Paleovirology: Blessing or Curse of Ancient Viruses: A Review

Shubhangi Warke, Prabhakar Tembhurne, Sumedha Bobade, V.C. Ingle

ABSTRACT
Ancient virus genomes preserved as fossils and carried by host within their genome. Although viral genomes evolve rapidly, their rate of change slows to the same pace as that of the host's DNA after insertion, making it possible to study viral DNA sequences that are many millions of years old. Paleovirology is the study of viral fossil records typically over prehistoric or geological timescales and the effects that these agents have had on the evolution of their hosts. Viruses sometimes heritably integrate into the genomes of their hosts, resulting in genomic features known as endogenous viral elements (EVEs). Using EVEs, the field of paleovirology investigates the long term evolution of viruses and their impact on hosts. One of the fruitful outcomes of high throughput genomics is the widespread availability of whole genome data, offering the unprecedented opportunity to investigate EVEs at a large scale. This review, highlights the utility of antiviral gene evolution for the study of paleovirology, the consequent surge in paleovirology and the main methodological techniques used to study them. EVEs can only be understood within an evolutionary framework and we outline a generalized workflow for conducting paleovirology studies.

Key words: Bovine immunodeficiency virus, Endogenous viral elements, Human endogenous retrovirus, Mouse leukemia viruses, Testicular germ cell tumours.

Paleovirology is the study of viral fossil records. Within the past century, a number of “emerging viruses” with pathogenic properties, such as HIV-1, SARS-CoV and several novel reassortments of influenza A, have entered the human population on a large scale. By identifying, sequencing and analysing the genetic fingerprints, it is possible to reconstruct viral ancestors and learn more about them. This new field of study is known as paleovirology.

Paleovirology is the study of ancient extinct viruses and the effects that these agents have had on the evolution of their hosts (Emerman and Malik, 2010). The emerging field of paleovirology aims to study the evolutionary age and impact of ancient viruses (paleoviruses) on host biology (Patel et al., 2011). Many viruses can enter the genomes of their hosts; such as retroviruses, do so as an obligate step during their replication process and others can occasionally do so, either by accident or as a latent part of their life cycle. Dramatic episodes of viral infections have challenged and shaped animal evolution for hundreds of millions of years. Ancient virus genomes preserved as fossils and carried by host within their genome. When viral integrations occur in the germ line of their host, they can be passed on to the next generation, potentially existing in the host population. When this occurs, the integrated endogenous virus genomes evolve at host rates of mutation and their sequence is relatively stably preserved (Katzourakis, 2013).

Paleovirus is an informal term for regions of genomes that originate from ancient germ line integration of viral genetic material. The scientific term for such regions is endogenous viral element or EVE. EVEs that originate from the integration of retroviruses are known as endogenous retroviruses, or ERVs and most viral fossils are ERVs. They may be traced to millions of years back, hence the terminology, although strictly speaking, it is impossible to detect an ancient virus in fossils. The presence of retroviral sequences in host genomes was noticed in the late 1960s, but these observations were initially greeted with skepticism. This special issue begins with the history of how endogenous retroviruses (ERVs) first came to light and outlines the discoveries that led to the current understanding of viral sequences in host genomes (Weiss, 2013). Recent work has shown that all known viral genomic structures and replication strategies are represented in this fossil record, greatly expanding the scope of paleovirology beyond just the retroviruses (Katzourakis and Gifford, 2010). NIRV (Non-retroviral Integrated RNA Viruses) impacted host genomes without leaving any direct trace of their existence Borna virus (Emerman and Malik, 2010).

Direct and indirect paleovirology
Approaches that rely on the identification of viral sequences in host genomes have been termed ‘direct paleovirology’. The presence of ancient paleoviruses by looking at differences in the activity of antiviral genes with a known target across host species and reconstituting the history of diversifying selection that is characteristic of conflict with a pathogen. This approach has been termed ‘indirect paleovirology’. Indirect paleovirology is most powerful where the interaction between a viral protein and host protein are known. It can be limited in its ability to formally prove the existence of a paleovirus and to rule out alternative scenarios; for example, in cases where the viruses that have...
shaped the evolution of a particular gene are unknown or extinct. (Aswad and Katzourakis, 2012; Katzourakis, 2013).

**Direct paleovirolgy**

It is based on the host integrated paleoviral infections and recovery of the viral ‘fossil record’. Approaches that rely on the identification of viral sequences in host genomes have been termed ‘direct paleovirolgy’. However, the existence of ancient viruses can also be inferred by investigation of their effects on the host genes that have evolved to control them (Daugherty and Malik, 2012; Duggal and Emerman, 2012).

