

Discovery and analysis of the first endogenous lentivirus

Aris Katzourakis*, Michael Tristem†, Oliver G. Pybus*, and Robert J. Gifford*‡§

*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom; †Division of Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, United Kingdom; and ‡Division of Infectious Diseases, Stanford University, Stanford, CA 94305

Edited by John M. Coffin, Tufts University School of Medicine, Boston, MA, and approved February 13, 2007 (received for review January 17, 2007)

The lentiviruses are associated with a wide range of chronic diseases in mammals. These include immunodeficiencies (such as HIV/AIDS in humans), malignancies, and lymphatic and neurological disorders in primates, felids, and a variety of wild and domesticated ungulates. Evolutionary analyses of the genomic sequences of modern-day lentiviruses have suggested a relatively recent date for their emergence, but the failure to identify any endogenous, vertically transmitted examples has meant that their longer term evolutionary history and origin remain unknown. Here we report the discovery and characterization of retroviral sequences belonging to a new lentiviral subgroup from the European rabbit (*Oryctolagus cuniculus*). These viruses, the first endogenous examples described, are >7 million years old and thus provide the first evidence for an ancient origin of the lentiviruses. Despite being ancient, this subgroup contains many of the features found in present-day lentiviruses, such as the presence of *tat* and *rev* genes, thus also indicating an ancient origin for the complex regulation of lentivirus gene expression. Although the virus we describe is defective, reconstruction of an infectious progenitor could provide novel insights into lentivirus biology and host interactions.

phylogeny | retrovirus | ERV | rabbit

Lentiviruses have been intensively studied since the emergence of human immunodeficiency viruses (HIV-1 and HIV-2) in the early 1980s. They have been divided into five subgroups, each restricted to a single mammalian order or family (1). All are characterized by extremely high rates of evolution, a feature that has facilitated the detailed reconstruction of recent evolutionary history for some subgroups. In particular, phylogenetic analysis of the primate lentiviruses has established the approximate timing and geographic locations of the transmission events that preceded the global HIV-1 and HIV-2 pandemics (2–6).

All known lentiviruses are exogenous (transmitted horizontally from host to host) and are only distantly related to endogenous, germ-line retroviruses (7–11). Relatively recent dates for the emergence of the modern lentiviral subgroups have been suggested, but these estimates are based on extrapolations of genetic distances observed among contemporary viruses, which are likely to be unreliable for evolutionary comparison of distantly related sequences (12–15). Thus, the lack of endogenous “genomic fossils” has precluded studies into the longer term evolutionary history and origin of lentiviruses (14, 15).

Here we report the discovery and analysis of the first endogenous lentivirus, present in the genome of the European rabbit, which we term rabbit endogenous lentivirus type K (RELK). The discovery of RELK extends the host range of lentiviruses to a fifth mammalian order and demonstrates that such viruses can also spread by vertical transmission and intragenomic proliferation, in addition to infectious horizontal transfer. By dating the invasion of the rabbit genome by the RELK virus, we provide the first direct evidence that lentiviruses are ancient in origin.

Results

Construction and Analysis of a Consensus Sequence. We systematically screened publicly available, genome scale sequence databanks

and trace archives using BLAST algorithms and sequence probes derived from the enzymatic and structural proteins of representative lentiviruses. This process identified numerous, highly significant matches within the whole genome shotgun assembled sequence of the European rabbit (*Oryctolagus cuniculus*). Further analysis revealed complete retroviral Gag, Pol, and Env coding domains, all of which were multiply defective, containing numerous in-frame stop codons and frameshift mutations.

To reconstruct the original genomic organization of RELK, we assembled a consensus sequence based on all rabbit genome matches with >2 kb of lentivirus sequence identity (34 in total). Sequences were aligned initially by using BlastAlign (16) and manually adjusted to identify ORFs. Because the whole genome shotgun rabbit sequences we recovered were typically 2–10 kb in length (thus rarely containing a complete viral sequence), some of these matches probably represent different parts of the same proviral insertion. Further analysis indicated that the low-coverage (2×) assembly of the rabbit genome currently available contains ≈25 full-length viruses and 150 solo LTRs; these are generated from full-length endogenous retroviruses via recombination between the two LTRs found flanking the internal coding region (17). Approximately one-third of the full-length RELK copies have resulted from segmental duplication of the rabbit genome (discussed below).

