, however, there is indirect evidence of viruses in the fossil record and viral sequences that have been recovered from ancient DNA (aDNA) (Box 1). One of the major challenges in paleovirology is the sequence divergence between related viruses. Over long timescales (hundreds of millions of years), viral sequences can diverge beyond recognition, but this can be overcome sometimes by performing comparisons of protein structure which is more conserved during evolution [13,14].

Our knowledge in paleovirology comes mostly from the study of **endogenous viral elements (EVEs)**. EVEs are viral integrations that have entered the germline of their hosts and are inherited as host alleles; their survival is determined by the balance between genetic drift and natural selection [15] (Figure 1). On some occasions EVEs are repurposed to perform a beneficial host function and become **exaptations** [4]. An important property of EVEs is that they can be used to calibrate the timing of virus evolution: if an EVE is orthologous across a number of species, this gives a minimum estimate for the age of the virus that integrated into the genome [16]. We can therefore think of EVEs as a genomic fossil record of viruses.

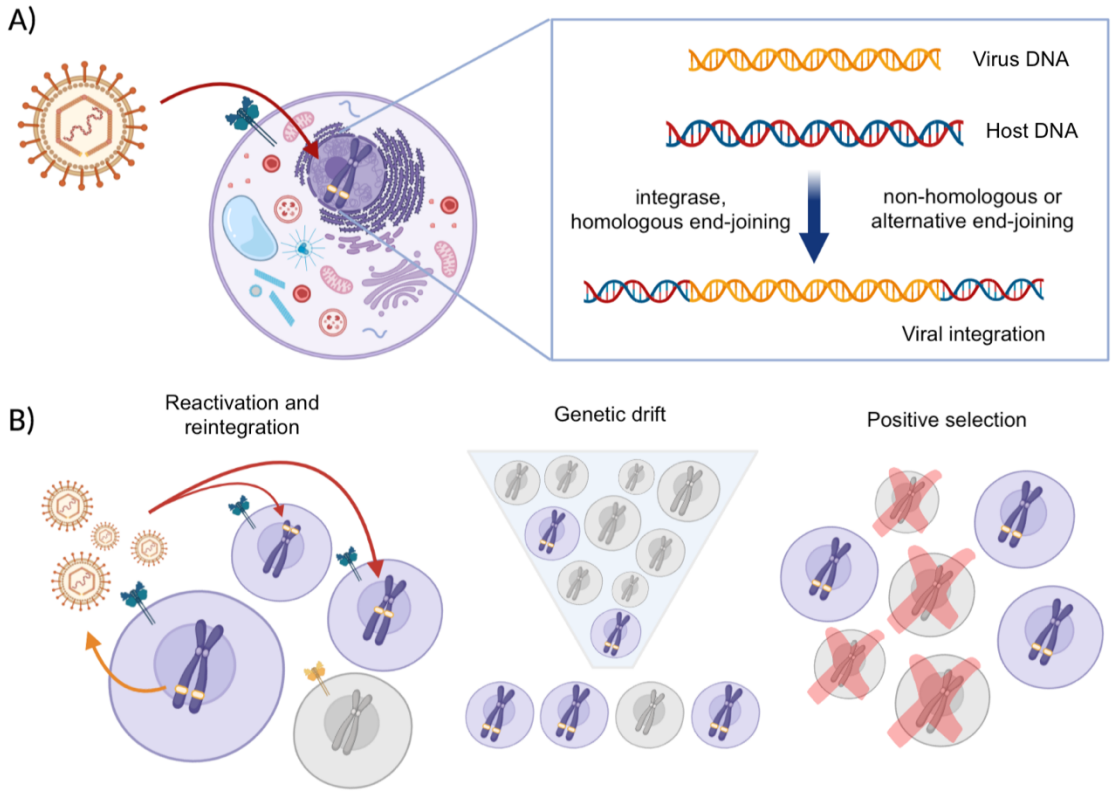


Figure 1. A) By infecting a susceptible cell, which expresses the required receptor for entry (blue), the virus genetic material can potentially gain access to the nucleus and be integrated into the cell's genome. This viral integration is considered an EVE once it has been inherited vertically. B). The persistence of an EVE throughout evolutionary time can result from different processes. If the virus has integrated an intact viral copy it might persist by continued cycles of reactivation and reintegration, leading to target-site polymorphisms. Cells might become non-susceptible by a mutation in the receptor protein (yellow),

restricting the virus' ability to colonise the genome. Alternatively, genetic drift may stochastically change the frequency of an EVE, here illustrated by a population-bottleneck. Finally, an EVE might confer a selective advantage on the host and will increase in frequency as a result of positive selection. Over evolutionary time, certain viruses such as bracoviruses have colonised all their hosts' chromosomes.

The idea of endogenous viral elements has its origins in the work with lysogenic bacteriophages by Lwoff and Gutmann [11], who showed that the virus had an endogenous stage which became a heritable feature of the host bacterium (termed a **prophage**). Wollman later suggested that the prophage is "intimately tied" to the genetic material of the bacterium, and is inherited as a host gene [12]. This general mechanism of integration into the host genome was later extended to some animal viruses: Lwoff proposed that polyomavirus could integrate into the chromosomes of cancerous cells [17], and Temin used this idea to construct the "provirus hypothesis" for *Rous sarcoma virus*, an oncogenic retrovirus of chicken [18]. For many years it was believed that retroviruses, were the only viruses to enter the eukaryotic germline given that integration into the host genome is a step required to complete viral replication (forming a **provirus**). However, the discovery of multiple geminivirus-like sequences in the genome of tobacco plants (*Nicotiana tabacum*), demonstrated that DNA viruses could also enter the germlines of eukaryotes [19]. Eventually, it was shown that viruses from all **Baltimore groups** could potentially integrate into eukaryotic genomes [16].

Compared to the RNA-retrotranscribing viruses, DNA virus EVEs occur less frequently [16]. Most families of eukaryotic DNA viruses do not encode an integrase, which limits direct integration into the host genome. Overall, the integration mechanisms for DNA virus EVEs are less well understood, although progress is being made (Box 2). Many families of DNA viruses have considerably larger genomes than RNA viruses, and integration of a full viral copy may be strongly counter-selected by the host. In general, it seems that the most abundant DNA virus EVEs in eukaryotic genomes are from families with smaller genome sizes (<10 kb), those which preferentially replicate in the nucleus and the families which encode an integrase [16,20–24]. The finding of very large herpesviral EVEs (>100 kb), is probably facilitated by the tendency of these viruses to establish latent infection in the host nucleus [25]. From an evolutionary standpoint, DNA viruses evolve more slowly and probably codiverge more often with their hosts [26–28], suggesting that their associations with hosts are more stable as compared to RNA viruses.