**Indirect paleovirolgy**

Virus-driven evolutionary adaptations within host proteins inferences about the viruses that drove this evolution (Patel et al., 2011). The activity of antiviral genes with a known target across host species and reconstructing the history of diversifying selection that is characteristic of conflict with a pathogen. The sequences of host ‘antiviral’ genes contain the signatures of epic evolutionary conflicts with viral antagonists. This approach has been termed ‘indirect paleovirolgy’ and is a powerful technique to trace both the ancient history of this conflict and how it has shaped susceptibility to modern viruses. Furthermore, this indirect approach can be the only way to study ancient infections in cases where EVEs cannot be identified. (Daugherty and Malik, 2012; Duggal and Emerman, 2012).

**Retroviral ‘Fossils’ in primate genome**

The human genome is a living document of ancient and now extinct viruses. Indeed, DNA of retroviral origin makes up 8% of human genome sequence. It is difficult to calculate exactly how many retroviral infections of the germ line led to the, 100,000 copies of endogenous retroviruses in the human genome because duplications, transpositions and other non-infectious events also contribute to this number. However, at a very minimum, each of the more than 31 families of endogenous retrovirus found in the human genome must have arisen from one or more separate paleoviruses that infected the ancestors of modern humans (Katzourakis et al., 2005). Since reinfections of the germ line with members of the same families occurred frequently (Belshaw et al., 2004), retroviral infections that impacted the genome had to have happened repeatedly during primate evolution with the most recent episode in humans between 100,000 and 1 million years ago (Bannert and Kurth, 2006).

“Paleovirolgy” is the study of ancient extinct viruses (called “paleoviruses”) and the effects that these agents have had on the evolution of their hosts. Thus far, the study of these viruses has mostly been limited to endogenous retroviruses that can be directly identified from their remnants in host genomes (Fig 1) (Emerman and Malik, 2010). The recent finding that at least one bornavirus gene has integrated in several mammalian genomes at multiple evolutionary periods demonstrates both the possibility of identifying and dating some other ancient classes of viral infections (Horie et al., 2010).

**Table 1:** EVEs identified in eukaryotic genomes (Feschotte and Gilbert, 2012).

| Family or genus | Taxa                                      | Number per haploid genome |
|-----------------|-------------------------------------------|---------------------------|
| Bornaviridae    | Vertebrates                                | 1 to 17                   |
| Herpesviridae   | Humans                                    | 1                         |
| Nudivirus       | Parasitic wasps                            | Several                   |
| Phycodnaviridae | Brown algae                                | 1                         |
| Circoviridae    | Mammals                                    | 1 to 2                    |
| Filoviridae     | Mammals                                    | 1 to 13                   |
| Reovirus        | *Aedes* spp. mosquitoes                    | 1                         |
| Bunyaviridae    | Ticks                                     | 14                        |
| Filoviridae     | Mammals                                    | 1 to 13                   |
| Hepadnavirus    | Passerine birds                            | 15                        |

During evolution, successive bursts of amplification within host germ lines have led to the formation of ERV repeat families that occupy a significant fraction of mammalian genomes (e.g. 8% and 10% in humans and mice, respectively). However, in some mammalian species, fully functional proviruses persist. Such ‘young’ ERVs can be reactivated to produce infectious virions that are transmitted horizontally and can even reinfect the germline to form proviruses that can become fixed as new ERVs at novel positions in the host genome. This is notably the case for the mouse leukemia viruses (MLVs) and the mouse mammary tumor viruses (Stocking and Kozak, 2008). Sheep genome harbour about 20 copies of endogenous betaretroviruses (DeMartini et al., 2003).

**EVEs shed light on the structural evolution of viruses**

As representatives of viral families fossilized at various evolutionary time points, endogenous viral elements (EVEs)
Paleovirology: Blessing or Curse of Ancient Viruses: A Review