The resulting consensus sequence [supporting information (SI) Fig. 5] contained long ORFs encoding Gag, Pol, and Env polyproteins. Two additional proteins, Tat and Rev, were identified by sequence similarity to the accessory proteins of other lentiviruses (see Figs. 1 and 2). The *tat* gene is located immediately downstream of *pol*, partially overlapping the *env* gene, whereas *rev* is located toward the 3' end of *env* in an alternative reading frame; *tat* and *rev* are found in similar locations in other lentiviruses. Other lentiviral features of the RELK consensus include (i) a putative TAR (transactivation responsive region) downstream of the viral promoter (SI Fig. 6a), (ii) a putative RRE (Rev responsive element) within *env* (SI Fig. 6b), (iii) a ribosomal frameshifting site at the start of *pol* (SI Fig. 6c), (iv) a dUTPase between RNaseH and Integrase in *pol* as found in nonprimate lentiviruses (SI Fig. 5), (v) a primer binding site with exact identity to the conserved lentiviral lysine tRNA (SI Fig. 5), and (vi) marked nucleotide composition bias, with a strong preference for adenine (A) over guanine (G) and thymine (T) over cytosine (C) across the entire coding region (SI Fig. 7).

Author contributions: A.K., M.T., and R.J.G. designed research; A.K., M.T., and R.J.G. performed research; A.K. and R.J.G. contributed new reagents/analytic tools; A.K., M.T., and R.J.G. analyzed data; and A.K., M.T., O.G.P., and R.J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: RELK, rabbit endogenous lentivirus type K; ML, maximum likelihood.

See Commentary on page 6095.

§To whom correspondence should be addressed. E-mail: rjmg@stanford.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0700471104/DC1.

© 2007 by The National Academy of Sciences of the USA

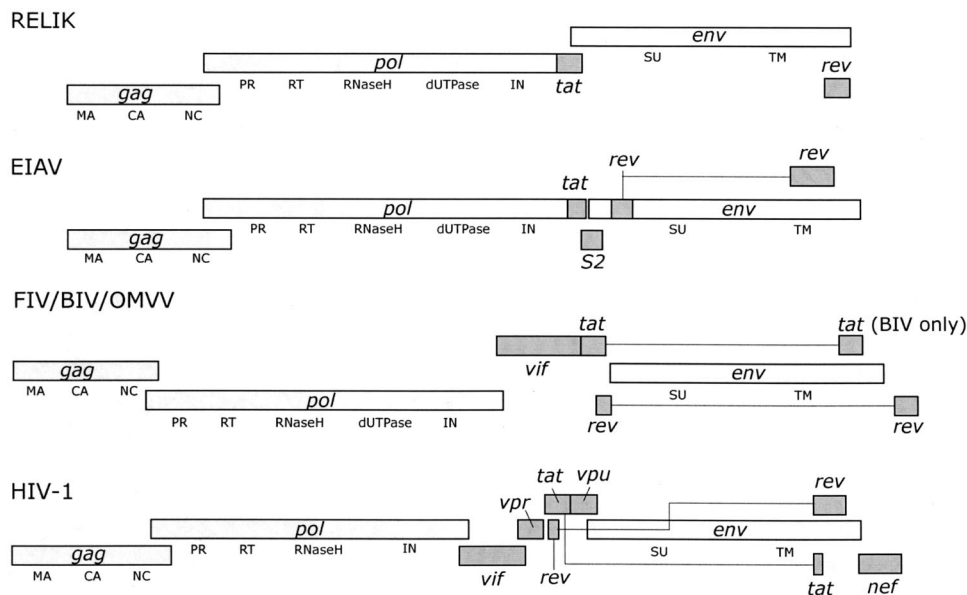


Fig. 1. Lentiviral genome organizations. The genome organization of the consensus RELIK provirus is shown in comparison to representative genome organizations of other lentiviral subgroups: equine (equine infectious anemia virus, EIAV), feline (feline immunodeficiency virus, FIV), bovine (bovine immunodeficiency virus, BIV), caprine/ovine (caprine arthritis–encephalitis virus, CAEV; ovine maedi–visna virus, OMVV), and primate lineages. Most primate lineages lack a *vpu* gene.