The diversity of DNA viruses found in eukaryotic genomes

DNA viruses in the genomes of eukaryotes fall into four viral kingdoms (*Bamfordvirae*, *Heunggongvirae*, *Shotokuvirae* and *Pararnavirae*) and two genera currently classified in the family "*Polydnaviridae*" (Figure 2). The kingdoms have been established by homology in protein structure and the molecular biology of the viruses. The kingdom *Bamfordvirae* comprises a diverse assemblage of viruses which share major and minor capsid proteins with vertical jelly-roll folds [29]. Viruses in the kingdom *Bamfordvirae* infect all major eukaryotic lineages and include the Nucleocytoplasmic Large DNA viruses (NCLDVs), virophages, *Mavericks*, *Polinton*-like viruses and adenoviruses. The kingdom *Heungonggvirae* includes the herpesviruses which infect animals as well as the tailed bacteriophages, which share the HK97 fold in their capsid proteins as well as similarities in the portal complex and the virus assembly pathways [13].

Another major lineage of eukaryotic viruses with endogenous elements is the kingdom *Shotokuvirae* (phyla *Cressdnaviricota* and *Cossaviricota*). The phylum *Cressdnaviricota*

comprises the families *Geminiviridae*, *Genomoviridae*, *Circoviridae* and *Nanoviridae*, which are known as the "circular, rep-encoding single-stranded DNA viruses" [7]. The family *Parvoviridae*, composed of viruses with linear single-stranded genomes and the families *Papillomaviridae* and *Polyomaviridae*, which have viruses with circular double-stranded DNA genomes, share replication proteins homologous to the ones in CRESS DNA viruses and thus have likely evolved from a common ancestor [30,31]. Hepadnaviruses and caulimoviruses, which infect vertebrates and plants respectively, are classified in the kingdom *Pararnavirae*. These viruses replicate by transcribing their genomes into RNA and then using a reverse transcriptase to make more dsDNA genomes [8]. Viruses in the family "*Polydnaviridae*" are found integrated in the genomes of parasitoid wasps; these viruses are used as gene transfer agents to deliver virulence factors into the host where the wasp larvae develop [32,33]. "*Polydnaviridae*" is an artificial grouping since the two genera *Bracovirus* and *Ichnovirus* have distinct evolutionary origins [33–35].

The integrations from these viruses have revealed two common themes: 1) some virus lineages have integrated more broadly than we would expect from the exogenous host range, indicating that these groups may infect a greater number of eukaryotic species than currently thought and 2) some integrations represent new viral families/lineages with no known exogenous counterparts, effectively increasing our knowledge of eukaryotic viral diversity. Circoviruses, for example, are known to infect vertebrates and arthropods, but their EVEs have also been identified in nematodes, gastropods and in *Giardia intestinalis* [16,36–38]. Geminiviruses and nanoviruses are viral pathogens of plants, their EVEs have been found in fungi, *Entamoeba hystolitica* [20], plants [39] and insects [40], while nanovirus-like sequences have been detected in plants, amphibians, crustaceans, insects, placozoans, green algae, diatoms, and in the *Giardia*, *Entamoeba* and *Blastocystis* protozoan parasites [20]. It has been noted that geminiviral and nanoviral-like EVEs in insect genomes probably derive from plant viruses that use these animals as vectors, in particular by the presence of movement protein genes [40]. Similarly, although hepadnaviruses are known to infect mammals and birds, hepadnavirus EVEs indicate that turtles and crocodilians are also potential hosts [41]. This is consistent with the discovery of exogenous hepadnaviruses (and 'nakednaviruses') infecting additional vertebrate hosts (amphibians and teleost fish) [42,43]. NCLDV EVEs have also been discovered which suggest a broad association with all major eukaryotic lineages and a greater phylogenetic diversity than presently acknowledged [28,44,45]. These insights gained from paleovirology can complement other sequence-based approaches such as metagenomics [46], as powerful tools for virus discovery.

Multiple novel viral lineages that have not been found in exogenous form, have been discovered by analysing the endogenous elements in host genomes. Naryaviruses, nenyaviruses and vilyaviruses are new clades of CRESS DNA viruses which have been identified in the genomes of the human protozoan parasites *Giardia duodenalis* and *Entoamoeba* spp. [47]. Their distinct placement on the *rep* phylogeny suggests they potentially represent 3 new viral families [47]. Another viral lineage distantly related to circoviruses and nanoviruses was identified in the genome of the pillbug *Armadillidium vulgare*, and a possible nonplant-infecting lineage of nanoviruses in the genomes of the molluscs *Crassostrea gigas* and *Lottia gigantea* [48]. In plants, a new lineage of endogenous caulimoviruses named 'florendoviruses' was discovered; they consist of bipartite genomes (unlike other caulimoviruses) and to exist at very high copy numbers (>0.5% total genome size) [49]. Importantly, new lineages of endogenous viruses (that is, functional viruses that persist integrated in the genomes of hosts), have also been discovered. *Mavericks* probably represent at least two new viral families [27]. They were previously described as large **transposons** (~20

kb), but they were later found to encode proteins homologous to capsid forming proteins of other bamfordviruses [50]. *Mavericks* found in the genomes of vertebrates undergo frequent cross-species transmissions and their core genes are evolving under strong purifying selection, further supporting their existence as endogenous viruses [27]. *Teratorn*s are another remarkable group of mobile genetic elements which encode proteins homologous to the proteins of alloherpesviruses [51]. *Teratorns* are extremely large elements (~180 kb) which have been found in the genomes of teleost fish [22,51,52]. They likely represent a new viral family which is the sister lineage to *Alloherpesviridae* [51]. Intriguingly, the ichnoviruses that have been domesticated by ichneumonid parasitoid wasps have an uncertain viral origin [34]. Although there is no sequence similarity to any other known lineage of viruses, their viral ancestry is supported by the absence of introns in their genes (more typical of virus rather than wasp genes) [34]. Therefore, it will be important to clarify the viral (or cellular) ancestry of ichnoviruses.

In general, oncogenic DNA virus integrations seem to be absent from the genomic fossil record of animals. Adenoviruses and polyomaviruses are known to integrate into somatic cells but have not been found in host germlines [53–55], there is only one report of a potentially integrated papillomavirus in the genome of the platypus (although an integration site has not been identified) [56], and herpesviral integrations have only been described in the tarsier genome or as low frequency (~1%) polymorphisms in the human population [25,57]. Despite the rich genomic fossil record of hepadnaviruses in vertebrates, no hepadnaviral integrations in mammals have been identified. Interestingly, the oncogenic X protein is specific to the *Orthohepadnavirus* genus that infects mammals, other vertebrate hepadnaviruses do not encode this protein and they are not known to be oncogenic to their hosts [41].