Paleovirology can provide insights into the origins of components found in modern viral genomes. For example, the origin of the accessory genes of complex retroviruses (such as nef or rev in HIV) has long been a mystery. These genes have diverse functions and, unlike the structural and enzymatic genes, gag, pol, and env, which are common to all retroviruses, they are only found in a subset of retroviral genera. Until recently, only the vpx gene that is present in HIV 2 and in various simian immunodeficiency viruses (SIVs) had been traced; it appears to have arisen by non-homologous recombination between two different SIVs. The discovery of an endogenous lentivirus, PSIV, in the genome of prosimian primates (namely, Malagasy lemurs) helped to shed light on the origin of a second lentiviral accessory gene. This gene, called orf2, shows substantial sequence similarity to the 32 end of the primate lentiviruses reverse transcriptase domain, suggesting that it might have arisen by partial duplication of this domain, possibly via template jumping during reverse transcription, interestingly, orf2 is of the same size and is located at the same position (within the pol-env intervening region) as the tat gene of modern lentviruses, suggesting that it may be a tat orthologue. The orf2 sequence, however, does not share any substantial similarity with known tat genes, which themselves are extremely diverged from each other and are barely alignable. Whether orf2 and tat are truly homologous therefore remains an open question that would need to be addressed by functional studies. Unlike vpx, which is only present in HIV 2 and in a subset of SIVs and must be of very recent origin, tat is present in all primate lentiviruses and in other lentiviruses (such as bovine immunodeficiency virus (BIV) from cows, equine infectious anaemia virus (EIAV) from horses and rabbit endogenous lentivirus type K (RELIK), which suggests it emerged much earlier than vpx. Thus, the link between tat and the 32 end domain of the lentiviral reverse transcriptase in PSIV might be our best opportunity to trace the origin of tat. LTR, long terminal repeat (Katzourakis et al., 2005).

Paleovirology, in this broader sense, may be able to correlate the existence of ancient infections with known phylogeographical events, such as extinctions, bursts of speciation and exchanges of fauna between continents, island isolations and population migrations. Paleovirology could be viewed as the study of ancient viruses that primate genomes encountered and defeated during the course of evolution. This view emphasizes that our current antiviral repertoire was not optimized to combat present infections, but rather is the product of selection for survival of our species' past infections. Thus, the selective changes that these antiviral genes incurred during these periods of evolutionary pressure might make them less competent to fight modern viral challenges (Duggal and Emerman et al., 2012). For example, the human TRIM5 gene does not inhibit HIV, although it was certainly selected to inhibit something else in our past. The analysis of amino acids on antiviral genes driven by selection of ancient pathogens can be used to identify the interface between the host protein and the virus in ways that could conceivably be used to design rational antiviral drugs or gene therapy strategies. Such analyses of the virus-host battles, on an evolutionary scale, can also explain the otherwise mysterious loss of antiviral activities. For instance, antiviral genes that serve no other cellular functions can incur significant fitness costs or relaxed selection and therefore can be lost due to the lack of a pathogen during extended periods of time (OhAinle et al., 2008; Venkataraman et al., 2009).

Modern consequences of ancient viruses

Although paleovirology focuses on ancient viruses, it has an important role to play in directing our approaches to modern viruses and the threats they pose. In the era of 'big data' and genomics, it has become feasible to begin characterizing the genetic diversity of viruses on a large scale and using this data to establish measures of 'viral risk' that are founded in empirically derived ecological and evolutionary principles.

Paleovirology is an important component of this endeavour, in the same way that paleontology is an essential component of efforts to recover and comprehend the evolutionary history of cellular organisms. Developing an understanding of the history of viral diseases can guide the development of new strategies for disease control, particularly where traditional approaches have proved ineffective or inappropriate. It can also underpin the rationale of efforts to protect endangered species from viral disease, both in the wild and in captive breeding programs.

EVEs and Cancer

HERVs may be involved in carcinogenesis by virtue of the expression of HERV mRNA. (Andersson et al., 1998) functional proteins (Sauter et al., 1995) or retroviral-like particles (Lower et al., 1993). They may also be associated with the generation of new promoters (Schulte et al., 1996) or the activation of proto-oncogenes. The expression of HERV-R mRNA is increased in some cases of small cell lung carcinoma (Andersson et al., 1998). In addition, a teratocarcinoma cell line has been shown to possess a HERV-K sequence and to secrete retroviral-like particles (Lower et al., 1993). Testicular germ cell tumours (TGCTs) have been shown to contain proteins of the HERV-K family and patients with TGCT often exhibit a specific immune response to gag and env proteins (Sauter et al., 1995, Sauter et al., 1996). It has been suggested that HERV-K may be important in the progression of TGCT through inhibition of an effective immune response. Overexpression is a common mechanism by which protooncogenes become activated, leading to subsequent neoplastic transformation and the HERV env genes have been shown to encode immunosuppressive proteins (Nelson et al., 2003).

