

Paleovirology, the genomic fossil record, and consequences of ancient viral infections

Aris Katzourakis, Department of Zoology, University of Oxford, Oxford, UK, OX1 3PS. E-mail: aris.katzourakis@zoo.ox.ac.uk

Fossils are invaluable sources of data for reconstructing the evolutionary history and interactions of organisms that may be long extinct. Viruses, however, are remarkably fragile entities that do not form a geological record. Thus, their ancient evolutionary history is hard to reconstruct, and their fast rates of evolution complicate approaches based on inference from extant viral sequence data. However, they do form a genomic fossil record as a result of occasional integration into the genomes of their hosts. This rich record of endogenous viral elements (EVEs), allows us to reconstruct ancient events in viral evolutionary history and in the interactions that viruses have had with their hosts. Furthermore, unlike the geological fossil record, these viral fossils contain genetic information as they preserve the sequence of ancient viruses, although sequence degradation occurs over time. Such a genomic record stretching back millions of years is not known to exist for any other organism.

The information provided by the analysis of EVEs has been instrumental in founding paleovirology, a field that focuses on the long term evolutionary history of viruses and their interactions with hosts. Endogenous retroviruses (ERVs) are the most common form of EVE in animal genomes, and by far the best studied. This is particularly the case for simple retroviruses, and these viruses have been studied in endogenised form for close to fifty years. Over a decade ago, the first endogenous lentivirus was discovered, a complex retrovirus encoding accessory genes, and discoveries since then have shown that various complex retroviruses have endogenised in several different animals over millions of years. Furthermore, EVEs have now been found for double and single stranded DNA and RNA viruses, for positive and negative sense RNA viruses, and for double stranded DNA viruses with an RNA intermediate, providing a fossil record of all known viral genome types and replication strategies.

Paleovirological analysis has revealed the evolutionary history of extant and extinct lineages of viruses. This type of research will continue to grow in pace as genomic data becomes ever more available, and the quality of our analytical tools improves. At the same time, in more recent years, paleovirology has increasingly encompassed unravelling the consequences of viral integrations for the evolution of their hosts. This increase in scope has involved more quantitative approaches and analytical tools that make use of technologies that require next generation sequencing and genomic analysis. The ten reviews collated in this issue of *Current Opinion in Virology* focus on a variety of topics, and synthesise current research in this growing field.

[Hayward](#) reviews work into the deep origins of retroviruses. New research supports a marine origin of the retroviruses, over half a billion years ago during the early Palaeozoic era. But what did the vertebrate retroviruses evolve from? Retroviruses share genomic features with LTR retrotransposons, a group of transposable elements that are confined to the cell, usually lacking the envelope gene that is necessary for contagion. Hayward highlights the importance of the Metaviridae, a group of viruses infecting organisms as

diverse as yeast, fruit flies and nematodes in determining the ultimate origins of the retroviruses. The metaviridae have acquired envelope genes on at least three separate occasions, and are thought to be the group most closely related to the retroviruses. Perhaps the retroviruses originated via envelope capture by an LTR retrotransposon-like lineage within the metaviridae during the early origins of the vertebrates. Hayward examines alternative hypotheses and analytical approaches that could reveal the ultimate origins of retroviruses and discusses the limitations of such studies, including issues associated with studying the promiscuous retroviral envelope gene.

The retroviral envelope gene evolves rapidly and shows discordance with the *pol* gene, indicating a history of recombination. [Sinha & Johnson](#), focus on one of the more extreme examples, namely the RD114-and-D-type-retrovirus (RDR) interference group. The RDR group shares a homologous *env* gene that uses a common cell-surface entry receptor, and the functional features and sequence characteristics of this group are well understood. The presence of the RDR *env* in otherwise distantly related viruses, or in viruses infecting distinct hosts, suggests that its acquisition may have facilitated cross-species transmission. Coupled with a rich genomic fossil record, the RDR group offers opportunities to combine evolutionary analyses with functional work to study how recombination, endogenisation, and cross-species transmission contribute to the evolution of viral lineages. Furthermore, viruses derived from the RDR group are involved in rabbit and human *syncytin* genes, a collection of disparate retroviral genes that have been exapted on multiple occasions independently by different mammalian hosts for roles in placentation. Thus, this group of viruses could reveal details of the exaptation of viral genes for host function.

The exaptation of viral genetic sequences for host function is a recurring theme in this special issue, and [Drezen and colleagues](#) review another intriguing example. Parasitoid wasps inject their eggs into their host as an obligate step in their replication cycle. Alongside these eggs, they also transfer the products of endogenous viruses derived from Polydnaviruses (PDVs), and these play important roles in their lifecycle. In the case of braconid wasps, the domesticated EVE is derived from a *Bracovirus*, while in ichneumonid wasps an *Ichnovirus* is involved. PDV particles are produced in the wasp ovaries, injected into the lepidopteran hosts during oviposition, and their gene products are subsequently expressed by cellular machinery where they function to counter the caterpillar immune response and protect the egg. The authors review this system, and outline the distinct evolutionary origins and distinct strategies to manipulate the host involved in this convergent viral capture. The recurrence of this type of virus capture suggests that it may be a general evolutionary mechanism, with more examples to be found in nature.

The theme of exaptation of viral gene sequences is also explored by [Horie](#) and by [Frank & Feschotte](#). [Horie](#) reviews the well-studied example of endogenous bornavirus-like elements (EBLs) of mammals. EBLs are the only described non-retroviral EVEs in human genomes, and have open reading frames that can be transcribed. Functional evidence of co-option in humans or other primates is lacking, but a recently integrated EBL in the thirteen-lined ground squirrel can express protein and inhibit mammalian bornavirus replication in a cell line (albeit not from the squirrel), suggesting likely co-option for an antiviral function. Recent research into mammalian EBLs suggests that some could have roles in cellular functions that are unrelated to antiviral activity. There are only a small number of known

exogenous bornaviruses, and analysis of EBL sequences could guide searches for such viruses, to better understand their interactions with antiviral EBLs.

