Published in final edited form as:
Curr Opin Virol. 2011 October 1; 1(4): 304–309. doi:10.1016/j.coviro.2011.06.007.

Paleovirology – Ghosts and gifts of viruses past

Maulik R. Patel[1], **Michael Emerman**[1,2], and **Harmit S. Malik**[1,3,*]

Maulik R. Patel: mpatel@fhcrc.org; Michael Emerman: memerman@fhcrc.org; Harmit S. Malik: hsmalik@fhcrc.org

[1]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle WA USA

[2]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle WA USA

[3]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle WA USA

Abstract

The emerging field of paleovirology aims to study the evolutionary age and impact of ancient viruses (paleoviruses) on host biology. Despite a historical emphasis on retroviruses, paleoviral 'fossils' have recently been uncovered from a broad swathe of viruses. These viral imprints have upended long-held notions of the age and mutation rate of viruses. While 'direct' paleovirology relies on the insertion of viral genes in animal genomes, examination of adaptive changes in host genes that occurred in response to paleoviral infections provides a complementary strategy for making 'indirect' paleovirological inferences. Finally, viruses have also impacted host biology by providing genes hosts have domesticated for their own purpose.

Introduction

Dramatic episodes of viral infections have challenged and shaped animal evolution for hundreds of millions of years. Paleovirology is the study of the impacts of ancient, now extinct, viruses on modern host genomes. A fraction of paleoviruses have left unmistakable traces- fossilized versions of themselves- now inherited as part of host genomes. Many other paleoviruses impacted host genomes without leaving any direct trace of their existence. Here, we review recent advances in the study of paleoviruses that rely both on direct and indirect traces of these viruses. The surprising lessons learned from them inform us both about virus evolution as well as how ancient host genomes successfully negotiated infections from severely pathogenic viruses.

Retroviral fossils

Retroviruses are unique among animal viruses because as an obligate step of their lifecycle, they integrate into host genomes as proviruses. If this integration event happens in germline cells, these proviruses can get passed from parent to offspring and vertically inherited within the species. Eight percent of the human genome is derived from over 100,000 retroviral fossils [1], many of which are present in all simian primates that last shared a common ancestor over 30 million years ago [2]. On the other end of the spectrum is the ongoing

Address correspondence to: Michael Emerman, 1100 Fairview Avenue N, C2-023, Seattle WA 98109, (206) 667-5058 (tel), (206) 667-6523 (fax), Harmit S. Malik, 1100 Fairview Avenue N, A2-205, Seattle WA 98109, (206) 667-5204 (tel), (206) 667-6522 (fax).
*(handling corresponding author for editorial inquiries)

invasion of koala genome by the KoRV retrovirus, which appears to have started in Northeast Australian populations 100 years ago, and is rapidly progressing towards Southern populations [3].

Because proviruses become part of the host genome upon integration, many phylogenetic tools normally used to study evolution of host genes can be employed to infer the history of ancient viruses using proviral imprints. For example, the recent discovery of RELIK, an endogenized lentivirus in rabbit genomes, expanded the host range of lentiviruses to include lagomorphs (rabbits and hares) and pushed back the minimum age of lentiviruses to more than 12 million years old [4–6].

Orthology of provirus integration sites in multiple host species is indicative of a single integration event in an ancestral host genome, and is one of the most reliable ways to date the minimum age of an endogenized virus (Figure 1A). However, inferences of viral age are more ambiguous if integration sites are not orthologous (Figure 1B). This ambiguity is well highlighted in the recent discovery of lentivirus fossils in Microcebus and Cheirogaleidae species of Malagasy lemurs[7,8]. Given the near identical sequence of these fossils, these might have represented the same endogenization events that occurred before the split of the two lemur genera 18.6 – 28 million years ago. However, these integration sites were not orthologous. To distinguish between a single infection followed by lineage sorting versus independent infection events (Figure 1B), the authors examined the sequence of the endogenized fossils, which suggested a relatively young age of endogenization. This led to the conclusion that the fossils in Microcebus and Cheirogaleidae were the result two independent infection events[8]. Thus, in absence of orthology, other means are required to accurately date ancient infections.