EVEs and autoimmunity

In 1990, an article appeared in the Times newspaper (24 November) with the title “AIDS-like virus may cause arthritis”. The report focused on Robert Garry’s research that identified retroviral particles in lip biopsies taken from patients with primary Sjogren’s syndrome (SS) (Garry et al., 1990). Similarly, in other autoimmune rheumatic diseases, such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), a plethora of articles added to this intriguing observation by providing evidence of retroviral
antigens at the site of disease or the presence of antiretroviral antibodies in the sera of patients. Mechanisms whereby HERVs could influence autoimmunity include molecular mimicry (HERVs sharing amino acids common to host proteins), superantigen motifs that bypass the normal MHC restrictive process of T cell stimulation, aberrant expression of antigens and the presence of neo-antigens, perhaps as a result of HERV and/or exogenous viral combinations. Evidence has been presented that HERV-K encodes an endogenous superantigen and that the development of insulin-dependent diabetes mellitus (IDDM) is associated with such a factor stimulating Vb7-carrying T cells, including those with an autoreactive potential.

Biological importance of EVEs

HERVs (Human Endogenous Retrovirus) have been present in our genome for a considerable period of time and perhaps have been retained because they performed a useful biological function. Alternatively, some HERVs may have been difficult to eliminate and thus persisted during evolution. To confer a selective advantage, a premise remains that HERVs produce products and/or augment mechanisms that benefit host survival.

Immunosuppressive peptide

The product derived from the env gene of mammalian type C retroviruses possesses motifs—for example, the fusion peptide, leucine zipper protein and immunosuppressive peptide (ISP)—that are essential for fusion and the infection of cells. In brief, the precursor env product is cleaved into two components: a surface protein (gp70) and a transmembrane protein (p15E) that contains an immunosuppressive region (Larsson et al., 1998). The presence of an ISP could be advantageous to a virally infected cell—in terms of shielding or “cloaking” itself from immunological attack—but may equally be important to a host. This is perhaps exemplified by HERV-R (ERV3), which is highly expressed in trophoblastic cells and results in high concentrations of env protein (~ 65 kDa) in syncytiotrophoblasts (Boyd et al., 1993). The immunosuppressive potential of this HERV and the fusogenic nature of placenta tissue suggests a possible involvement in normal placental function, in protecting the developing fetus from maternal immune response. Furthermore, it is possible that HERVs may change the pattern of gene expression during embryo development by altering different rates of development of different parts of the embryo (Mi et al., 2000).

HERVs constitute only a part of what are termed “transposable elements”, a generic term encompassing both DNA sequences that can be excised and reinserted at another site and retroelements. The term retroelements describes any sequence that can replicate itself by a process involving reverse transcription and includes HERVs, retrotransposons (which mostly lack an env gene), retroposons and retrosequences. Retroposons and retrosequences are exemplified by long stretches of related sequences (up to 6 kbp) called LINES and very short interspersed repeat elements of about 300 bp, respectively. Thus far, from being a fixed, immutable structure, the genome of a eukaryotic cell can harbour many sequences that move from one site on a chromosome to a completely different position. This phenomenon of plasticity is considered important because it permits rapid changes in our genome that could not be afforded by mutations alone. Furthermore, retroelements may carry regulatory sequences to new sites in the genome and thus alter the expression of existing adjacent genes (Speek, 2001).

Paleoviral ‘gifts’ from ancient viruses

A ‘domesticated’ viral gene is one that has been co-opted by host genomes for their own use. Such domesticated genes, found among the viral fossils in the genome, have maintained their open reading frame intact for millions of years despite mutational decay that one might expect for a gene under neutral selection (Patel et al., 2011).

Perhaps the best known of these domesticated viral genes is the syncytin gene, which is the surviving remnant of a HERV-W (for human endogenous retrovirus from W family) insertion into primate genomes more than 35 million years ago (Mi et al., 2000). While the bulk of this provirus shows mutational decay, the envelope gene has been preserved intact in all hominoids. It turns out that its molecular roles in membrane apposition and fusion are conserved but now for a completely different biological process, that of trophoblast development in the placenta (Mallet et al., 2004).

CONCLUSION

Viral fossils embedded in animal genomes provide unique insight into the evolutionary age of viral lineages. ERVs possess a similar genomic organisation (gag–pol– env) to present day exogenous retroviruses but are not infectious. ERVs may be of benefit to the host but could also be harmful and may be involved in certain autoimmune diseases and cancers indirect paleovirology investigates adaptive evolution in host genes to infer the action of an ancient virus. In future perspective Understanding the history of viral diseases can guide the development of new strategies for disease control. ERV and exogenous RV interference can help in developing novel strategy for anti-retroviral targets.