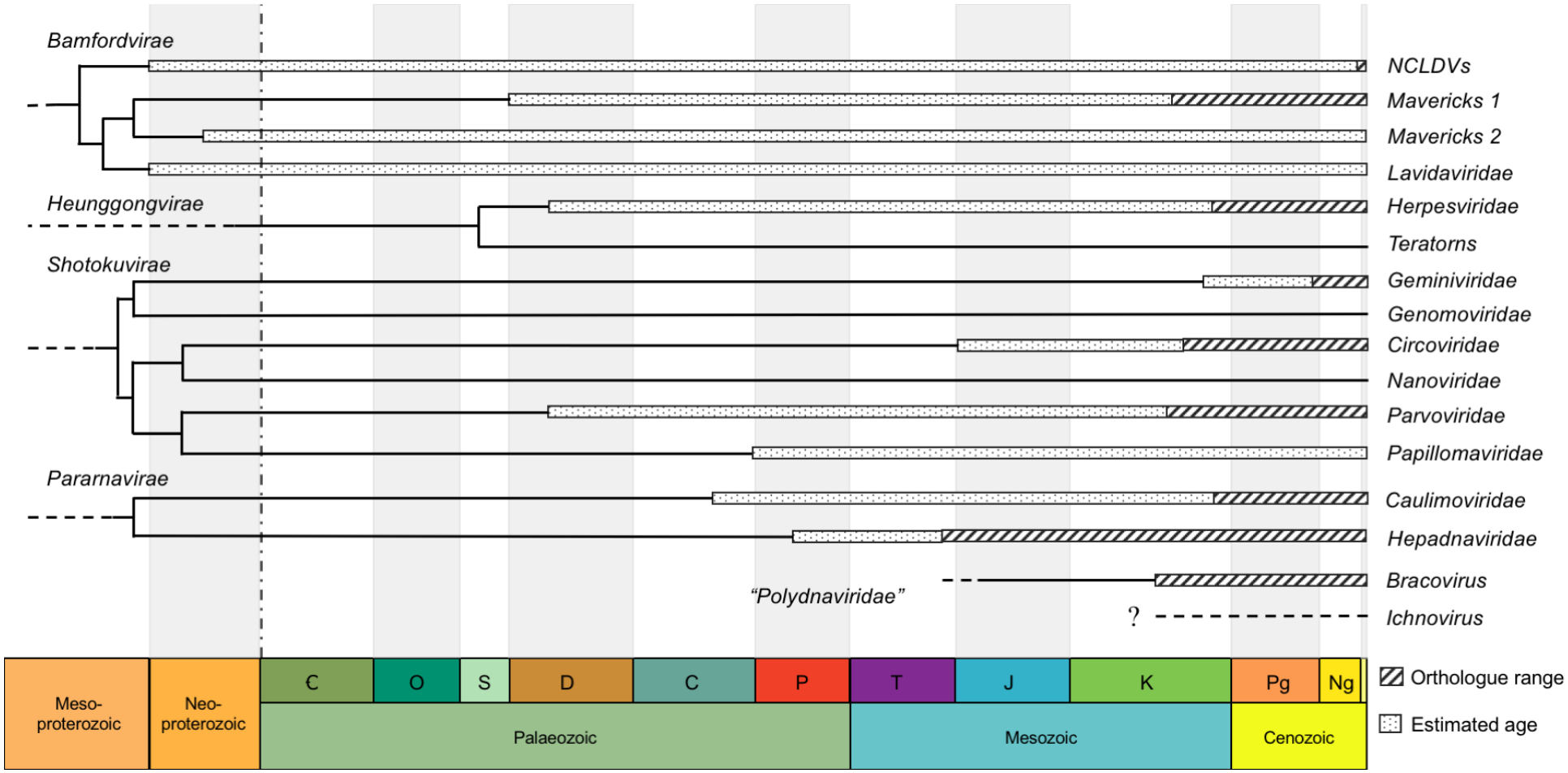


Figure 2. Evolutionary trees of the major lineages of DNA viruses found in eukaryotic genomes. Cross-hatched boxes represent the age of the oldest viral integrations identified by orthology in host species, while dotted boxes represent the age estimated for the groups based on molecular dating, genetic exchanges, codivergence or biogeographic data. The cladistic relationships shown for *Bamfordvirae* and *Shotokuvirae* are based on current evolutionary hypotheses [31,58]. Node ages reflect the divergence date for the hosts infected by the different viral lineages, but this codivergence assumption should be handled with caution.

Coevolution of DNA viruses with eukaryotes

The DNA virus EVEs of eukaryotes have also revealed the long timescales of virus-host coevolution. Orthologous EVEs provide direct evidence for the existence of viral lineages tens to hundreds of millions of years in the past (Figure 2). The oldest orthologue dated is an hepadnavirus EVE integrated in the genomes of turtles 207-230.7 Mya [41]. Another hepadnavirus EVE in the genomes of birds and the tuatara could be potentially 280 My old, but authors caution against this finding since orthology was difficult to assess [21]. Two orthologous *Maverick* integrations also occur in the genomes of turtles with an estimated age of 95 My [27]. In humans, a parvovirus capsid gene is found integrated in an intron of the *Ellis van Creveld syndrome 2* gene, the integration occurred at least 98 Mya since orthologous EVEs are found in non-human primates, carnivores, ungulates and dolphins [59]. Orthologous circovirus EVEs have also been dated to 72-90 My based on their presence in pythons and vipers [36]. In plants, an endogenous caulimovirus orthologue has been discovered in the genomes of *Pinus taeda* and *Pinus lambertiana*, indicating an age of at least 75 My [23].

Evidence for the ancient coexistence of eukaryotes and DNA viruses has also been obtained from analyses of sequence divergence or by close correspondence of the virus/host phylogenies. By using relaxed molecular clocks on a phylogeny calibrated by the dates of genomic orthologues, *Mavericks* have been estimated to have infected vertebrates since at least 419 Mya [27]. In a group of parvoviruses called “chapparvoviruses”, endogenised elements mostly follow the host phylogeny, suggesting they have infected their insect hosts for at least 400 My [60]. Similarly, the evolution of circovirus EVEs in ray-finned fish agrees with the host phylogeny, suggesting they potentially integrated >200 Mya [36]. From the divergence in the repeat regions (which are identical at integration) and using rates of neutral evolution reported for mammals, an endogenous herpesvirus found in the genome of the tarsier (*Tarsius syrichta*) has been dated to 56-76 My [57]. In humans, herpesviruses 6A and 6B (HHV6A/B) can infect cells of the germline and can be inherited vertically [61]. An HHV6B element, present in the genomes of Maasai, Pakistani and Native American people, was dated at ~84,000-324,000 years, corresponding to the time before the modern human migrations out of Africa [62].

Additional evidence has been obtained for the existence of NCLDV and **virophages** more than a billion years ago. Phylogenetic analyses indicate that the DNA-dependent RNA polymerase was exchanged on several occasions between NCLDVs and proto-eukaryotes, which occurred 1-2 billion years ago [63]. Comparison of different phylogenetic hypotheses using the sequences of the core viral proteins further suggest that virophages evolved to parasitise NCLDVs short after the origin of the NCLDV ancestor [58]. This is consistent with the close correspondence between the promoter and poly-A sequences of virophages and NCLDVs, which would have been inherited from a recent common ancestor and this specificity maintained throughout evolutionary time [64,65]. This lineage of primitive virophages would have further diversified into the modern virophages (lavidaviruses), *Mavericks*, *Polinton*-like viruses, cytoplasmic linear plasmids and adenoviruses [58].

Another relevant aspect in DNA virus evolution is the role of cross-species transmissions. Although DNA viruses tend to have stable host associations and to co-speciate with their hosts, cross-species transmissions can occur frequently in some lineages. For example, it was shown that although *Mavericks* in cyprinid fish had mainly evolved by co-speciation (65%), cross-species transmissions occurred 25% of the time [27]. Similarly, although the evolution of

hepadnaviruses is characterised by stable host associations and they mirror the host phylogeny, there is evidence suggesting multiple host-jumps between vertebrates and even a cross-class transmission between fish and mammals [43]. The macroevolutionary mode in caulimoviruses, however, seems to be dominated by cross-species transmissions rather than co-speciation; these include cross-division transfers in plants [23,66]. The reason for this difference could be due to the transmission mode of caulimoviruses which use insect vectors to spread to new plants; the finding of endogenous caulimoviruses in the genomes of these insects could support this hypothesis. Paleovirological evidence can thus be used to identify hidden vectors/intermediates during cross-species transmissions.

