Macroevolution of Complex Retroviruses

Aris Katzourakis,[1,2]* Robert J. Gifford,[1]* Michael Tristem,[3] M. Thomas P. Gilbert,[4] Oliver G. Pybus[1]

Retroviruses are unusual among microorganisms in leaving a "fossil record" in the form of endogenous viral insertions within their hosts' genomes. However, insertions before the radiation of mammalian orders lack both retroviral accessory genes and contemporary infectious relatives (*1*) and thus provide limited insight into the coevolution of complex retroviruses and mammalian innate antiviral responses (*2*). We have found that foamy viruses, complex retroviruses currently infecting many mammals, are present within the genomes of sloths. By combining phylogenetic, genomic, and biogeographic methods, we have established that foamy viruses circulated among ancestral mammals >100 million years ago (Ma), demonstrating the survival of an infectious lineage of complex retroviruses across an entire geological era.

We screened all available mammalian genomes and identified foamy virus insertions only in the two-toed sloth, *Choloepus hoffmanni*, which belongs to the basal mammalian group Xenarthra. These sequences, termed SloEFV (sloth endogenous foamy virus), group robustly with contemporary infectious foamy viruses in phylogenies (*3*). We constructed a consensus ~11.5-kb SloEFV genome that could be aligned with contemporary strains and that exhibited typical foamy virus characteristics, including sequences similar to the accessory genes *tas* and *bet* (*3*).

Genomic analysis points to an ancient SloEFV origin. *C. hoffmanni*'s genome contains hundreds of SloEFV elements, most of which have lost their coding regions through recombination. Only 72 elements contained >1 kb of coding region, all of which carried numerous stop codons, frame shifts, insertions, and deletions. We used two independent approaches to determine the age of the SloEFV germline invasion. First, we measured genetic distances among eight elements that were identified as having arisen via host genome duplication events. By applying a xenarthran-specific rate of neutral evolution (*3*), we obtained a minimal SloEFV age of 39 million years (range from 34 to 45 million years; minimal because the earliest duplication event must postdate SloEFV germline invasion). Second, we determined the taxonomic distribution of SloEFV by polymerase chain reaction screening of tissue samples from five further xenarthran species: Insertions were detected in two- and three-toed sloths but were absent from anteaters and armadillos. SloEFV sequences from four sloth species did not cluster phylogenetically; indeed, we observe phylogenetically robust grouping of elements from different species (*3*), conservatively indicating that SloEFV entered the xenarthran germline before the divergence of two- and three-toed sloths (~21 Ma) but after the anteater-sloth split (~55Ma) (*4*). Before this, foamy viruses circulated in xenarthran ancestors in an exogenous form.

These minimum ages are highly important in the context of mammalian biogeography: Xenarthrans diverged from other mammals ~105 Ma (*5*) and were isolated in South America as landmasses separated throughout the early Cenozoic (Fig. 1A). Mammals such as primates and rodents that evolved outside South America likely arrived there after 30 Ma (*6*), precluding them from having introduced SloEFV to xenarthrans upon their arrival. Phylogenetic comparison of foamy viruses and their hosts strongly supports a coevolutionary history tracing back to the origin of mammals: Topologies of virus and host phylogenies exactly match (Fig. 1B), and viral branch lengths are significantly correlated with mammalian divergence times [R^2 = 0.74; $P < 0.0001$, robust to the exclusion of the sloth branch and of primates (*7*)]. See (*3*) for details.

The descendants of foamy viruses that infected ancestral mammals during the Cretaceous thus diverged in concert with their hosts throughout the Cenozoic and have persisted in a surprisingly unchanged form until today, supporting the idea that evolutionary constraint can maintain viral genomic conservation over many millions of years despite exceptionally high short-term rates of mutation (*8*). Furthermore, SloEFV's identification indicates that retrovirus accessory genes and mammalian mechanisms of innate immunity will be best understood when considered as the joint products of macroevolutionary conflict played out over a geological time scale.