Phylogenetic Reconstruction. To establish the level of support for the placement of RELIK within the lentiviruses, we aligned the relatively conserved reverse transcriptase domain of Pol together with a diverse selection of exogenous retroviral reverse transcriptases. Phylogenetic analysis using both maximum likelihood (ML) and Bayesian methods placed RELIK within the lentiviruses with robust phylogenetic support (Fig. 3a). To date, lentiviruses have been isolated from a relatively restricted range of species, and the identification of RELIK extends lentiviral host range to a fifth mammalian order.

To explore the relationship of RELIK to other lentiviruses in more detail, we reconstructed phylogenies from an alignment of the entire *pol* gene (Fig. 3b). This showed some support for the clustering of RELIK with equine infectious anemia virus, which is also consistent with their similar genomic organization (Fig. 1). However, as is commonly observed in lentiviral phylogenies, support for the deeper phylogenetic relationships was much weaker than for the grouping of species within subgroups (9). Trees based on Gag and the transmembrane (TM) region of Env did not reveal any obvious recombination events between RELIK and other lentiviruses. The genetic divergence of the RELIK clade is sufficiently great to indicate that it should be classified as a new lentiviral subgroup.

Mode of Proliferation. Genomic proliferation of endogenous retroviruses can occur via several mechanisms, including intracellular

retrotransposition, either in *cis* or in *trans*, and retroviral reinfection of germ-line cells (18). These mechanisms can be distinguished by looking for evidence of past purifying selection on *gag*, *pol*, and *env*, as described in previous studies (19, 20). The signature of the different mechanisms of replication can be observed in the internal branches of endogenous retrovirus phylogenetic trees allowing the largely neutral evolution represented by the terminal branches to be ignored (19, 20). Genes not used by a particular mechanism of proliferation will evolve neutrally and thus will quickly accumulate inactivating mutations. Purifying selection on the *env* gene is indicative of extracellular reinfection because this gene is necessary only for movement between cells (19). We found that both *gag* and *pol* were under significant purifying selection, having ratios of nonsynonymous-to-synonymous substitution rates (dN/dS) of 0.36 on internal branches ($n = 10$; the neutral evolution hypothesis of dN/dS = 1 was rejected at $P = 0.001$) and 0.26 ($n = 9$, dN/dS = 1 was rejected at $P < 0.0001$), respectively. The *env* gene data set was subdivided into two because all but five elements contained an ≈ 1 -kb-long deletion. The five elements containing a full-length *env* gene were under significant purifying selection (dN/dS = 0.34, $n = 5$, dN/dS = 1 was rejected at $P < 0.0001$ on internal branches), whereas the remainder did not depart significantly from a dN/dS of 1 (dN/dS = 0.82, $n = 8$, $P = 0.63$ on internal branches). It therefore appears that the subset of RELIK elements containing full-length *env* genes under purifying selection has been replicating by reinfection of germ-line cells (i.e., involving the release of exogenous viral particles). However, a higher proportion of elements have been replicating via intracellular retrotransposition in *cis*, as indicated by the relaxed selection on their *env* genes, but purifying selection on their *gag* and *pol* genes.

Dating the Invasion of the Rabbit Genome by RELIK. Multiple lines of evidence indicate that RELIK invaded the rabbit genome millions of years ago. First, RELIK sequences are all multiply defective, and several also contain short interspersed nuclear element insertions. For example, one element (present on AAGW01148698) has five in-frame stop codons, 11 frameshift mutations (in coding regions), and two short interspersed nuclear element insertions compared with the consensus sequence.

RELIK (1-57)	MEYQLQLSCQKNSCVCKKCR--YHCQLCPLQKELGISYSRARTKELQKWOQQQTEKCT
HIV-1 (13-68)	HPGSQPKTACTN-CYCKKCC--FHCQVCFITKALGISYGRKKRRRQRRAPQGSQTHQVS
SIVAGM (28-79)	EEYYQLQACENKCKWCKKCC--FHCMLCFHKKGLGIRYHVY--RKRGP---GTNKKIPG
HIV-2 (41-95)	AQLHRPLEPCTNKCYCKRCS--FHCQLCFSSKGLGISYERKGRRRRTP--RKTTPSPS
BIV (29-82)	NTCKGPRRDCPH-CCCPICCS--WHCQLCPLQKLNGLVNYGSGPRRGRTR--GKGRRIIRT
EIAV (26-77)	SSGGVPGQNTG--GQEARPN--YHCQLCPLR-SLGDYLDASLRKKNK--QLKAIQQG
FIV (43-END)	IWRDRFKVFNPRGCLCWWCKLYYQLE--QSTLSISSA-----QRKAIQQG
OMVV (63-END)	WWMKNANAPIRRNCGRLCPNGWGSQVRDVEL-----