Beyond retroviruses

Whole genome sequencing of multiple animal species has now revealed DNA remnants of multiple non-retroviral fossils[9]. The first of these discoveries to be published was that of endogenous viral elements (or EVEs) from Bornaviruses, which are negative strand RNA viruses associated with neurological pathology[10,11]. Using orthologous EVEs, Bornavirus fossils were determined to be more than 40 million years old in primates, and even older (>90 million years old) based on EVEs from Afrotherian species (elephants, hyrax, tenrec) [9].

Orthologous EVEs from Filoviruses, a family which consists of the deadly Ebola and Marburg viruses, suggest a minimum age for these viruses of 30 million years in rodents[9,11,12]. Filoviral imprints have been found in a broad swathe of other mammalian genomes[9,11,12], greatly expanding our knowledge of their ancient host range. Animal genomes also contain EVEs from RNA viruses of wide representation: *Bunyaviridae*, *Rhabdoviridae*, *Orthomyxoviridae*, *Reoviridae* and *Flaviviridae*[9] as well as from DNA viruses *Parvoviridae*, *Circoviridae* and *Hepadnaviruses*[9,13,14]. Remarkably, some paleoviruses (*Flaviridae*, *Rhabdoviridae*, *Parvoviridae*) can be almost entirely reconstructed from their EVEs [9]. This will be useful in experiments that can resurrect these ancient paleoviruses and how host genes defeated them.

None of these non-retroviruses possess the machinery to integrate their genes into host genomes but a variety of mechanisms may account for their EVEs. For example, Bornaviral EVEs may have occurred as a result of their N gene transcript hijacking the retrotransposition machinery of Line1 elements in host germlines [10,15]. Hepadnaviral EVEs may have resulted from aberrant non-homologous end-joining [13]. Despite their idiosyncratic differences, given the diversity of EVEs already found [9], it is likely that paleoviruses represent nearly the entire breadth of virology. Ironically, double stranded

DNA viruses such as Herpesviruses, which are suspected of having a long co-evolutionary history with animals [16], have not (yet) left paleoviral fossils in any animal genomes. However, given that paleovirology is in its infancy, it is likely a matter of time before a systematic search turns up fossils from these viruses.

Discordance in evolutionary rates

All proviruses and EVEs were discovered largely based on the fact that they are similar in sequence to extant viruses. This is surprising because one expects the currently circulating viruses used as search queries to have changed so much in their sequence that they should have quickly lost resemblance to their distant ancestors. One explanation for this observation is that the high virus evolution rates based on extant viral sequences represent short-term (years to decades) evolutionary rates that are not a true reflection of long-term (millions of years) virus evolution. Presumably, if long-term substitution rates could be measured, they would turn out to be much slower.

The recent discovery of hepadnaviral fossils in bird genomes provided an opportunity to directly measure long-term substitution rates of these viruses [13]. Hepadnaviruses are small DNA viruses (~3 kb), whose substitution rates, as calculated based on extant viral sequences, are on the same order as many rapidly evolving RNA viruses. Hepadnavirus fossils in bird genomes are up to 75% identical in sequence to contemporary hepadnaviruses, yet can be dated via orthology to be at least 19 million years old. The long-term substitution rate that can account for the accumulation of only 25% difference over 19 million years would have to be a thousand fold lower than estimates based on currently circulating hepadnaviruses [13]. Why would short-term versus long-term substitution rates of viruses be so different? One explanation is that the mutational space of a given virus is limited so that it is unable to change beyond a certain degree of similarity and still retain functional identity (Figure 2). If this explanation were true, short-term virus evolution would not accurately capture the constraint and long-term purifying selection under which viruses evolve. Thus, paleovirology is dramatically altering our perceptions of not only how old viruses are but also their long-term evolutionary trajectories.

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.

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[17]. 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[18]. Remarkably, retroviral envelope genes have been domesticated multiple times in eutherian mammals. Primates have also domesticated *syncytin-2*[19] (from HERV-FRD), rodents have domesticated *syncytin-A* and *syncytin-B*[20,21] (from a HERV-F like virus), while lagomorphs have also domesticated the envelope gene from the Env-Ory1 virus as *syncytin-Ory1*[22]. Thus, placental mammals may owe an important aspect of their reproductive function to retroviral gene domestication.

The *Friend-virus-susceptibility-1*, or *Fv1* gene derived from retroviral *gag* gene provides another example of domestication[23]. In this case, host genomes have domesticated the

capsid gene of a MuERV-L retrovirus for function as an antiviral factor 7 million years ago [24]. Fv1 may represent a prototypic example of host genomes domesticating viral components to interfere with either virus-virus or virus-host interactions necessary for successful virus infection.