Transitions between viruses and transposons

The DNA viruses of eukaryotes have given rise to new lineages of endogenous viruses and replication competent transposons. Endogenous viruses encode the proteins required to make an infectious particle, while transposons do not. The *Teratorn* lineage of endogenous viruses originated when an exogenous virus related to alloherpesviruses acquired the integrase of a *piggyBac* transposon [51,52]. Similarly, the *Mavirus*-like virophages as well as *Mavericks*, share a retroviral-like integrase that was acquired from an ancestral retroelement [24,29]. Endogenous virophages are an example of DNA virus EVEs exapted to function in immunity (Box 3). The bracoviruses, which are dispersed across the chromosomes of braconid parasitoid wasps, originate from the domestication of endogenous nudiviruses that integrated into the wasp genome ~103 Mya [35,67]. In fact, the nudivirus ancestors of bracoviruses encode tyrosine recombinases and have integrated extensively across a broad range of arthropod genomes [68].

Multiple lineages of eukaryotic transposons also descend from ancestral DNA viruses. In the black truffle, *Tuber melanosporum*, geminivirus *rep*-like sequences have been found in association with 2 transposase ORFs and flanked by terminal inverted repeats (TIRs); these elements form their own clade and seem to have amplified within the truffle genome [20]. Another *rep*-like element with TIRs appears in the planarian, *Schmidtea mediterranea*; there are multiple copies which are similar to the extrachromosomal PEVE element [20]. Circovirus *rep*-like elements have undergone independent intragenomic amplifications in carnivores in association with LINE-1 retroelements [36]. In birds, certain endogenous hepadnaviruses have reached more than 300 copies in their host genomes and were hypothesised to have been mobilised by flanking TEs [21].

The transition from transposons to viruses is illustrated by hepadnaviruses and caulimoviruses. The reverse transcriptases of viruses form a monophyletic group in the phylogeny of retroelements [8,69,70], suggesting that they ultimately descend from a transposable element that acquired a capsid protein. However, while the capsid protein of caulimoviruses is homologous to the proteins of retroviruses, metaviruses and pseudoviruses (orterviruses), hepadnaviruses encode a non-homologous capsid protein [70]. Recently, a lineage of hepadnavirus-like retroelements (HEART1/2) was discovered in the genomes of arthropods [69]. HEART elements are believed to be retrotransposons since they lack the core protein required for capsid formation, appear in more than 300 copies in some insect genomes and have a conserved linear genetic organisation [69]. Discovery of the HEARTs raises two possible scenarios for the origins of retrotranscribing viruses: 1) either the orterviruses and hepadnaviruses arose independently from an ancestral retroelement by acquiring different capsids or 2) orterviruses and hepadnaviruses descend from a common viral ancestor that exchanged the ancestral capsid gene in one of the lineages (implying that HEARTs evolved

from an ancestral virus by reductive evolution). The question about the nature of the ortervirus/hepadnavirus ancestor (either virus or retrotransposon), could be illuminated by searching for other closely related lineages of endogenous viruses or retrotransposons in the genomes of eukaryotes.

Functional significance and potential uses of DNA virus EVEs

In addition to the roles of endogenous virophages, bracoviruses and ichnoviruses as defence and offence mechanisms in immunity, DNA virus EVEs have been an important source of evolutionary novelty in eukaryotes. These include novel DNA binding-proteins, the eukaryotic DNA-dependent RNA polymerases as well as a role in horizontal transfer of genetic material (Box 4). Importantly, the new viral diversity uncovered by paleovirology could be a source of useful biotechnological applications. For example, adeno-associated viruses (*Parvoviridae*) are well-established vectors used in genetic engineering; however, since they are small viruses, transgene sizes are limited to about 5 kb [71]. As an alternative, retroviral vectors can package up to 11 kb of genetic material [72]. Since they have the capacity to integrate and also infect vertebrates, *Mavericks* and *Teratorns* seem to be attractive candidates to develop new viral vectors that would deliver larger/multiple transgenes to the host chromosome. These endogenous viruses could also be potentially used to develop new viral-based vaccines or vectors for reprogramming T-cells against cancer [73,74].

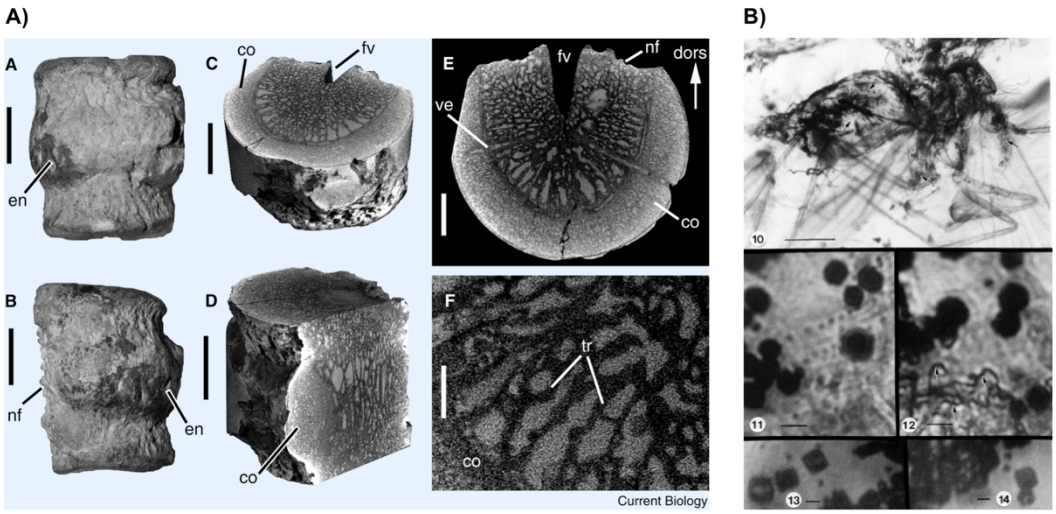
Concluding Remarks and Future Perspectives

Paleovirology is advancing our understanding of the diversity and deep origins of eukaryotic DNA viruses. We have recognised endogenous elements from numerous viral families and we are constantly uncovering a hidden diversity of viral lineages integrated across the genomes of eukaryotes. These endogenous viral elements are giving us insights into the interactions between viruses and hosts over geological timescales, which span a very detailed record in the millions to tens of millions of years and a few integrations in the hundreds of millions of years. The inferences from these findings are starting to give us insights into virus evolution in the scale of billions of years in the past. Research in paleovirology has also helped to disentangle the blurry evolutionary boundaries between eukaryotic DNA viruses and transposons. We are learning increasingly more about their impact on host biology, and especially about their role in immunity. Many questions are still open to enquiry (‘see Outstanding Questions’). New algorithms that bypass the need for detecting homology to known viral proteins could greatly increase our knowledge of viral diversity and evolution. In addition, it will be important to study the molecular virology of elements such as *Mavericks* and *Teratorns*, given that they could potentially be used as new vectors in genetic engineering, gene therapy or for the development of new vaccines.

