Ancient lentiviruses leave their mark

Edward C. Holmes*

Center for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University, University Park, PA 16802; and Fogarty International Center, National Institutes of Health, Bethesda, MD 20892

One of the most remarkable insights from the human genome project was that more genetic material is devoted to vertically transmitted and defective remnants of once-infectious retroviruses, commonly referred to as endogenous retroviruses (ERVs), than to human protein-coding DNA (1). Current estimates place the ERV component of the human genome at 5–8%, comprising at least 31 distinct families, the largest of which, HERV-H, carries ≈1,300 full-length copies (2). Similarly high numbers are likely to hold for other mammalian species and paint a picture in which ERVs are ubiquitous genomic elements; thus, the exogenous ancestors of ERVs must have plagued mammalian species for millennia. The paper by Katzourakis *et al.* (3) in this issue of PNAS now shows that this is true of the retroviruses associated with the one of the most serious of all human diseases.

Despite such a wealth of retroviral DNA in the mammalian genome, in only a few cases has it proven possible to identify the infectious progenitors of these now-endogenous passengers. To date, the best-documented examples are koala retrovirus (KoRV), in which the process of endogenization has been dramatically captured in real time (4), as well as mouse mammary tumor virus (MMTV), murine leukemia virus (MLV), and Jaagsiekte sheep retrovirus (JSRV). Highly notable absentees were the lentiviruses, most famous in their manifestation as the human immunodeficiency viruses (HIV-1 and -2), the causative agents of AIDS, as well as a rogue's gallery of viruses responsible for chronic mammalian diseases that have been familiar to veterinarians for decades (the prefix "lenti" is derived from the Latin for slow, *lentus*, marking their distinctive pathology). Katzourakis *et al.* (3) fill this gap, providing the first description of an endogenous member of the lentivirus group, christened RELIK.

Using a suite of bioinformatic tools based on the BLAST algorithm, Katzourakis *et al.* (3) identify 25 full-length but clearly defective copies of RELIK in the genomes of European rabbits, members of the order *Lagomorpha*. This is the fifth mammalian order identified to carry a lentivirus in some form. However, no exogenous lentiviruses have yet been identified in rabbits, and in no case do the same species carry both endogenous and exogenous lentiviruses. Through an intricate analysis of patterns of nucleotide substitution at synonymous and nonsynonymous sites as markers of selection pressure, Katzourakis *et al.* are also able to show that RELIK has spread by both the production of exogenous virus, indicative of its infectious past, and intracellular retrotransposition, common in other ERVs.

The Time Scale of Lentivirus Evolution

The discovery of RELIK is of more importance than simply filling an annoying gap in retrovirus biodiversity. Most notably, it sheds new light on the antiquity of the lentiviruses, an issue that has intrigued virologists since their discovery. The major difficulty in estimating the age of an ERV such as RELIK is that rates of evolutionary change differ dramatically between endogenous (low-rate) and exogenous (high-rate) retroviruses. Consequently, if it is assumed that ERVs evolve at the same rate as genomic DNA but in fact have incorporated mutations that have arisen through replication with the notoriously error-prone retroviral enzyme reverse transcriptase, rate estimates would be seriously biased. To bypass this problem, Katzourakis *et al.* (3) estimate the number of substitutions between duplicated copies of RELIK, reasonably assuming that they have evolved at the same rate as the rest of the rabbit genome since their divergence. Such an analysis leads to an estimated time of origin of RELIK of at least 7 million years. Although there is necessarily some degree of uncertainty in such estimates, it is likely that RELIK has indeed been passively associated with rabbit genomes for millions of years. Such antiquity sits in marked contrast to studies of exogenous lentiviruses, such as HIV. Although there are long-standing suggestions that the primate lentiviruses, HIV and its simian relatives the simian immunodeficiency viruses (SIVs), have codiverged with their primate hosts over many millions of years (5), such claims are at odds with "molecular clock" estimates of their divergence times, which are usually no more than a few thousand years, even under the most liberal assumptions (6). Indeed, that all exogenous retroviruses studied to date, with the notable exception of the simian foamy viruses (7), evolve at very rapid rates means that studying cellular replicating endogenous genomes is the only realistic way to explore their origin and deep phylogenetic history; rapid evolutionary rates mean that the signal of ancestry history in infectious retroviruses is quickly eroded by mutation accumulation. More starkly, there are growing examples of more recent cross-species transmission events within the primate lentiviruses. Although that from chimpanzees (SIVcpz) to humans (HIV-1) is the most famous, species jumping now seems responsible for the origin of SIV in both chimpanzees (8) and, more recently, gorillas (SIVgor; ref. 9). Similarly, patterns of genetic diversity in another group of exogenous lentiviruses, the feline immunodeficiency viruses (FIVs) infecting cougars in North America, seemingly reflect changes in host-population demography on the timescale of decades (10).

