An Endogenous Foamy Virus in the Aye-Aye (*Daubentonia madagascariensis*)

Guan-Zhu Han and Michael Worobey

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA

We report the discovery and analysis of an endogenous foamy virus (PSFVaye) within the genome of the aye-aye (*Daubentonia madagascariensis*), a strepsirrhine primate from Madagascar. Phylogenetic analyses indicate that PSFVaye is divergent from all known simian foamy viruses, suggesting an association between foamy viruses and primates since the haplorhine-strepsirrhine split. The discovery of PSFVaye indicates that primate foamy virus might be more broadly distributed than previously thought.

Foamy viruses are nonpathogenic complex retroviruses that infect a variety of mammalian species, including Old World and New World primates (7). Simian foamy viruses (SFVs) appear to have codiverged with Old World primates for more than 30 million years (Myr) (13). However, little is known about the origin of SFVs and foamy virus infections in primates outside the infraorder Simiiformes. Retroviruses can integrate into host genomes as endogenous elements, which provide valuable insights for studying the deep history of retroviruses (4). The recent discovery of endogenous foamy virus elements within sloth genomes suggests that foamy viruses and their mammal hosts likely share a coevolutionary history that stretches back at least to the origin of placental mammals (6). Despite the vast quantity of genomic sequences available, endogenous foamy virus-like elements have been discovered only within the genomes of sloths (6) and a coelacanth (G.-Z. Han and M. Worobey, unpublished data), suggesting that germ line invasion by foamy virus is very rare (6).

To search for additional endogenous foamy viruses, we screened all available animal whole-genome shotgun sequences from GenBank and identified several foamy virus-like insertions within the genome of the aye-aye (*Daubentonia madagascariensis*), which we designate “aye-aye prosimian foamy virus” (PSFVaye). We found only four contigs containing sequences homologous to the three main open reading frames (ORFs) of foamy virus: *gag*, *pol*, and *env* (Table 1; also, see Fig. S1 in the supplemental material). No paralogous PSFVaye copies were found within the aye-aye genome. Although we identified three sequences homologous to foamy virus Pol protein (Table 1), they have no or only a several-amino-acid residue overlap. Because the aye-aye genome sequence used in this study is a high-coverage (~38×) assembly, it is likely that there is only a single copy of PSFVaye present in the aye-aye genome. These PSFVaye elements seem to be functionally defective, given that all the elements contain numerous in-frame stop codons and frameshift mutations.

To determine the relationship between PSFVaye and other retroviruses, a maximum-likelihood (ML) phylogeny was inferred from a conserved region of the Pol protein sequences of PSFVaye and various retroviruses. The sequences for this and the additional analyses (see below) were aligned using MUSCLE (2). We used Gblocks 0.91b to eliminate the ambiguous regions and extract the conserved regions from the alignments (14). The

phylogeny was built using PHYML 3.0 (3) under an rtREV model (1). ML support was evaluated via nonparametric bootstrap analyses with 1,000 pseudoreplicates. The phylogenetic tree shows that PSFVaye groups with previously described foamy virus with robust support (see Fig. S2 in the supplemental material). Therefore, PSFVaye is indeed an endogenous foamy virus.

To further elucidate the relationship between PSFVaye and other foamy viruses, we conducted a separate phylogenetic analysis using conserved *pol* gene nucleotide sequences with a Bayesian approach. The best-fitting model of nucleotide substitution was determined using jModelTest (9), and a GTR+Γ4 substitution model was used. The Bayesian analyses were performed with MrBayes 3.1.2 (11) using 1,000,000 generations in four chains and with sampling of posterior trees every 1,000 generations. MCMC (Markov chain Monte Carlo) convergence was indicated by an effective sample size of >500, as calculated in the program Tracer v1.5. All the SFVs cluster together and form a monophyletic group with strong support (posterior probability of 1.00). PSFVaye is divergent from all known simian foamy viruses, which makes it highly unlikely that PSFVaye originated from a simian foamy virus via recent host switching. However, the precise phylogenetic position of PSFVaye was unresolved, likely because exogenous foamy viruses have evolved under strong functional constraints but PSFVaye might have experienced extensive within-host neutral evolution after endogenization. Nevertheless, the phylogenetic position of PSFVaye (Fig. 1) is compatible with the ancestral codivergence of foamy viruses and their primate hosts, potentially dating back to approximately 85 Myr ago (to the haplorhine-strepsirrhine split). In this scenario, foamy virus would have arrived in Madagascar along with the ancient African strepsirrhines around 50 to 80 Myr ago (8, 16, 17). Although

Received 15 March 2012 Accepted 30 April 2012

Published ahead of print 9 May 2012

Address correspondence to Guan-Zhu Han, guan-zhu@email.arizona.edu, or Michael Worobey, worobey@email.arizona.edu.