Box 1. Evidence of ancient viruses from fossils and aDNA

The study of ancient viruses, i.e. viruses which have infected organisms of the past, can also be informed by the geological fossil record and molecular fossils, most notoriously, by the isolation and sequencing of aDNA (genetic material which has been preserved under exceptional circumstances). Two reports have been made of indirect evidence of viruses in the geological fossil record: in non-avian dinosaurs and insects. The vertebra of a Jurassic dinosaur (dated at ~150 My) was described in 2011 which showed characteristic signs of Paget’s disease [75]. Although it is not entirely clear what causes Paget’s disease, it seems

underlying genetic factors may interact with paramyxoviruses to produce an altered histology at the site of infection [76]. Structures resembling viral occlusion bodies (similar to the ones produced by baculoviruses) were reported in the gut of amber-fossilised insects from the Cretaceous Period (~100 My) [77]. Although this evidence is indirect, it hints to the possibility of finding actual fossilised **virions** or molecular fossils in the future. Direct sequencing of virus genomes is indeed possible from archaeological samples. A study of ancient DNA from human Bronze and Iron age samples (800-4,500 years), recovered sequences from hepatitis B viruses (family *Hepadnaviridae*) infecting these individuals [78]. Their analyses revealed an extended diversity of hepatitis B viruses which is not seen today, and concluded that present day diversity arose only after the split of the New and Old World genotypes [78]. In exceptional circumstances more ancient viruses have been recovered. A 30,000-year old giant virus, *Pithovirus sibericum*, was isolated from permafrost after culture in *Acanthamoeba castellanii* [79]. This new virus combined a *Pandoravirus* amphora-like capsid morphology with a gene content more similar to icosahedral viruses [80]. A later study from the same permafrost sediment samples characterised *Mollivirus sibericum*, another nonicosahedral giant virus which appeared to be a distant relative of pandoraviruses or from another lineage which acquired pandoravirus genes from extensive horizontal gene transfers [81]. These examples illustrate a field of paleovirology which remains largely unexplored.



Box 1 figure. A) Sections of a vertebra from *Dysalotosaurus lettowvorbecki* showing characteristic signs of Paget's disease: bone outgrowth, thickening of the bone cortex and formation of internal trabeculae. B) Specimen of an amber-fossilised sandfly (Diptera: Phlebotomidae), showing structures which resemble, in size and geometry, viral occlusion bodies present in the gut. Images reproduced with permission from Elsevier [75,77].

Box 2. Mechanisms of DNA virus integration in eukaryotes

DNA virus EVEs have formed by a number of mechanisms in eukaryotes. *Mavericks*, *Teratorns* and virophages, which encode their own integrase/transposase/recombinase, can mediate direct integration of the element into the host genome [22, 24, 65]. The DNA circles of bracoviruses, endogenous viruses of parasitoid wasps, also integrate into host immune cells using a tyrosine recombinase and follow a unique mechanism involving sequence cleavage and loss at the host integration motifs (HIMs) [82,83]. Human herpesviruses 6 A

and B (HHV6A/B) integrate full genome copies into telomeres, probably by using the homologous recombination pathway (either break-induced replication, requiring Rad51 or single-strand annealing, requiring MRN/RPA/Rad52), which depends on the sequence similarity of the telomeric repeats in the virus (pTMR) and the host [25]. Most other integrations are thought to occur via the non-homologous end-joining or microhomology-mediated pathways. Adeno-associated virus (AAV), a **virus satellite** of adenovirus in the family *Parvoviridae* (genus *Dependoparvovirus*), can integrate via site-specific nicking of the host DNA by the viral Rep, followed by host DNA synthesis, strand displacement and template switching to the AAV genome, which generates an endogenous copy of AAV [84]. In humans, AAV can latently infect cells by integrating into chromosome 19q13.42, causing a partial duplication of the *myosin-binding subunit 85* (MBS85) gene [84]. Proteins involved in the nonhomologous end-joining pathway, Ku70/80, PARP1 and Rad50, have been shown to participate in the site-specific integration mechanism of AAV [85]. In addition, endogenous viral elements from hepadnaviruses have been proposed to integrate via non-homologous end-joining induced after double-strand breaks [86]. Caulimoviruses have also been suggested to integrate via either the non-homologous end-joining or microhomology mediated pathways [23, 49]. It is plausible that additional DNA virus EVEs have arisen by integration through these repair mechanisms, which are mainly found as genome fragments, but experimental evidence for specific cases is still lacking. The synthesis-dependent end-joining pathway, which relies on regions of microhomology and the presence of inverted or direct repeats [87] at virus-host junctions, could also account for the integration of some DNA viruses. Another possibility is that DNA virus EVEs may integrate by interaction of RNA transcripts with the machinery of retroelements in cells, such as has been found to occur extensively for RNA virus EVEs [16]. The hallmarks for integration via LINEs for example, are a poly-A tail and flanking target-site duplications [16], however, signs for this integration pathway have not been characterised for DNA virus EVEs.

Box 3. Endogenous DNA viruses in immunity

A number of endogenous DNA virus integrations function in immunity. That is the case for provirophages, or the integrated form of virophages. The *Mavirus* virophage can integrate into the genome of its host *Cafeteria roenbergensis*, and upon coinfection with its giant virus *Cafeteria roenbergensis virus* (CroV), it reactivates and interferes with viral replication [88]. Provirophage reactivation leads to decreased CroV titres, and although it does not protect the eukaryotic host from dying, release of newly produced virophages protects the wider host population [89]. As such, integrated provirophages may be seen as an altruistic defence system against giant virus infection in unicellular eukaryotes [88]. Another remarkable case is that of polydnaviruses, a group of endogenous viral elements in the genomes of parasitoid wasps (Superfamily Ichneumonoidea) [33]. Polydnavirus EVEs have been domesticated by parasitoid wasps on at least three separate occasions, to be expressed in the ovaries of females and package virulence genes which are delivered to the caterpillar host [35]. These are essential in suppressing the host immune system for development of the wasp larvae. Elements from the *Bracovirus* genus, which infect 'microgastroid complex' wasps, seem to be monophyletic and diverged from betanudiviruses ~190 My ago (101-340 My)[33, 35]. In contrast, polydnaviruses in the genus *Ichnovirus*, which infect wasps in the Campopleginae and Banchinae subfamilies, do not appear to be monophyletic and suggest two independent acquisitions [90]. In addition, some DNA virus EVEs appear to contribute to immune defence by RNA interference mechanisms. In the *Ornithodoros* soft ticks, small-interfering RNAs (siRNA) and PIWI-interacting RNAs (piRNAs) are produced from endogenous

African Swine Fever virus-like integrations which may confer these vectors some protection against infection by the virus [91]. In a study of arthropod genomes [40], piRNA clusters were found to be enriched in EVEs and 18.49% of these integrations resembled parvoviruses. However, the small RNAs did not have a high similarity to the known viruses, which may indicate that they might interfere against uncharacterised exogenous viruses or that they had this function in the past [91]. Intriguingly, ancient HHV6A/B integrations have been found closely linked to the *MOV10L1* locus (in the subtelomere of chromosome 22q) in East Asians, which is required to make piRNAs [92]. Hypothetically, piRNAs from this locus could function in restricting further integrations by exogenous HHV6 [92].