Fig. 2. Alignment of the putative RELIK Tat protein with the Tat proteins of other lentiviruses. Numbers refer to the relative positions aligned. Boxes denote identical amino acid residues, and bold type indicates residues with similar properties. Lentiviral Rev proteins share little sequence similarity, but we did identify putative nuclear localization and nuclear export signals (34), which are underlined in SI Fig. 5.

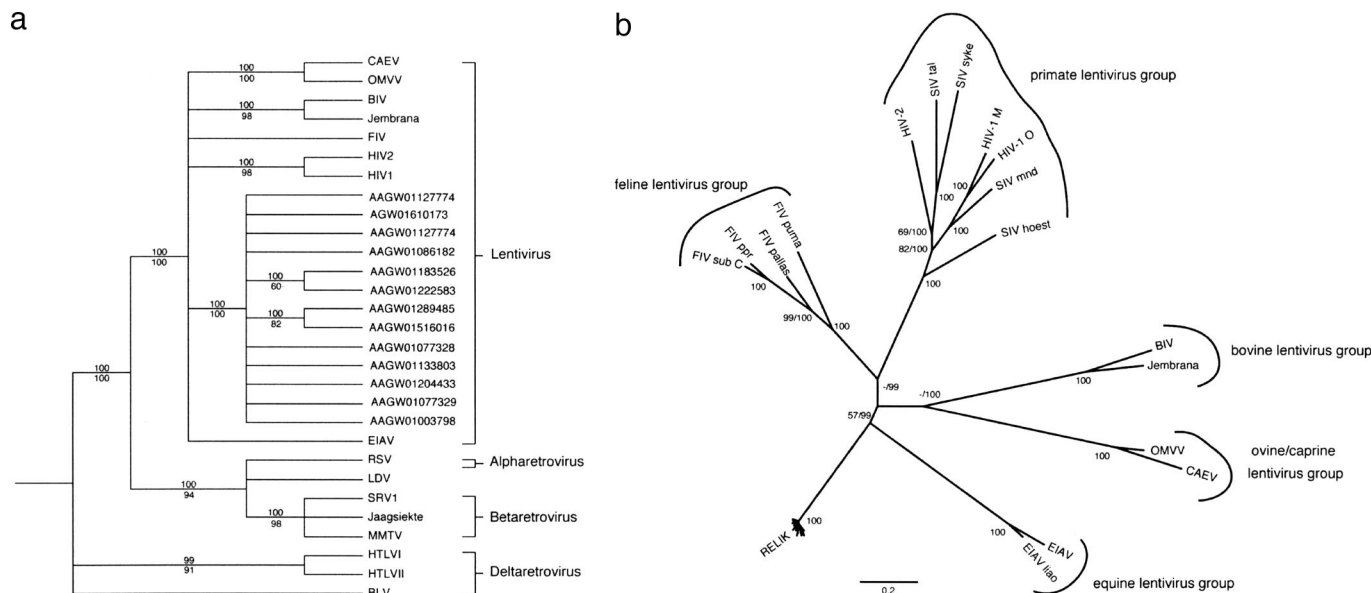


Fig. 3. Phylogenetic relationships of RELIK to other retroviruses. (a) Phylogeny of RELIK and other lentiviruses together with a sample of nonlentiviral exogenous retroviruses, rooted on BLV, HTLV1, and HTLV2. Support for the ML trees was assessed via 1,000 nonparametric bootstrap replicates, and posterior clade probabilities were assessed for the Bayesian phylogeny. Both support indices are indicated as values out of 100, with posterior clade probabilities indicated above the branches and bootstrap scores below. Branches with posterior probabilities <95% were collapsed. The RELIK sequences are indicated by their accession numbers. (b) Unrooted phylogeny showing the relationships of the lentiviruses. Bootstrap scores and Bayesian posterior probabilities are indicated to the left and right of the forward slash, respectively, and nodes with only 100 indicated showed maximal support under both measures.