REFERENCES

Andersson, A., Svensson, A., Rolny, C., et al. (1998). Expression of human endogenous retrovirus ERV3 (HERV-R) mRNA in normal and neoplastic tissues. Int J Oncol. 12:309–13.
Aswad, A. and Katzourakis, A. (2012). Paleovirology and virally derived immunity. Trends in Ecology and Evolution. DOI 10.1016/j.tree.2012.07.007.
Bannett, N. and Kurth, R. (2006). The evolutionary dynamics of human endogenous retroviral families. Annu Rev Genomics Hum Genet. 7: 149–173.
Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G, Paces, J., et al. (2004). Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci USA. 101: 4894–4899.
Boyd, M.T., Bax, C.M.R., Bax, B.E., et al. (1993). The human endogenous retrovirus ERV-3 is upregulated in differentiating placental trophoblast cells. Virology. 196: 905–9.
Daugherty, M.D. and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annu. Rev. Genet. 46: 677–700.
DeMartini, J. C., Carlson, J. O., Leroux, C., Spencer, T., Palmarini, M. (2003). Endogenous retroviruses related to jaagsiekte sheep retrovirus. Curr Top Microbiol Immunol. 275: 117–137.

Duggal, N.K. and Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. Nat. Rev. Immunol. 12:687–695.

Emerman, M. and Malik, H.S. (2010). Paleovirology-Modern Consequences of Ancient Viruses. PLoS Biol. 8(2): e1000301. doi:10.1371/journal.pbiol.1000301.

Feschotte, C. and Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. 13: U283–U288. (doi:10.1038/nrg3199)

Garry, R.F., Fermin, C.D., Darenn, J.H., et al. (1990). Detection of a human intracisternal particle antigenically related to HIV. Science. 250: 1127–9.

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., et al. (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature. 463: 84–87.

Katzourakis, A., Rambaut, A. and Pybus, O.G. (2005). The evolutionary dynamics of endogenous retroviruses. Trends Microbiol. 13: 463–468.

Katzourakis, A. (2013). Paleovirology: inferring viral evolution from host genome sequence data. Philos Trans R Soc Lond B Biol Sci. 368(1626): 20120493.

Katzourakis, A. and Gifford, R.J. (2010). Endogenous viral elements in animal genomes. PLoS Genet. 6, e1001191. (doi:10.1371/journal.pgen.1001191).

Larsson, E. andersson, G. (1998). Beneficial role of human endogenous retroviruses: facts and hypotheses. Scand J Immunol. 48: 329–38.

Lower, R., Lower, J., Tondera-Koch, C., et al., (1993). A general method for identification of transcribed retrovirus sequences (R-US PCR) reveals the expression of the human endogenous retrovirus loci HERV-H and HERV-K in teratocarcinoma cells. Virology. 192: 501–11.

Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L., Mandr and B. (2004). The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proc Natl Acad Sci USA. 101:1731–1736.

Mi, S., Lee, X., Li X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., et al., (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature. 403: 785–789.

Nelson, P.N., Carnegie, P.R., Martin, J., Davari Ejtehadi, H., Hooley, P., Roden, D., Rowland-Jones, S., Warren, P., Astley, J., Murray P.G. (2003). Dernystified. Human endogenous retroviruses. Mol Pathol. 56: 11–18.

OhAinle, M., Kerns, J.A., Li, M.M., Malik, H.S., Emerman, M. (2008). Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell Host Microbe. 4: 249–259.

Patel, M.R., Emerman, M. and Malik, H.S. (2011). Paleovirology – Ghosts and gifts of viruses past. Curr Opin Virol. 1(4): 304–309. doi:10.1016/j.coviro.2011.06.007.

Sauter, M., Roemer, K., Best, B., et al., (1996). Specificity of antibodies directed against Env protein of human endogenous retroviruses in patients with germ cell tumours. Cancer Res. 56:4362–5.

Sauter, M., Schommer, S., and Kremmer, E., et al., (1995). Human endogenous retrovirus K10: expression of gag protein and detection of antibodies in patients with seminomas. J Virol. 69: 414–21.

Schulte, A.M., Shoupeng, L. and Kurtz, A., et al., (1996). Human tropoblast and choricarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. Proc Natl Acad Sci USA. 93: 14759–64.

Speck, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Biol Cell. 21:1973–85.

Stockey, C. and Kozak, C. A. (2008). Murine endogenous retroviruses, Cell Mol Life Sci. 65(21): 3383–3398. Doi:10.1007/s00018-008-8497-0.

Venkataraman, N., Cole, A.L., Ruchala, P., Waring, A.J., Lehrer, R.I., et al., (2009). Reactivated retrocytosis: ancestral human defenses active against HIV-1. PLoS Biol. 7: e95.

Weiss, R.A. (2013). On the concept and elucidation of endogenous retroviruses. Phil. Trans. R. Soc. B. 368: 20120494. (doi:10.1098/rstb.2012.0494).