Box 4. Virus-host horizontal gene transfers and evolutionary novelty

The exchanges of genetic material between hosts and viruses are the sources of significant evolutionary novelty. One striking case is that of dinoflagellates, whose chromatin is depleted in histones and which instead use DNVPs (dinoflagellate/viral nucleoproteins) as their primary DNA-binding protein [93]. The only homologue of DNVPs occurs in viruses of the family *Phycodnaviridae*, it appears these proteins were acquired by the ancestors of dinoflagellates through a phycodnavirus infection, however the precise evolutionary source of these proteins remains unclear [94]. Interestingly, when expressed in the yeast *Saccharomyces cerevisiae*, DNVPs have a toxic effect by impairing transcription and disrupting nucleosomal chromatin, this effect is relieved by a reduction in the number of histones [94]. Thus, it might be the case that histone depletion was an adaptation to chromatin invasion by DNVPs [63]. Another remarkable case is that of the largest subunits of the DNA-dependent RNA polymerase of eukaryotes and NCLDVs [63]. Phylogenetic analyses of the homologous proteins, revealed that these polymerases were exchanged between protoeukaryotes and NCLDVs, and the viral polymerases gave rise to two of the three DNA-dependent RNA polymerases (polymerases I and II) used by modern eukaryotes [63]. NCLDVs are also prominent for their extensive capture of host gene homologues, which have been acquired independently by different lineages and seem to be a driving force for viral gigantism [95]. DNA viruses have also been implicated in mediating horizontal gene transfer between eukaryotes. Nine different transposon superfamilies have been found integrated in the genome of the baculovirus *Autographa californica multiple nucleopolyhedrovirus* (AcMNPV) [96], and some present signs of horizontal transfer between multiple sympatric species of insects [97]. Hitchhiking has also been reported for a *Maverick* integrated into a bracovirus in the genome of *Cotesia congregata*, raising the possibility of horizontal transfer of an endogenous virus mediated by bracoviruses between the wasp and lepidopteran host genomes [98]. Furthermore, DNA viruses may be implicated as the main mechanism of eukaryote-to-eukaryote gene transfer, especially because there is extensive evidence for both host gene capture and cross-species transmission events occurring during viral evolution [99]. A compelling example is that of C-type lectins from hymenopterans which seem to have been transferred on at least two occasions to their lepidopteran hosts via infecting bracoviruses [100]. The existence of these integrations as polymorphisms in virus populations will require the use of ultra-deep sequencing to uncover the magnitude of host gene capture by viruses.

Acknowledgements

JGNB is supported by the "Dr. Jose Gregorio Hernandez" Award from the National Academy of Medicine of Venezuela and Pembroke College, Oxford.

Glossary

Virus: a DNA or RNA genetic element which encodes a capsid for horizontal transmission between cells; viruses are obligate parasites of cells.

Reverse transcription: the process by which an RNA molecule is transcribed into a complementary DNA copy.

Virion: infectious viral particle that can successfully parasitise a susceptible and permissive host cell.

Prophage: the endogenous form of a bacteriophage integrated into a bacterial chromosome.

Provirus: the endogenous form of a retrovirus integrated into a eukaryotic chromosome.

Provirophage: the endogenous form of a virophage integrated into the genome of a eukaryotic or giant virus host.

Endogenous viral elements (EVEs): a viral sequence integrated into the genome of a cell in the germline, which has been vertically inherited (from parent to offspring). EVEs may retain the ability to form infectious viral particles, they may be degraded (fossils) or have acquired a new function.

Baltimore groups: classification of viruses based on the nature of their genome and mRNA expression strategy. Viruses are classified into 7 Baltimore groups: dsDNA viruses (Group I), ssDNA viruses (Group II), dsRNA viruses (Group III), plus-sense ssRNA viruses (Group IV), minus-sense ssRNA viruses (Group V), reverse transcribing ssRNA viruses (Group VI) and reverse transcribing dsDNA viruses (Group VII).

Exaptation: biological trait that has been repurposed by natural selection to perform a new function.

Virophage: a virus satellite that has a negative (parasitic) effect on the replication of the helper virus. The virophages described so far are all parasites of giant viruses and belong to the family *Lavidaviridae*.

Giant virus: a virus in the phylum *Nucleocytoviricota* with a genome > 500 kb. All giant viruses discovered so far belong to this clade, but this does not exclude the possibility of other uncharacterized lineages of giant viruses.

Transposon (or transposable element): a sequence that can be copied into a new genomic location, either by encoding the enzymes required for transposition (autonomous) or by encoding the replication signals to use the enzymes from another element (non-autonomous). Unlike viruses, transposons do not encode capsids for horizontal transmission between cells.

Virus satellite: a virus that depends on coinfection of a cell by a helper virus in order to replicate.

References

1 Koonin, E. V. et al. (2015) Origins and evolution of viruses of eukaryotes: the ultimate modularity. Virology 479–480, 2-25.
2 Wigington, C.H. et al. (2016) Re-examination of the relationship between marine virus and microbial cell abundances. Nat. Microbiol. DOI: 10.1038/nmicrobiol.2015.24.
3 Koonin, E. V. (2016) Viruses and mobile elements as drivers of evolutionary transitions. Philos. Tran.s R Soc. Lond. B Biol. Sci. 371, 555–571.
4 Koonin, E. V. and Krupovic, M. (2018) The depths of virus exaptation. Curr. Opin. Virol. 31, 1-8.
5 Opriessnig, T. et al. (2020) Porcine circoviruses: current status, knowledge gaps and challenges. Virus Res. DOI: 10.1016/j.virusres.2020.198044.
6 Brandes, N. and Linial, M. (2019) Giant viruses-big surprises. Viruses. DOI: 10.3390/v11050404.
7 Krupovic, M. et al. (2020) Cressdnaviricota: a virus phylum unifying seven families of rep-encoding viruses with single-stranded, circular DNA genomes. J. Virol. DOI: 10.1128/JVI.00582-20.
8 Krupovic, M. et al. (2018) Ortervirales: new virus order unifying five families of reverse transcribing viruses. J. Virol. DOI: 10.1128/JVI.00515-18.
9 Senkevich, T.G. et al. (2015) Mapping vaccinia virus DNA replication origins at nucleotide level by deep sequencing. Proc. Natl. Acad. Sci. U. S. A. 112, 10908-10913.
10 Lynch, M. and Marinov, G.K. (2015) The bioenergetic costs of a gene. Proc. Natl. Acad. Sci. U. S. A. 112, 15690-15695.
11 Lwoff, A. and Gutmann, A. (1950) Recherches sur un Bacillus megatherium lysogène. Ann. Inst. Pasteur 78, 711-739.
12 Wollman, E.L. (1953) Sur le determinisme génetique de la lysogénie. Ann. Inst. Pasteur 84, 281-293.
13 Baker, M.L. et al. (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. J. Virol. 79, 14967-14970.
14 Krupovic, M. and Bamford, D.H. (2008) Virus evolution: how far does the double β-barrel viral lineage extend? Nat. Rev. Microbiol. 6, 941–948.
15 Charlesworth, B. (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10, 195–205.
16 Katzourakis, A. and Gifford, R.J. (2010) Endogenous viral elements in animal genomes. PLoS Genet. DOI: 10.1371/journal.pgen.1001191.
17 Lwoff, A. (1960) Tumor viruses and the cancer problem: a summation of the conference. Cancer Res. 20, 820-829.
18 Temin, H.M. (1964) The participation of DNA in Rous sarcoma virus production. Virology 23, 486–494.
19 Bejarano, E.R. et al. (1996) Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proc. Natl. Acad. Sci. U. S. A. 93, 759-764.
20 Liu, H. et al. (2011) Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. BMC Evol. Biol. DOI: 10.1186/1471-2148-11-276.
21 Lytras, S. et al. (2021) Ancient evolution of hepadnaviral paleoviruses and their impact on host genomes. Virus Evol. DOI: 10.1093/ve/veab012.
22 Inoue, Y. et al. (2017) Complete fusion of a transposon and herpesvirus created the Teratorn mobile element in medaka fish. Nat. Commun. DOI: 10.1038/s41467-017-00527-2.
23 Gong, Z. and Han, G.Z. (2018) Euphyllophyte paleoviruses illuminate hidden diversity and macroevolutionary mode of Caulimoviridae. J. Virol., DOI: 10.1128/JVI.02043-17.
24 Pritham, E.J. et al. (2007) Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene 390, 3-17.