Second, the high number of RELIK full-length and solo LTR sequences points to an established germ-line infection. Finally, approximately one-third of RELIK sequences arose via segmental duplication events within the rabbit genome; hence, the RELIK germ-line invasion must be older than these events.

We identified and dated three pairs of RELIK elements that obviously arose by segmental duplication (each duplicate pair contained at least 1.5 kb of RELIK sequence and an unambiguous shared flanking site) (see Fig. 4 and SI Fig. 8). The identification of several indels, including a unique short interspersed nuclear element insertion into one of the paired RELIK elements but not the other, and the presence of a pronounced transition/transversion mutation bias, confirmed that each pair

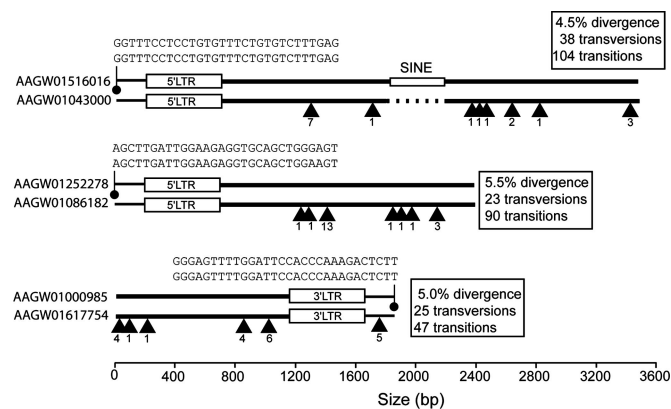


Fig. 4. Schematic of the three segmental duplications used for dating RELIK insertions. Open boxes show RELIK LTRs and the location of the single short interspersed nuclear element insertion, and flanking sequences are indicated by thinner lines. Indels between the sequences are represented by solid triangles, with the number of inserted/deleted bases involved below. Thirty bases of flanking sequences are shown for each segmentally duplicated pair. Percentage divergence was calculated by using the GTR model with all parameters estimated from the data.

resulted from segmental duplication, rather than by genomic sequencing errors (Fig. 4). The RELIK sequences within each pair differed by 0.045, 0.05, and 0.055 substitutions per site, the result of a gradual accumulation of neutral mutations during host replication. Half of the genetic distance between the two members of a segmentally duplicated pair, calibrated according to the host neutral substitution rate, therefore provides an accurate estimate of the date of the duplication event, and hence a minimum estimate for the age of the RELIK elements (pairwise distances were estimated by using the general time reversible model of substitution). The rate of neutral substitution for the European rabbit has been estimated at $\approx 4 \times 10^{-9}$ substitutions per site per year (21), intermediate between the human and mouse rates of 2.2×10^{-9} and 4.5×10^{-9} , respectively (22, 23). This corresponds to estimated minimum integration dates for the three pairs of 5.7, 6.3, and 7 million years. It is unlikely that invasion of the rabbit germ line occurred significantly earlier than the oldest of these dates because the median root-to-tip distance in a phylogenetic tree of RELIK sequences is 0.043 substitutions per site, corresponding to a maximum age of 10.8 million years assuming neutral evolution. However, our analysis of dN/dS ratios shows that some of this divergence is due to nucleotide substitution occurring during viral replication (i.e., at a much faster rate than host-induced mutation).

Discussion

Although numerous retroviral insertions have been characterized in vertebrate genomes, none has previously revealed significant homology to exogenous lentiviruses (1, 8–10, 24, 25). This has been taken as reflecting a recent origin for the lentiviral genus (such that there has not been sufficient time for genome invasion to occur), or a biological barrier to genome invasion, arising either from lentiviral mechanisms for *trans* regulation of gene expression (10) or the lack of specific receptors for these viruses on germ cells (11). The discovery of endogenous lentiviral insertions containing putative *tat* and *rev* genes and corre-

sponding RNA secondary structural motifs demonstrates that germ-line infection by lentiviruses can occur and that lentiviral mechanisms for complex regulation of expression do not preclude successful germ-line colonization.