An intriguing instance of a non-retroviral domestication might be that of bornavirus derived EBLN-1 (for endogenous Bornavirus-like nucleoprotein 1), which inserted into host genome before the divergence of strepsirrhine primates 54 million years ago[9]. Given the diminishingly small probability that an open reading frame the size of EBLN-1 would survive intact for this long, and its expression in cell lines, it is possible that EBLN-1 contributes to host biology. However, further functional studies will be required to determine the precise molecular function of EBLN-1.

What's next for direct paleovirology?

The great promise of paleovirology is a glimpse not only into the past but also into the unknown. All paleoviruses discovered so far have relied on some present-day counterpart. New methods for database searches are needed to identify related sequences found in many whole genome sequencing projects of animals that don't 'belong'. The search for ancient paleoviruses that have no extant versions can then begin in earnest.

In coming years, the emerging science of paleovirology is likely to advance beyond mere description of ancient infections. Specific hypotheses about the nature of ancient viruses or the impact they had on host genes can be tested with more ease since individual viral genes or entire viral genomes can be reconstructed *in silico* and then tested *in vitro*[25–28]. These experiments promise to reveal not only which viruses our ancestors encountered during their evolution, but potentially also how these ancient infections were successfully negotiated. Deciphering these ‘rules of combat’ could hold great promise to understand host-virus interactions in general.

Indirect paleovirology- evolutionary ‘echoes’ of ancient conflicts

Notwithstanding the recently discovered cases of non-retroviral EVEs, the majority of viruses are unlikely to have left behind a fossil record. Furthermore, while viral fossils tell us that infections took place, they do not inform us whether these viruses were pathogenic. A complementary approach to the study of viral fossils is to study virus-driven evolutionary adaptations within host proteins to make inferences about the viruses that drove this evolution[29].

Host genomes need to constantly adapt to counteract either new viruses or new 'adaptive' variants of viruses previously ‘defeated’ by the host. This will result in recurrent and hence rapid evolution of host proteins that come in direct contact with viral proteins. This evolutionary signature of positive selection can be used to make paleovirological inferences [29]. For example, TRIM5 is an antiviral factor that has evolved under positive selection. Given that TRIM5 is known to specifically restrict only retroviruses, one can infer that adaptive changes in TRIM5 were likely driven by ancient retroviruses[30,31]. By dating when these adaptive changes confined to the capsid-interaction domain occurred[32,33], it is also possible to infer the minimum age of retroviruses that the selection was in response to.

In contrast to TRIM5, antiviral proteins like Protein Kinase R antagonize a wide array of viruses [34]. At first glance, genes like PKR might appear to be poor paleoviral markers. However, distinct classes of viruses might have imposed selective pressures on distinct domains or residues of PKR. Thus, reconstructing when a particular domain of PKR evolved

under positive selection can provide a great deal of specificity about the type of paleovirus that may have caused this selection.

Indirect paleovirology has thus far been limited to analyses of individual antiviral genes. As more and more genomes are analyzed in detail, correlated episodes of positive selection in multiple host genes will add greater specificity to the hypothesis of a paleovirus, and provide insight into ancient host-virus interactions, which may have present-day repercussions. Ultimately, it is the synergy of indirect and direct paleovirology that is likely to yield the most robust insights into how these ancient viruses have affected the evolution of modern host genomes and their current susceptibility to present-day viruses.

Highlights

- Viral fossils embedded in animal genomes provide unique insight into the evolutionary age and mutation rate of viral lineages
- Paleoviruses may span the currently known breadth of animal virology
- Some viral genes have been domesticated by host genomes for housekeeping or defense functions
- Indirect paleovirology investigates adaptive evolution in host genes to infer the action of an ancient virus

Acknowledgments

We thank Efrem Lim, Matt Daugherty and Ray Malfavon-Borja for comments on this article. We are funded by a postdoctoral fellowship from the Helen Hay Whitney Foundation (MRP), NIH grant R37 AI30937 (ME) and an NSF CAREER grant (HSM). HSM is an Early Career Scientist of the Howard Hughes Medical Institute.

