A proviral puzzle with a prosimian twist

Welkin E. Johnson[1]

Department of Microbiology and Molecular Genetics, New England Primate Research Center, Harvard Medical School, Southborough, MA 01772

Viruses are obligatory intracellular parasites, and as such they cannot exist in the absence of a host. It follows that the natural history of the viruses is inseparable from that of their animal hosts, and to reconstruct the evolutionary past of a virus is to understand much about the history of those species it colonizes. Unfortunately, viruses do not leave behind physical remnants of their presence: once a viral species becomes extinct, it vanishes without a trace. At best, the evolutionary history of most viruses can only be inferred indirectly through the phylogenetic comparison of modern, living viruses. There is, however, one prominent exception: the *Retroviridae*. The genomes of all animal species have accumulated (over hundreds of millions of years) the proviral remnants of ancient, largely extinct, retroviral species. This vast archive of viral "fossils" comprises millions upon millions of elements, a tiny fraction of which are coming to light as a consequence of genome sequencing efforts. From the genome of the gray mouse lemur, a diminutive primate (<100 g) found only on the island of Madagascar, Gifford and colleagues (1) have now unearthed a retroviral fossil unambiguously related to the modern AIDS viruses, as reported in this issue of PNAS. How it came to be there is an intriguing, and as yet unsolved, evolutionary mystery.

The Lemur's Tale

The lentiviruses constitute a genus within the *Retroviridae* and include the primate lentiviruses of humans, apes, and Old World monkeys, as well as lentiviruses that have been isolated from sheep, goats, cattle, horses, and cats (2). The most notorious of the primate lentiviruses are the human immunodeficiency viruses (HIV-1 and HIV-2), but there are also some 30 or more lentiviruses indigenous to African primates (the simian immunodeficiency viruses or SIVs). Endemic SIV infections of African primates are generally nonpathogenic, probably reflecting a substantial period of virus/host coevolution. Thus far, lentiviruses have not been detected in the Asian apes or monkeys, or in any species of New World monkey (primates of Central and South America).

Discovery of the first endogenous lentivirus, in the genome of *Oryctolagus cuniculus*, the European rabbit, was reported last year (3). Using an iterative BLAST procedure, Gifford and colleagues (1) have now found small sequence fragments resembling lentiviral sequences in the archives of the *Microcebus murinus* (gray mouse lemur) genome project. Multiple proviral fragments were uncovered, and together with additional sequences recovered by PCR, the authors were able to piece together a consensus viral genome. The reconstructed provirus, which they refer to as pSIVgml (prosimian immunodeficiency virus of gray mouse lemur), clusters with modern primate lentiviruses in phylogenetic analyses incorporating representatives of all known lentiviral taxa. In addition, a comparison of structural features provides new insight into lentiviral evolution. For example, pSIVgml contains a dUTPase domain embedded in the Gag-Pol polyprotein. A dUTPase is present in the same location in some nonprimate lentiviruses, but is not found in HIV/SIV, suggesting that it is an ancestral feature that was lost in the lineage leading to the modern primate lentiviruses. pSIVgml also lacks some of the accessory genes found among the modern primate lentiviruses, such as *Vpr*, *Vpx*, and *Vpu*, indicating that these genes are recently derived features of modern HIV/SIV. pSIVgml has an additional ORF in approximately the same location as the *nef* gene of modern primate lentiviruses, although the predicted amino acid sequence does not significantly resemble any known Nef homologue. It should prove informative, from an evolutionary perspective, to compare and contrast cellular functions of resurrected pSIVgml proteins with their modern counterparts.

ERVs have been likened to fossils, with modern genomes filling the role of geological strata.

Genesis

After a retrovirus enters a cell, the viral RNA genome is converted into double-stranded DNA and inserted irreversibly and at random into the cell's chromosomal DNA. In the event that this process occurs in germ-line DNA—either directly by infection of germ-line tissue or indirectly by infection of embryonic cells destined to differentiate into germ-line tissue—the newly-integrated provirus constitutes a heritable, insertional mutation. Proviral sequences in the germ line are referred to as endogenous retroviruses or ERVs (the term is misleading, because it refers strictly to a sequence's retroviral origin and does not imply the capacity to express infectious virions). Because an ERV-containing locus can be inherited by any or all lineages descending from the original host species, the distribution of the ERV among related taxa is an indication of its relative age. In this regard, ERVs have been likened to fossils, with modern genomes filling the role of geological strata (4).