[Frank & Feschotte](#) review how viral coding and regulatory sequences have been repurposed by their hosts. The authors discuss various examples of EVE-derived immunity, and other roles of exapted ERVs including in early embryonic development. Intriguingly, as with immunity, the role of ERVs in embryonic development also shows signs of convergence in humans and mice. The authors review work into how the placenta may provide fertile ground for the co-option of viral sequences as EVEs are frequently transcribed there. [Romer et al](#) focus their review on the HERV-H family that has been exapted for the regulation of human pluripotency. HERV-H provides binding sites for pluripotency transcription factors and gives rise to transcripts, some of which have been shown to have a functional role in the expression of pluripotency and the stem cell state. Despite the crucial role of HERV-H in human pluripotency, this function likely arose relatively recently in primate evolution. We do not yet know what proportion of HERV-H elements have functional roles in the genome, and whether some elements have other roles in the development of the early embryo.

A crucial requirement for the minimisation of the deleterious consequences of carrying potentially active viral sequences in the genome, which could also influence their exaptation, is the ability to control their expression. The expression of ERV nucleic acids and proteins can lead to an immune response, as the immune system recognises pathogen-associated molecular patterns (PAMPs). [Tie & Rowe](#) review how endogenous retroviruses and retrotransposons can be epigenetically controlled in differentiated cells, a process that begins in early embryogenesis and is maintained in postembryonic tissues. A variety of processes are involved in controlling ERV expression, including histone methylation and the KAP1 and KRAB-ZNF system that is also involved in gene regulation. Loss of ERV control could have autoimmune and tumorigenic consequences for the host, but targeted epigenetic modulation could potentially be a tool to stimulate immune surveillance in cancer patients.

The reviews in this issue highlight the importance of virus-to-host horizontal gene transfer (HGT), sometimes leading to functional consequences for the host. At the same time, host-to-virus HGT has long been recognised since the discovery of the *src* oncogene in retroviruses, and there are several other examples, but this process has received less attention than virus-to-host HGT. [Gilbert and Cordaux](#) review both of these processes, and also argue that viruses should be studied as potentially important intermediates of host-to-host HGT. The frequency of both of these processes in nature may be vastly underestimated due to the transient nature of such events. The authors highlight the importance of population level data and laboratory studies to recapitulate such events in controlled conditions. A growing number of studies indicate that the importance of virally mediated host-to-host HGT of transposable elements could be underestimated, and future work should quantify such events including the transfer of *bona fide* genes.

HGT is at the centre of many fundamental questions in virus evolution, including the origins of polintons, virophages and their relatives. [Koonin and Krupovic](#) emphasise that polintons, polinton-like viruses and parasites of the giant viruses termed transpovirons, are a group of related elements that can integrate into genomes; most of these elements also possess

features of *bona fide* viruses. At the centre of this scheme is a hypothesised ancestor of these viruses and transposable elements, termed polintoviruses. The authors also review evidence that virophages can confer resistance to giant viruses when integrated into their eukaryotic hosts. [Campbell et al](#) focus their review on polintons and virophages, and outline the evidence for the evolutionary scenario whereby polintons are derived from integrated virophages. They emphasise the importance of combining metagenomic discoveries with evolutionary techniques from paleovirology in order to disentangle the evolutionary history of polintons and virophages in a gene-by-gene approach. The questions addressed by these two reviews echo long standing unresolved debates about the ultimate origins of viruses and cellular life, as the events being investigated could date back to before the origins of the cell. Paleovirological discoveries and approaches offer unique opportunities to observe the ancient past of viruses, and could yet help resolve some of these longstanding puzzles.

The capture of viral sequences for the manipulation of host immunity appears to be a recurrent and widespread event in nature, in hosts as diverse as insects, mammals and single celled eukaryotes, and further research into this area is warranted to understand this phenomenon. The recurring exaptation of EVEs for roles in host immunity could be because viruses have already evolved to interface with host immunity, and the host exploits many pre-existing adaptations by co-opting them for immune related roles. This could in turn lead to secondary adaptive contributions from the co-opted EVE sequence, such as roles in placentation, pluripotency, or other cellular functions. Perhaps co-opted EVEs that do not serve an immune modulatory function in the present day were initially selected for such functions. Future research can exploit the insights generated from these early examples to systematically discover new ones, and determine the scale of this process.

The articles in this issue demonstrate the breadth of current approaches in paleovirology. Insights into the evolutionary biology of virus/host interactions offer a framework within which to incorporate observations about such interactions in the present day. This will help us understand and combat viruses that pose an existing health burden, and tackle future cross species transmissions. Furthermore, the integration of paleovirology with metagenomics and viral discovery will prove invaluable to understanding the diversity of both pathogenic and non-disease causing viruses. Some reviews highlight the potential for paleovirology to elucidate very early events in the evolution of viruses. Many articles highlight the transition towards understanding the consequences of ancient viral integration for the evolution of their hosts, and the central role of exaptation for the manipulation of host immunity. Future paleovirology research will continue to uncover novel roles of EVEs in host biology, as well as quantify their impact on the evolution of life.

Biography

Aris Katzourakis is Associate Professor of evolutionary genomics in the department of zoology, at the university of Oxford, and a Royal Society university research fellow. He received his BSc and PhD degrees at Imperial College London. His research interests include viral evolution, genome evolution, and transposable elements. He leads a research group that is centred on paleovirology, a field that focuses on the evolutionary history of viruses and on their interactions with their hosts.