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]
2. Medstrand P, Mager DL. Human-specific integrations of the HERV-K endogenous retrovirus family. J Virol. 1998; 72:9782–9787. [PubMed: 9811713]
3. Tarlinton RE, Meers J, Young PR. Retroviral invasion of the koala genome. Nature. 2006; 442:79–81. [PubMed: 16823453]
4. van der Loo W, Abrantes J, Esteves PJ. Sharing of endogenous lentiviral gene fragments among leporid lineages separated for more than 12 million years. J Virol. 2009; 83:2386–2388. [PubMed: 19109386]
5. Keckesova Z, Ylinen LM, Towers GJ, Gifford RJ, Katzourakis A. Identification of a RELIK orthologue in the European hare (Lepus europaeus) reveals a minimum age of 12 million years for the lagomorph lentiviruses. Virology. 2009; 384:7–11. [PubMed: 19070882]
6. Katzourakis A, Tristem M, Pybus OG, Gifford RJ. Discovery and analysis of the first endogenous lentivirus. Proc Natl Acad Sci U S A. 2007; 104:6261–6265. [PubMed: 17384150]
7. Gifford RJ, Katzourakis A, Tristem M, Pybus OG, Winters M, Shafer RW. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. Proc Natl Acad Sci U S A. 2008; 105:20362–20367. [PubMed: 19075221]
8. Gilbert C, Maxfield DG, Goodman SM, Feschotte C. Parallel germline infiltration of a lentivirus in two Malagasy lemurs. PLoS Genet. 2009; 5 e1000425.
9. Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. PLoS Genet. 2010; 6 e1001191. ** A comprehensive survey from published animal genomes of endogenized non-

retroviral fragments, which bolsters the view that paleoviruses reflect the breadth of animal virology.

10. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, Ikuta K, Jern P, Gojobori T, Coffin JM, et al. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature. 2010; 463:84–87. [PubMed: 20054395] ** The first demonstration of an endogenous non-retrovirus, implicating ancient Bornaviral infections in animals (also in ref. 9, 11).
11. Belyi VA, Levine AJ, Skalka AM. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. PLoS Pathog. 2010; 6 e1001030. ** Multiple, rampant filoviral EVEs in animal genomes dramatically increase age and host range estimates for these medically important viruses.
12. Taylor DJ, Leach RW, Bruenn J. Filoviruses are ancient and integrated into mammalian genomes. BMC Evol Biol. 2010; 10:193. [PubMed: 20569424] ** Multiple, rampant filoviral EVEs in animal genomes dramatically increase age and host range estimates for these medically important viruses.
13. Gilbert C, Feschotte C. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. 2010; 8 ** In addition to identifying hepadnavirus fossils in bird genomes and significantly increasing the estimated age of these viruses, the authors provide the first estimate of long-term evolutionary rate for any virus using viral fossils.
14. Kapoor A, Simmonds P, Lipkin WI. Discovery and characterization of mammalian endogenous parvoviruses. J Virol. 2010; 84:12628–12635. [PubMed: 20943964]
15. Feschotte C. Virology: Bornavirus enters the genome. Nature. 2010; 463:39–40. [PubMed: 20054386]
16. Davison AJ. Evolution of the herpesviruses. Vet Microbiol. 2002; 86:69–88. [PubMed: 11888691]
17. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature. 2000; 403:785–789. [PubMed: 10693809]
18. Mallet F, Bouton O, Prudhomme S, Cheynet V, Oriol G, Bonnaud B, Lucotte G, Duret L, Mandrand B. The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proc Natl Acad Sci U S A. 2004; 101:1731–1736. [PubMed: 14757826]
19. Blaise S, de Parseval N, Benit L, Heidmann T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proc Natl Acad Sci U S A. 2003; 100:13013–13018. [PubMed: 14557543]
20. Dupressoir A, Marceau G, Vernochet C, Benit L, Kanellopoulos C, Sapin V, Heidmann T. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proc Natl Acad Sci U S A. 2005; 102:725–730. [PubMed: 15644441]
21. Dupressoir A, Vernochet C, Bawa O, Harper F, Pierron G, Opolon P, Heidmann T. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. Proc Natl Acad Sci U S A. 2009; 106:12127–12132. [PubMed: 19564597] * The first genetic evidence for the necessity of endogenized retroviral genes for placental function in mice, a likely exemplar for all the other syncytin genes in eutherian mammals.
22. Heidmann O, Vernochet C, Dupressoir A, Heidmann T. Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new "syncytin" in a third order of mammals. Retrovirology. 2009; 6:107. [PubMed: 19943933]
23. Best S, Le Tissier P, Towers G, Stoye JP. Positional cloning of the mouse retrovirus restriction gene Fv1. Nature. 1996; 382:826–829. [PubMed: 8752279]
24. Yan Y, Buckler-White A, Wollenberg K, Kozak CA. Origin, antiviral function and evidence for positive selection of the gammaretrovirus restriction gene Fv1 in the genus Mus. Proc Natl Acad Sci U S A. 2009; 106:3259–3263. [PubMed: 19221034]
25. Goldstone DC, Yap MW, Robertson LE, Haire LF, Taylor WR, Katzourakis A, Stoye JP, Taylor IA. Structural and functional analysis of prehistoric lentiviruses uncovers an ancient molecular interface. Cell Host Microbe. 2010; 8:248–259. [PubMed: 20833376]