Furthermore, the identification of lentiviral genomic fossils enables us to provide the first estimates for the minimum age of the lentivirus group based on genomic data. Estimation was assisted by the identification of pairs of insertions that had arisen through segmental duplication of the host genome (Fig. 4). Analysis of these duplicated insertions provides a minimum estimate for the age of the RELIK subgroup of between 5.7 and 7 million years, significantly predating the estimated ages of other lentiviral subgroups. For example, estimates based on epidemiological and phylogenetic studies of exogenous viruses have suggested that the feline and primate subgroups originated probably no more than 1–2 million years ago (13, 14). Thus, it appears that the lentivirus genus may be significantly older than has previously been demonstrable.

Retroviruses are thought to have evolved increasing genomic complexity over time (26, 27), and it therefore seems likely that the precursor to the modern lentiviruses had a relatively simple genomic organization. Consistent with this, and the apparently ancient origin of the endogenous insertions, we note that RELIK has the simplest genomic organization of all the lentiviral subgroups and lacks a *vif* gene, a characteristic it shares with equine infectious anemia virus.

Endogenous lentiviruses appear to be extremely rare. The European rabbit is the first species identified as harboring lentiviral insertions despite the vast quantity of mammalian genome sequence data now available. We screened 46 mammalian genome libraries containing $>10^8$ bp of sequence data, of which 34 contained $>10^9$ bp. Of these, only the European rabbit genome appears to harbor endogenous lentiviruses.

Based on the above, it is tempting to speculate that lentiviruses in rabbits, or perhaps other, related lagomorphs, might represent the precursors of modern exogenous lentiviruses. In accordance with this idea, the ancestral geographic range of the European rabbit (southern Europe and northwest Africa) overlaps that of many species now harboring exogenous lentiviruses, including cattle, horses, and (wild) cats (28). This hypothesis requires RELIK to be placed basal in lentiviral phylogenies. The phylogenies presented (Fig. 3) are similar to those described previously (9), in that the deeper phylogenetic relationships show low levels of support, and the precise order of branching events is therefore difficult to ascertain. Future identification of related lentiviruses in other lagomorph or mammalian genomes may shed light on this issue and provide further support for the timescale of lentiviral evolution that we propose.

Further biological characterization of RELIK may deliver advances in efforts to treat and prevent lentiviral disease.

Although RELIK is an ancient lentivirus and only defective copies were identified in this analysis, recent research has shown that it is possible to reconstruct infectious progenitors of such viruses (29). Thus, the reconstruction of an infectious RELIK lentivirus could potentially provide a novel small animal model for experimental lentiviral research. It also remains possible that exogenous, RELIK-like lentiviruses are still circulating in some rabbit populations or other lagomorphs.

Materials and Methods

Phylogenetic Reconstruction. Two phylogenies were reconstructed, the first to establish the level of support for the placement of RELIK within the lentiviral genus, and the second to determine the relationship of RELIK to the five previously identified lentiviral subgroups. For the first phylogeny we aligned 159 amino acids of the relatively conserved reverse transcriptase domain in Pol from a diverse selection of exogenous retroviruses. We used the rtREV amino acid substitution matrix, developed specifically for retroviral phylogenetic inference (30). Both ML and Bayesian methods were used by using the programs PHYML and MrBayes 3 (31, 32). Bayesian analyses were run for 5 million steps, sampling trees every 100 steps, and discarding the first 5,000 trees. Support for the ML trees was assessed via 1,000 nonparametric bootstrap replicates. For the second data set we aligned the whole Pol polyprotein and identified 590 unambiguously aligned amino acids from the reverse transcriptase, RNaseH, and Integrase domains. Phylogenies were reconstructed by using the underlying nucleotide sequence data, under the GTR+ Γ model of substitution. ML phylogenies were estimated in PAUP, with support estimated as for the first data set.

Selection Analysis and dN/dS Ratios. We used the “two-ratio” model in PAML (33) to estimate the ratio of nonsynonymous-to-synonymous substitutions (dN/dS) in the *gag*, *pol*, and *env* genes of RELIK, having first removed sequences that arose through segmental duplication of the host genome and overlapping gene regions (*rev* is almost entirely overlapped, and *tat* is too short for this analysis). This model allows the largely neutral evolution represented by the terminal branches of the phylogeny to be ignored (20). We tested for deviation from neutrality of the internal branches by fixing their dN/dS to 1 and comparing the difference in likelihood when this parameter was estimated by using a likelihood ratio test, comparing twice the difference in log likelihood to a χ^2 distribution with one degree of freedom. For the three genes, ML phylogenies were reestimated for a subsample of the RELIK sequences under the GTR+ Γ model, containing at least 50% of the coding region relative to the consensus.