25 Aimola, G. et al. (2020) Current understanding of human herpesvirus 6 (HHV-6) chromosomal integration. Antiviral Res. DOI: 10.1016/j.antiviral.2020.104720.
26 Geoghegan, J.L. et al. (2017) Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathog. DOI: 10.1371/journal.ppat.1006215.
27 Barreat, J.G.N. and Katzourakis, A. (2021) Phylogenomics of the Maverick virus-like mobile genetic elements of vertebrates. Mol. Biol. Evol. DOI: 10.1093/molbev/msaa291.
28 Gallot-Lavallée, L. and Blanc, G. (2017) A glimpse of Nucleo-cytoplasmic Large DNA virus biodiversity through the eukaryotic genomics window. Viruses. DOI: 10.3390/v9010017.
29 Krupovic, M. and Koonin, E. v. (2015) Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution. Nat. Rev. Microbiol. 13, 105–115.
30 Hickman, A.B. et al. (2002) Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. Mol. Cell 10, 327-337.
31 Kazlauskas, D. et al. (2019) Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. Nat. Commun. DOI: 10.1038/s41467-019-11433-0.
32 Drezen, J.M. et al. (2017) Endogenous viruses of parasitic wasps: variations on a common theme. Curr. Opin. Virol. 25, 41–48.
33 Gauthier, J. et al. (2018) The recurrent domestication of viruses: Major evolutionary transitions in parasitic wasps. Parasitology, 145, 713-723.
34 Volkoff, A.N. and Cusson, M. (2020) The unconventional viruses of ichneumonid parasitoid wasps. Viruses. DOI: 10.3390/v12101170.
35 Thézé, J. et al. (2011) Paleozoic origin of insect large dsDNA viruses. Proc. Natl. Acad. Sci. U. S. A. 108, 15931-15935.
36 Dennis, T.P.W. et al. (2019) The evolution, distribution and diversity of endogenous circoviral elements in vertebrate genomes. Virus Res. 262, 15–23.
37 Cui, J. et al. (2014) Low frequency of paleoviral infiltration across the avian phylogeny. Genome Biol. DOI: 10.1186/s13059-014-0539-3.
38 Gilbert, C. et al. (2014) Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. Proc. Biol. Sci. DOI: 10.1098/rspb.2014.1122.
39 Filloux, D. et al. (2015) The genomes of many yam species contain transcriptionally active endogenous geminiviral sequences that may be functionally expressed. Virus Evol. DOI: 10.1093/ve/vev002.
40 ter Horst, A.M. et al. (2018) Endogenous viral elements are widespread in arthropod genomes and commonly give rise to PIWI-interacting RNAs. J. Virol. DOI: 10.1128/JVI.02124-18.
41 Suh, A. et al. (2014) Early Mesozoic coexistence of amniotes and Hepadnaviridae. PLoS Genet. DOI: 10.1371/journal.pgen.1004559.
42 Lauber, C. et al. (2017) Deciphering the origin and evolution of Hepatitis B viruses by means of a family of non-enveloped fish viruses. Cell Host Microbe 22, 387-399.
43 Dill, J.A. et al. (2016) Distinct viral lineages from fish and amphibians reveal the complex evolutionary history of hepadnaviruses. J. Virol. 90, 7920-7933.
44 Maumus, F. and Blanc, G. (2016) Study of gene trafficking between acanthamoeba and giant viruses suggests an undiscovered family of amoeba-infecting viruses. Genome Biol. Evol. 8, 3351-3363.
45 Gong, Z. et al. (2020) Molecular fossils reveal ancient associations of dsDNA viruses with several phyla of fungi. Virus Evol. DOI: 10.1093/ve/veaa008.
46 Schulz, F. et al. (2020) Giant virus diversity and host interactions through global metagenomics. Nature 578, 432-436.
47 Kinsella, C.M. et al. (2020) Entamoeba and Giardia parasites implicated as hosts of CRESS viruses. Nat. Commun. DOI: 10.1038/s41467-020-18474-w.
48 Thézé, J. et al. (2014) Remarkable diversity of endogenous viruses in a crustacean genome. Genome Biol. Evol. 6, 2129-2140.
49 Geering, A.D.W. et al. (2014) Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. Nat. Commun. DOI: 10.1038/ncomms6269.
50 Krupovic, M. et al. (2014) Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. Biol. Direct. DOI: 10.1186/1745-6150-9-6.
51 Aswad, A. and Katzourakis, A. (2017) A novel viral lineage distantly related to herpesviruses discovered within fish genome sequence data. Virus Evol. DOI: 10.1093/ve/vex016.