In the beginning, chromosomes bearing a newly-formed ERV will be exceedingly rare relative to wild-type chromosomes (i.e., lacking the provirus). An upstream battle against the flow of random genetic drift ensues, and the vast majority of ERVs are probably lost to antiquity. With time and luck, an ERV may be passed on more often than not, spreading slowly through the gene pool and eventually achieving fixation [if the ERV happens to confer some benefit on the organism, it might even get a boost along the way from positive selection (5)]. For these reasons, the age of ERV loci may vastly postdate the initial incursion of a retrovirus into a new host. Despite the dismal prospects for any newly-formed provirus, this outcome has been consummated many millions of times during metazoan evolution, and ERVs can outnumber actual genes in the genomes of modern species [including humans (6)].

The retroviral provirus is bracketed by two long terminal repeats (LTRs), with the viral genes arrayed in between. The mechanism of reverse transcription ensures that the 5′ and 3′ LTRs are identical at the moment of integration. As a component of the nuclear genome,

Author contributions: W.E.J. wrote the paper.

The author declares no conflict of interest.

See companion article on page 20362.

[1]E-mail: wjohnson@hms.harvard.edu.

an ERV will accumulate sequence changes over time, and consequently the two LTR sequences will diverge in rough proportion to the age of the ERV. Given a reasonable calibration of the molecular clock, the genetic distance between the LTRs of an ERV can be used to estimate its age. If ERVs are molecular fossils, then divergence between the LTRs is the counterpart to radioisotope dating.

So how and when did this virus arrive in Madagascar? Endogenization does not necessarily follow immediately on the heels of cross-species transmission, and analysis of ERV loci can only provide a minimum estimate of when a virus first colonized an ancestral host. Unfortunately, perhaps because the *M. murinus* genome project is still a work in progress, a pSIVgml provirus with both LTRs intact was not found. Instead, Gifford *et al*. (1) used LTR remnants from two pSIVgml alleles unambiguously descended from the same original provirus to estimate the age of the insertion. One allele still contained what was clearly a 3′ LTR, whereas the other allele was a solo LTR (formed by homologous recombination between the 5′ and 3′ LTRs with concomitant deletion of the intervening sequence). Because recombination could have resolved at any point along the LTR sequence, it is likely that formation of the solo LTR erased some or all of the accumulated differences. Additional substitutions occurring after solo-LTR formation would have further contributed to divergence between the two alleles. Although the interallelic estimate involves a great deal of uncertainty, it provides a conservative lower bound to the true age of the ERV. Using this approach, the authors calculated that the pSIVgml locus in question formed no less than 1.9–3.0 million years ago (MYA).

M. murinus is just one of 50 or more lemur species, and branches on the lemuriform tree represent estimated divergence times ranging from deep (>50 MYA) to more recent (<5 MYA) (7, 8). It should be possible to independently estimate the age of pSIVgml based on the distribution of individual loci among extant species, particularly other mouse lemurs. A thorough assessment could give a precise estimate of when pSIVgml-related viruses last roamed Madagascar, in the process providing a lower bound for the most recent common ancestor of the primate lentiviruses. As more sequence data accumulate from the gray mouse lemur, proviruses containing both LTRs may be identified, which can be used to make additional estimates of insertion times. ERV sequences are also uniquely suited for use as phylogenetic markers, and accurately estimating the age of the SIV-related ERV in lemurs could contribute to investigations of the biogeographical history of Madagascar (9).