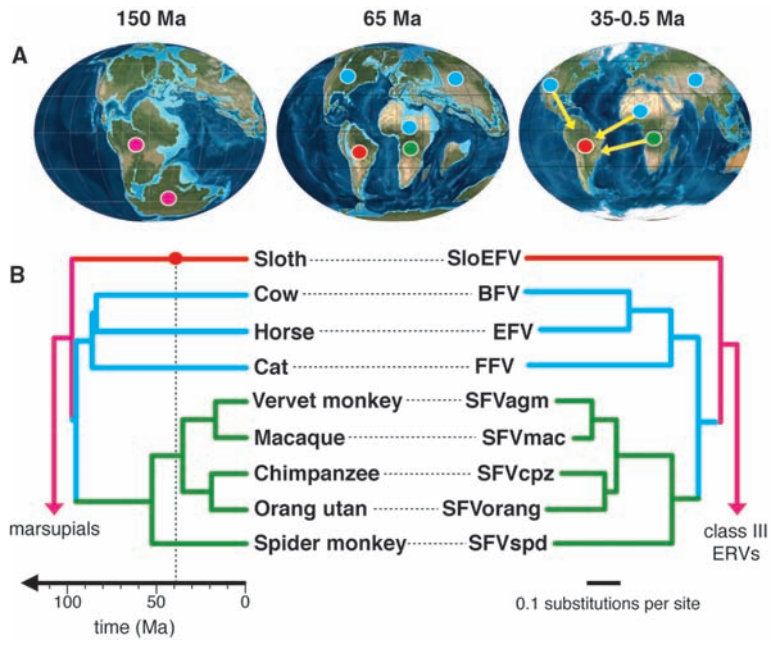

Fig. 1. (**A**) Estimated positions of Cenozoic landmasses and mammalian lineages. At 150 Ma, Old and New World landmasses were connected, and ancestral eutherian mammals (pink) existed within this range. From ~100 to ~30 Ma, xenarthrans (red) were geographically isolated from primates (green) and other boreoeutherian mammals (blue). SloEFV invasion of the sloth genome (>39 Ma, range from 34 to 45 Ma) thus occurred before the arrival of boreoeutherians in South America. (**B**) Foamy virus phylogeny (right; scale bar indicates amino acid changes per site) is congruent with that of mammals (left; scale bar, Ma) (*5*). Red circle denotes 39 Ma, our estimate of the latest possible date of SloEFV genome invasion.

References and Notes

1. R. A. Weiss, *Retrovirology* **3**, 67 (2006).
2. D. Wolf, S. P. Goff, *Annu. Rev. Genet.* **42**, 143 (2008).
3. Materials and methods are available as supporting material on *Science* Online.
4. F. Delsuc, S. F. Vizcaino, E. J. P. Douzery, *BMC Evol. Biol.* **4**, 11 (2004).
5. O. R. P. Bininda-Emonds *et al.*, *Nature* **446**, 507 (2007).
6. D. Huchon, F. M. Catzeflis, E. J. P. Douzery, *Proc. R. Soc. London Ser. B* **267**, 393 (2000).
7. W. M. Switzer *et al.*, *Nature* **434**, 376 (2005).
8. E. C. Holmes, *J. Virol.* **77**, 3893 (2003).
9. We thank the Washington University St. Louis Genome Sequencing Center for the *C. hoffmanni* genome assembly; B. Voirin, A. Hedayat, and the University of Copenhagen Zoological Museum for xenarthran tissue samples; and R. Blakey for tectonic maps. A.K. was funded by the James Martin 21st Century School, M.T.P.G. by the Danish Natural Science Research Council, and O.G.P. by the Royal Society. GenBank accession numbers are GQ169506 to GQ169527.

Supporting Online Material

www.sciencemag.org/cgi/content/full/325/5947/1512/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References

27 March 2009; accepted 9 July 2009
10.1126/science.1174149

[1]Zoology Department, University of Oxford, Oxford OX1 3PS, UK. [2]Institute for Emergent Infections, University of Oxford, Oxford OX1 3PS, UK. [3]Division of Biology, Imperial College London, London SL5 7PY, UK. [4]Natural History Museum of Denmark, Copenhagen University, 1350 Copenhagen, Denmark.

*To whom correspondence should be addressed. E-mail: aris.katzourakis@zoo.ox.ac.uk (A.K.); robert.gifford@zoo.ox.ac.uk (R.J.G.)