Supplemental material for this article may be found at <http://jvi.asm.org/>.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.00650-12

TABLE 1 Aye-aye genome contigs that match foamy virus gene sequences

Contig	Position (nucleotides)	Genomic region represented	Best hit	% identity	BLASTP E-value
contig_1842496	1154–249	<i>gag</i>	CAD92795 (feline foamy virus)	41	3e-63
contig_2073143	335–591	<i>pol</i>	ABV59399 (spider monkey foamy virus)	42	4e-10
	726–1259	<i>pol</i>	ADE06000 (common marmoset foamy virus)	64	7e-69
contig_1522645	75–1990	<i>pol</i>	NP_054716 (equine foamy virus)	44	2e-158
contig_362513	657–2009	<i>env</i>	ABM55472 (bovine foamy virus)	37	2e-85

foamy virus has been detected in galagos, a strepsirrhine species native to mainland Africa, no sequence of galago foamy virus has been determined (5). Although it is still possible that galagos acquired foamy virus by direct transmission from simian species, determining the sequence of galago foamy virus might shed additional light on the diversity and deep history of primate foamy viruses.

Lacking both PSFVaye paralogous copies or long terminal repeat (LTR) sequences, we were not able to estimate the integration time based on sequence divergence (though the multiple stop codons suggest a considerable time span). Notably, we did not identify any foamy virus-like elements in the genome of *Microcebus murinus*, another lemur species, suggesting that the endogenization event of PSFVaye might not predate the common ancestor of lemur species.

The Indo-Madagascar continent split from continental Africa

around 165 Myr ago and has remained in its present position since approximately 121 Myr ago (10, 15). The Indian subcontinent and Madagascar subsequently separated nearly 88 Myr ago (12). All living lemurs originated from a single common ancestor that colonized Madagascar from mainland Africa around 50 to 80 Myr ago (8, 16, 17). The aye-aye is phylogenetically basal to all extant lemurs (8, 16, 17). If foamy virus indeed arrived in Madagascar along with ancient African strepsirrhines as discussed above, foamy virus may have codiverged with lemur species there. It is thus possible that there are undetected foamy viruses still circulating in other lemurs in Madagascar.

In this study, we established the presence of a distinct foamy viral lineage in a basal primate in Madagascar. It appears that the foamy viruses are more widely distributed than has previously been thought, both phylogenetically and geographically. The discovery of PSFVaye provides important insights into the origin of

FIG 1 Bayesian phylogenetic analysis of PSFVaye and other endogenous and exogenous foamy viruses. The phylogenies are 50% majority-rule consensus trees generated via MrBayes 3.1.2. The node labels are posterior probabilities, which indicate the proportion of trees in the posterior distribution containing the node. Branch lengths are in expected changes per site. Host species are as follows: SFVgor, gorilla; SFVcpz, chimpanzee; SFVmac, macaque; SFVagm, African green monkey; and SFVspm, spider monkey. SloEFV, sloth endogenous foamy virus; EFV, equine foamy virus; BFV, bovine foamy virus; FFV, feline foamy virus. Aye-aye photo courtesy of Tom Junek.

SFVs and fills in a gap between simian viruses and other mammalian foamy viruses.

ACKNOWLEDGMENTS

This work was supported by NIAID grant R01 AI084691 and the David and Lucile Packard Foundation.

REFERENCES

1. Dimmic MW, Rest JS, Mindell DP, Goldstein D. 2002. rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *J. Mol. Evol.* 55:65–73.
2. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic. Acids Res.* 32:1792–1797.
3. Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307–321.
4. Johnson WE, Coffin JM. 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci. U. S. A.* 96:10254–10260.
5. Johnston PB. 1971. Taxonomic features of seven serotypes of simian and ape foamy viruses. *Infect. Immun.* 3:793–799.
6. Katzourakis A, Gifford RJ, Tristem M, Gilbert MT, Pybus OG. 2009. Macroevolution of complex retroviruses. *Science* 325:1512.
7. Meiering CD, Linial ML. 2001. Historical perspective of foamy virus epidemiology and infection. *Clin. Microbiol. Rev.* 14:165–176.
8. Perelman P, et al. 2011. A molecular phylogeny of living primates. *PLoS Genet.* 7:e1001342. doi:10.1371/journal.pgen.1001342.
9. Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25:1253–1256.
10. Rabinowitz PD, Coffin MF, Falvey D. 1983. The separation of Madagascar and Africa. *Science* 220:67–69.
11. Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
12. Storey M, et al. 1995. Timing of hot-spot related volcanism and the breakup of Madagascar and India. *Science* 267:852–855.
13. Switzer WM, et al. 2005. Ancient co-speciation of simian foamy viruses and primates. *Nature* 434:376–380.
14. Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56:564–577.
15. Yoder AD, et al. 2003. Single origin of Malagasy Carnivora from an African ancestor. *Nature* 421:734–737.
16. Yoder AD, Cartmill M, Ruvolo M, Smith K, Vilgalys R. 1996. Ancient single origin for Malagasy primates. *Proc. Natl. Acad. Sci. U. S. A.* 93:5122–5126.
17. Yoder AD, Yang Z. 2004. Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol. Ecol.* 13:757–773.