We thank Robert Shafer, Austin Burt, Donald Quicke, and Paul Harvey for helpful discussion and comments on the manuscript. A.K. was funded by a Medical Research Council fellowship.

- van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, *et al.* (2000) *Seventh Report of the International Committee on Taxonomy of Viruses* (Academic, San Diego).
- Salemi M, Strimmer K, Hall WW, Duffy M, Delaporte E, Mboup S, Peeters M, Vandamme AM (2001) *FASEB J* 15:276–278.
- Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T (2000) *Science* 288:1789–1796.
- Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, Bibollet-Ruche F, Chen Y, Wain LV, Liegeois F, *et al.* (2006) *Science* 313:523–526.
- Lemey P, Pybus OG, Rambaut A, Drummond AJ, Robertson DL, Roques P, Worobey M, Vandamme AM (2004) *Genetics* 167:1059–1068.
- 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.
- Gifford R, Tristem M (2003) *Virus Genes* 26:291–315.
- Katzourakis A, Tristem M (2005) in *Retroviruses and Primate Genome Evolution*, ed Sverdlov ED (Landes Bioscience, Georgetown, TX), pp 186–203.
- Foley BT (2000) in *HIV Sequence Compendium 2000*, eds Kuiken C, Foley B, Freed E, Hahn B, Korber B, Marx PA, McCutchan F, Mellors JW, Mullins JJ, Sodroski J, *et al.* (Los Alamos Natl Lab, Los Alamos, NM), pp 35–43.
- Lower R, Lower J, Kurth R (1996) *Proc Natl Acad Sci USA* 93:5177–5184.
- Stoye JP (2006) *Genome Biol* 7:241.
- Holmes EC (2003) *J Virol* 77:3893–3897.
- Brown EW, Yuhki N, Packer C, O'Brien SJ (1994) *J Virol* 68:5953–5968.
- Sharp PM, Bailes E, Robertson DL, Gao F, Hahn BH (1999) *Biol Bull* 196:338–342.
- Sharp PM, Bailes E, Gao F, Beer BE, Hirsch VM, Hahn BH (2000) *Biochem Soc Trans* 28:275–282.
- Belshaw R, Katzourakis A (2005) *Bioinformatics* 21:122–123.
- Hughes JF, Coffin JM (2004) *Proc Natl Acad Sci USA* 101:1668–1672.
- Katzourakis A, Rambaut A, Pybus OG (2005) *Trends Microbiol* 13:463–468.
- Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, Tristem M (2004) *Proc Natl Acad Sci USA* 101:4894–4899.
- Belshaw R, Katzourakis A, Paces J, Burt A, Tristem M (2005) *Mol Biol Evol* 22:814–817.

21. Matthee CA, van Vuuren BJ, Bell D, Robinson TJ (2004) *Syst Biol* 53:433–447.
22. 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.
23. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, et al. (2002) *Nature* 420:520–562.
24. Vogt PK (1997) in *Retroviruses*, eds Coffin JM, Hughes SH, Varmus HE (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 27–69.
25. Gifford R, Kabat P, Martin J, Lynch C, Tristem M (2005) *J Virol* 79:6478–6486.
26. Boeke JD, Stoye JP (1997) in *Retroviruses*, eds Coffin JM, Hughes SH, Varmus HE (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 343–435.
27. Goff SP (2001) in *Fields Virology*, eds Knipe DM, Howley PM (Lippincott, Williams, and Wilkins, Philadelphia), pp 1871–1939.
28. Wilson DE, Reeder DM (1993) *Mammal Species of the World: A Taxonomic and Geographic Reference* (Smithsonian Institution Press, Washington, DC).
29. Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, Pierron G, Heidmann T (2006) *Genome Res* 16:1548–1556.
30. Dimmic MW, Rest JS, Mindell DP, Goldstein RA (2002) *J Mol Evol* 55:65–73.
31. Guindon S, Gascuel O (2003) *Syst Biol* 52:696–704.
32. Ronquist F, Huelsenbeck JP (2003) *Bioinformatics* 19:1572–1574.
33. Yang Z (1997) *Comput Appl Biosci* 13:555–556.
34. Pollard VW, Malim MH (1998) *Annu Rev Microbiol* 52:491–532.