A key task in mammalian evolutionary genomics should be ongoing surveillance.

The discovery of RELIK clearly puts the lentiviral group deeper in mammalian evolution than previously imagined and in doing so complicates theories for their origin. An antiquity of 7 million years necessarily means that infectious lentiviruses must be of at least the same age and, given their propensity to jump species boundaries, it seems evident that the number of mammalian orders found to harbor lentiviruses will steadily increase (despite being one of our closest relatives, it took >20 years of surveying primate species to identify SIVgor). Although Katzourakis *et al.* (3) suggest that the European rabbit might represent the progenitor species of all lentiviruses,

Author contributions: E.C.H. wrote the paper.

The author declares no conflict of interest.

See companion article on page 6261.

*E-mail: ech15@psu.edu.

that expansive genome-sequence data sets are to date available for only a tiny fraction of mammalian species makes it highly likely that other endogenous (and exogenous) lentiviruses will be discovered as more genome-sequencing projects are completed.

The discovery of RELIK is also important for understanding the evolution of genome complexity in retroviruses. Although RELIK contains some of the accessory genes that characterize the lentiviruses and make them rather more complex than other retroviruses, notably *tat* and *rev*, it notably lacks the *vif* gene present in all other lentiviruses identified to date, with the exception of equine infectious anemia virus (EIAV) from horses. Therefore, either EIAV and RELIK share a common ancestry and *vif* evolved after their separation, perhaps by a host gene capture, or *vif* was independently lost in EIAV and RELIK. Unfortunately, extensive sequence divergence means that there is insufficient signal at the base of the lentivirus tree to help tease apart these different evolutionary scenarios. As such, the future sampling of lentiviruses from other host species represents the most powerful way to resolve this important episode of genome evolution. The study of *vif* is of more importance than simply resolving phylogenetic history; *vif* has become a top target for study by retrovirologists because it represents a viral counterstrategy to host-induced deamination mutagenesis, a process mediated by the APOBEC family of genes found in mammals (11). Although it is tempting to speculate that the lack of *vif* was in part responsible for the successful host control of the exogenous form of rabbit lentivirus, the sequences of RELIK are again so divergent as to prevent a clear-cut examination of the telltale monotonous G-to-A mutational changes that mark the action of APOBEC.

Reconstructing the Past

Although the discovery of RELIK has undoubtedly shed important new light on the origin and evolution of the lentiviruses, perhaps the most lasting lesson is that we are still far from the bottom of the well of retroviruses in general and of lentiviruses in particular. Indeed, a key task in mammalian evolutionary genomics should be ongoing surveillance for remnants of these highly important pathogens as well as for research on their intrinsic evolutionary dynamics and, more intriguingly, how they might affect the patterns and processes of host evolution. Crucially, these analyses can go far beyond the relatively straightforward inference of phylogenetic history because it is now possible to repair and then reconstruct ancient retroviruses by using the consensus sequences of their current endogenous relatives, therein providing a unique insight into retroviral diseases of the past (12). Because Katzourakis *et al.* (3) have compiled a complete consensus of RELIK, reconstructing an infectious copy of this virus is now a distinct possibility. Such a reconstructed virus will provide vital new information on the basics of lentivirus function and evolution, including the evolutionary implications of a lack of *vif*, the intrinsic rates of mutation and recombination, and the determinants of host species range, and in doing so will further blur the boundary between the endogenous and infectious forms of these ubiquitous parasites.

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al.* (2001) *Nature* 409:860–921.
2. Katzourakis A, Tristem M (2005) in *Retroviruses and Primate Genome Evolution*, ed Sverdlov ED (Landes Bioscience, Austin, TX), pp 186–203.
3. Katzourakis A, Tristem M, Pybus OG, Gifford RJ (2007) *Proc Natl Acad Sci USA* 104: 6261–6265.
4. Tarlinton RE, Meers J, Young PR (2006) *Nature* 442:79–81.
5. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, *et al.* (1999) *Nature* 397: 436–441.
6. Holmes EC (2003) *J Virol* 77:3893–3897.
7. Switzer WM, Salemi M, Shanmugam V, Gao F, Cong M-E, Kuiken C, Bhullar V, Beer B, Vallet D, Gautier-Hion A, *et al.* (2005) *Nature* 434:376–380.
8. Bailes E, Gao F, Bibollet-Ruche F, Courgnaud V, Peeters M, Marx PA, Hahn BH, Sharp PM (2003) *Science* 300:1713.
9. Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, Liu W, Loul S, Butel C, Liegeois F, Bienvenue Y, *et al.* (2006) *Nature* 444:164.
10. Biek R, Drummond AJ, Poss M (2006) *Science* 311:538–541.
11. Sheehy AM, Gaddis NC, Choi JD, Malim MH (2002) *Nature* 418:646–650.
12. Lee YN, Bieniasz PD (2007) *PLoS Pathog* 3:e10.