26. Lee YN, Bieniasz PD. Reconstitution of an infectious human endogenous retrovirus. PLoS Pathog. 2007; 3:e10. [PubMed: 17257061] * These studies utilize viral fossils to resurrect ancient, now extinct paleoviruses to characterize their infection properties and pathogenic potential.
27. Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, Pierron G, Heidmann T. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Res. 2006; 16:1548–1556. [PubMed: 17077319] * These studies utilize viral fossils to resurrect ancient, now extinct paleoviruses to characterize their infection properties and pathogenic potential.
28. Kaiser SM, Malik HS, Emerman M. Restriction of an extinct retrovirus by the human TRIM5alpha antiviral protein. Science. 2007; 316:1756–1758. [PubMed: 17588933]
29. Emerman M, Malik HS. Paleovirology--modern consequences of ancient viruses. PLoS Biol. 2010; 8 e1000301.
30. Sawyer SL, Wu LI, Emerman M, Malik HS. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci U S A. 2005; 102:2832–2837. [PubMed: 15689398]
31. Goldschmidt V, Ciuffi A, Ortiz M, Brawand D, Munoz M, Kaessmann H, Telenti A. Antiretroviral activity of ancestral TRIM5alpha. J Virol. 2008; 82:2089–2096. [PubMed: 18077724]
32. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature. 2004; 427:848–853. [PubMed: 14985764]
33. Sebastian S, Luban J. TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. Retrovirology. 2005; 2:40. [PubMed: 15967037]
34. Elde NC, Child SJ, Geballe AP, Malik HS. Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature. 2009; 457:485–489. [PubMed: 19043403]

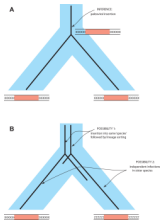

Figure 1. Deducing age of paleoviruses based on insertion in host genomes

(A) When a viral imprint is found in exactly the same genomic location in two species' genomes, then the paleoviral insertion is inferred to have occurred in a common ancestor. We can infer the minimum age of the paleovirus based on speciation times of the two most distantly related host species that bear the 'orthologous' provirus. (B) If the insertions are seen in 'non-orthologous' positions, there is a possibility that this reflects unequal lineage sorting in which Species A inherited one EVE while Species B inherited the other, despite the fact that both EVEs resulted from the same episode of infection in an ancestral species. More likely though, this reflects independent insertions after the species had already separated. Other corroborative evidence (e.g. estimates of mutations) is needed to formally distinguish between these two possibilities.

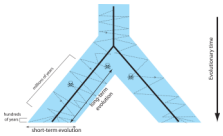

Figure 2. Apparent discordance in short versus long-term evolutionary rates of viruses

As viruses evolve over long periods of time, bulk lineages (bold lines) diverge and diversify. However, within each lineage, there are limits to which a viral lineage can evolve before it begins to lose function. As a result, given enough time and mutational saturation, long-term evolution of viruses may be bounded by functional constraints (dotted line represents evolutionary trajectory of a single viral lineage). As a result, any paleoviral fossils (denoted by 'skull and bones') derived from this 'trunk' of evolution will reflect the evolutionary boundedness of mutational space. In contrast, short-term mutation rates are less affected by mutational saturation and may not reflect the 'purifying selection' imposed long-term on viral lineages[13]. Thus, even though they occur at vastly different time-scales, extant viruses evolving under short-term mutation rates can be compared to, and used to identify, ancient paleoviral fossils.