The Itinerant Lentivirus

More than 250 miles of open ocean separate Madagascar from mainland Africa, the state of affairs for more than 120 million years. The progenitor of the Malagasy primates arrived sometime later, circa 50–80 MYA (7). To explain the origin of pSIVgml in the genome of the modern gray mouse lemur, while accounting for the presence of related SIVs on the African continent, the authors offer three equally thought-provoking scenarios (others may be envisioned, but they are likely to constitute variations on one of these themes). In the first scenario, the primate lentiviruses predate the arrival of the lemuriform ancestor, and pSIVgml descends from a virus that came along for the ride. If true, the most recent common ancestor of pSIVgml and HIV-1 would have to have been around more than 65 MYA. The second scenario posits a more recent introduction, involving contact between SIV-infected African fauna and the ancestors of the Malagasy primates during hypothetical windows of transient terrestrial access between Madagascar and the mainland (10). In the third scenario, a lentivirus was smuggled from Africa to Madagascar by an unknown third party, perhaps an aerial vector capable of crossing the Mozambique Channel. Vectored transmission has not been documented for any extant retrovirus, making this the most speculative scenario.

Introduction of a lentivirus into a naïve host on Madagascar by either the second or third scenario raises many interesting questions. What might the result of transmission and colonization of the Malagasy primates, long separated from the rest of Africa and the other primate lineages, have been? From modern examples of cross-species transmission, it is clear that such events can be overtly pathogenic (witness HIV-1 and HIV-2 in humans). It could be interesting to ask whether integration times correspond to known extinctions or other major phylogenetic events in lemur history. Is there evidence for cross-species transmissions during the lemuriform radiation? Do the lemur orthologues of antiretroviral genes such as *TRIM5*, *BST2*, and *APOBEC3* reveal distinct episodes of positive selection? Does an exogenous lentivirus, perhaps with the potential for zoonotic transmission, still lurk among the modern lemurs?

A Distant Mirror

The lemurs are endangered, and major efforts are being made to help conserve these species in all their biodiverse glory (http://lemur.duke.edu/conservation/). There are many reasons, ethical and philosophical as well as scientific, for desiring preservation. For evolutionary biologists, lemurs play a significant role in helping to understand our own evolutionary past and our relationship to other primates (11). They are a reference for phylogenetic comparisons, and a baseline for inferring ancestral and derived morphological, physiological, and behavioral characteristics. Now, the lemurs may present a novel opportunity to glimpse the way in which events unfolded during the millennia following introduction of a lentivirus into a primate host, a distant mirror of the modern AIDS epidemic. Hopefully an opportunity gained, not lost.

1. Gifford RJ, *et al.* (2008) A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc Natl Acad Sci USA* 105:20362–20367.
2. Desrosiers RC (2007) Nonhuman lentiviruses. *Field's Virology,* eds Knipe DM, Roizman B, Howley PM, Straus SE, Griffin DE (Lippincott Williams & Wilkins, Philadelphia), pp 2215–2243.
3. Katzourakis A, Tristem M, Pybus OG, Gifford RJ (2007) Discovery and analysis of the first endogenous lentivirus. *Proc Natl Acad Sci USA* 104:6261–6265.
4. Coffin JM (2004) Evolution of retroviruses: fossils in our DNA. *Proc Am Philos Soc* 148:264–280.
5. Arnaud F, Varela M, Spencer TE, Palmarini M (2008) Endogenous retroviruses: Coevolution of endogenous Betaretroviruses of sheep and their host. *Cell Mol Life Sci* 65:3422–3432.
6. Lander ES, *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
7. Horvath JE, *et al.* (2008) Development and application of a phylogenomic toolkit: Resolving the evolutionary history of Madagascar's lemurs. *Genome Res* 18:489–499.
8. Yoder AD (2007) Lemurs. *Curr Biol* 17:R866–R868.
9. Johnson WE, Coffin JM (1999) Constructing primate phylogenies from ancient retrovirus sequences. *Proc Natl Acad Sci USA* 96:10254–10260.
10. Tattersall I (2006) Historical biogeography of the strepsirhine primates of Madagascar. *Folia Primatol (Basel)* 77:477–487.
11. Horvath JE, Willard HF (2007) Primate comparative genomics: Lemur biology and evolution. *Trends Genet* 23:173–182.