52 Inoue, Y. et al. (2018) Fusion of piggyBac-like transposons and herpesviruses occurs frequently in teleosts. Zoological Lett. DOI: 10.1186/s40851-018-0089-8.
53 Doerfler, W. The fate of the DNA of adenovirus type 12 in baby hamster kidney cells. Proc. Natl. Acad. Sci. U. S. A. 60, 636-643.
54 Sutter, D. et al. (1978) Patterns of integration of viral DNA sequences in the genomes of adenovirus type 12-transformed hamster cells. Cell 14, 569-585.
55 Feng, H. et al. (2008) Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 319, 1096-1100.
56 Cui, J. and Holmes, E.C. (2012) Evidence for an endogenous papillomavirus-like element in the platypus genome. J. Gen. Virol. 93, 1362–1366.
57 Aswad, A. and Katzourakis, A. (2014) The first endogenous herpesvirus, identified in the tarsier genome, and novel sequences from primate rhadinoviruses and lymphocryptoviruses. PLoS Genet. DOI: 10.1371/journal.pgen.1004332.
58 Barreat, J.G.N. and Katzourakis, A. (2021) Evidence for a billion-years arms race between viruses, virophages and eukaryotes. bioRxiv. DOI: 10.1101/2021.04.20.440574.
59 Liu, H. et al. (2011) Widespread endogenization of densoviruses and parvoviruses in animal and human genomes. J. Virol. 85, 9863-9876.
60 Pénzes, J.J. et al. (2019) An ancient lineage of highly divergent parvoviruses infects both vertebrate and invertebrate hosts. Viruses. DOI: 10.3390/v11060525.
61 Wight, D.J. et al. (2020) Unbiased optical mapping of telomere-integrated endogenous human herpesvirus 6. Proc. Natl. Acad. Sci. U. S. A. DOI: 10.1073/pnas.2011872117.
62 Aswad, A. et al. (2020) Evolutionary history of endogenous human herpesvirus 6 reflects human migration out of africa. Mol. Biol. Evol. 38, 96-107.
63 Guglielmini, J. et al. (2019) Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. Proc. Natl. Acad. Sci. U. S. A. DOI: 10.1073/pnas.1912006116.
64 Campbell, S. et al. (2017) Disentangling the origins of virophages and polintons. Curr. Opin. Virol. 25, 59-65.
65 Fischer, M.G. and Suttle, C.A. (2011) A virophage at the origin of large DNA transposons. Science 332, 231-234.
66 Diop, S.I. et al. (2018) Tracheophyte genomes keep track of the deep evolution of the Caulimoviridae. Sci. Rep. DOI: 10.1038/s41598-017-16399-x.
67 Gauthier, J. et al. (2021) Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. Commun. Biol. DOI: 10.1038/s42003-020-01623-8.
68 Cheng, R.L. et al. (2020) Nudivirus remnants in the genomes of arthropods. Genome Biol. Evol. 12, 578-588.
69 Gong, Z. and Han, G.Z. (2018) Insect retroelements provide novel insights into the origin of hepatitis B viruses. Mol. Biol. Evol. 35, 2254-2259.
70 Krupovic, M. and Koonin, E. v. (2017) Homologous capsid proteins testify to the common ancestry of retroviruses, caulimoviruses, pseudoviruses, and metaviruses. J. Virol. DOI: 10.1128/JVI.00210-17.
71 Naso, M.F. et al. (2017) Adeno-associated virus (AAV) as a vector for gene therapy. BioDrugs 31, 317-334.
72 Sweeney, N.P. et al. (2017) Delivery of large transgene cassettes by foamy virus vector. Sci. Rep. DOI: 10.1038/s41598-017-08312-3.
73 Lauer, K.B. et al. (2017) Multivalent and multipathogen viral vector vaccines. Clin. Vaccine Immunol. DOI: 10.1128/CVI.00298-16.
74 Milone, M.C. and O'Doherty, U. (2018) Clinical use of lentiviral vectors. Leukemia, 32, 1529-1541.
75 Witzmann, F. et al. (2011) Paget disease of bone in a Jurassic dinosaur. Curr. Biol. 21, R647-648.
76 Kravets, I. (2018) Paget's disease of bone: diagnosis and treatment. Am. J. Med. 131, 1298-1303.
77 Poinar, G. and Poinar, R. (2005) Fossil evidence of insect pathogens. J. Invertebr. Pathol. 89, 243-250.
78 Mühlemann, B. et al. (2018) Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature 557, 418-423.
79 Legendre, M. et al. (2014) Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. Proc. Natl. Acad. Sci. U. S. A. 111, 4274–4279.
80 Legendre, M. et al. (2013) Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341, 281–286.

81 Legendre, M. et al. (2015) In-depth study of Mollivirus sibericum, a new 30,000-yold giant virus infecting Acanthamoeba. Proc. Natl. Acad. Sci. U. S. A. 112, E5327-E5335.
82 Chevignon, G. et al. (2018) Cotesia congregata bracovirus circles encoding PTP and ankyrin genes integrate into the DNA of parasitized Manduca sexta hemocytes. J. Virol. DOI: 10.1128/JVI.00438-18.
83 Burke, G.R. et al. (2014) Widespread genome reorganization of an obligate virus mutualist. PLoS Genet. DOI: 10.1371/journal.pgen.1004660.
84 Henckaerts, E. et al. (2009) Site-specific integration of adeno-associated virus involves partial duplication of the target locus. Proc. Natl. Acad. Sci. U. S. A. 106, 7571-7576.
85 Daya, S. et al. (2009) Adeno-associated virus site-specific integration is mediated by proteins of the nonhomologous end-joining pathway. J. Virol. 83, 11655-11664.
86 Gilbert, C. and Feschotte, C. (2010) Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. DOI: 10.1371/journal.pbio.1000495.
87 Yu, A.M. and McVey, M. (2010) Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res. 38, 5706-17.
88 Fischer, M.G. and Hackl, T. (2016) Host genome integration and giant virus-induced reactivation of the virophage Mavirus. Nature 540, 288–291.
89 Berjón-Otero, M. et al. The dual lifestyle of genome-integrating virophages in protists. Ann. N. Y. Acad. Sci. 1447, 97–109.
90 Béliveau, C. et al. (2015) Genomic and proteomic analyses indicate that Banchine and Campoplegine polydnaviruses have similar, if not identical, viral ancestors. J. Virol. DOI: 10.1128/JVI.01001-15.
91 Forth, J.H. et al. (2020) Identification of African swine fever virus-like elements in the soft tick genome provides insights into the virus' evolution. BMC Biol. DOI: 10.1186/s12915-020-00865-6.
92 Liu, X. et al. (2020) Endogenization and excision of human herpesvirus 6 in human genomes. PLoS Genet. DOI: 10.1371/journal.pgen.1008915.
93 Gornik, S.G. et al. (2012) Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. Curr. Biol. 22, 2303–2312.
94 Irwin, N.A.T. et al. (2018) Viral proteins as a potential driver of histone depletion in dinoflagellates. Nat. Commun. DOI: 10.1038/s41467-018-03993-4.
95 Koonin, E.V. and Yutin, N. (2018) Multiple evolutionary origins of giant viruses. F1000Research. DOI: 10.12688/f1000research.16248.1.
96 Loiseau, V. et al. (2020) Wide spectrum and high frequency of genomic structural variation, including transposable elements, in large double-stranded DNA viruses. Virus Evol. DOI: 10.1093/ve/vez0.
97 Gilbert, C. et al. (2014) Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. Nat. Commun. DOI: 10.1038/ncomms4348.
98 Dupuy, C. et al. (2011) Transfer of a chromosomal Maverick to endogenous bracovirus in a parasitoid wasp. Genetica 139, 489-496.
99 Gilbert, C. and Cordaux, R. (2017) Viruses as vectors of horizontal transfer of genetic material in eukaryotes. Curr. Opin. Virol. DOI: 10.1007/s10709-011-9569-x.
100 Gasmi, L. et al. (2015) Recurrent domestication by Lepidoptera of genes from their parasites mediated by bracoviruses. PLoS Genet. DOI: 10.1371/journal.pgen.1005470.