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

Guan-Zhu Han* & Michael Worobey*

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA

* Corresponding authors

Guan-Zhu Han
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA
Tel.: (520)-621-4881; Fax: (520)-621-9190; E-mail: guanzhu@email.arizona.edu

Michael Worobey
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA
Tel.: (520) 626-3456; Fax: (520)-621-9190; E-mail: worobey@email.arizona.edu

Running title: Aye-aye endogenous foamy virus

Abstract word count: 67

Text word count: 946

Figures: 1
Abstract

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 haplorrhine-strepsirrhine split. The discovery of PSFVaye indicates that primate foamy virus might be more broadly distributed than previously thought.
Foamy viruses are non-pathogenic 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foamy virus-like elements have been discovered only within the genomes of sloths (6) and a coelacanth (Han and Worobey, unpublished data), suggesting germline invasion of foamy virus is very rare (6).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, respectively (Table 1; Fig. S1). No paralogous PSFVaye copies were found within the aye-aye genome. Although we identified three sequences homologous to foamy virus Pol protein (Table 1), there is no or only several amino acid
residue overlap with each other. Because the aye-aye genome sequence used in this study is a high-coverage (~38×) assembly, it is unlikely that there is more than 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 frame-shift mutations.

To determine the relationship between PFVaye 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 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 1000 pseudoreplicates. The phylogenetic tree shows that PSFVaye groups with previously described foamy virus with robust support (Fig. S2). 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, sampling posterior trees every 1000 generations. MCMC convergence was indicated by an effective sample size >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). The 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 co-divergence of foamy viruses and their primate hosts, potentially dating back to approximately 85 Myr ago (to the haplorrhine-strepsirrhine split). Under this scenario, foamy virus would have arrived in Madagascar along with the ancient African strepsirrhines around 50-80 Myr ago (8, 16, 17). Although foamy virus has been detected in galagos, strepsirrhine species native to mainland Africa, no sequence of galago foamy virus has been elucidated (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 either 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 consequently separated nearly 88 Myr ago (12). All living
lemurs originated from a single common ancestor that colonized Madagascar from
mainland Africa around 50-80 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 have 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 SFVs and fills in a missing link
between simian viruses and other mammal foamy viruses.

Acknowledgments

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

References


**Figure legend**

**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. For the following simian foamy viruses (SFVs), host species is shown in parentheses: 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.

**Table 1. The matching contigs identified in aye-aye genome**

<table>
<thead>
<tr>
<th>Contig Number</th>
<th>Position</th>
<th>Genomic Regions Represent</th>
<th>Best Hit</th>
<th>Identity</th>
<th>BLASTP E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>contig_1842496</td>
<td>1154-249</td>
<td>gag</td>
<td>CAD92795 (Feline foamy virus)</td>
<td>41%</td>
<td>3e-63</td>
</tr>
<tr>
<td>contig_2073143</td>
<td>335-591</td>
<td>pol</td>
<td>ABV59399 (Spider monkey foamy virus)</td>
<td>42%</td>
<td>4e-10</td>
</tr>
<tr>
<td></td>
<td>726-1259</td>
<td>pol</td>
<td>ADE06000 (Common marmoset foamy virus)</td>
<td>64%</td>
<td>7e-69</td>
</tr>
<tr>
<td>contig_1522645</td>
<td>75-1990</td>
<td>pol</td>
<td>NP_054716 (Equine foamy virus)</td>
<td>44%</td>
<td>2e-158</td>
</tr>
<tr>
<td>contig_362513</td>
<td>657-2009</td>
<td>env</td>
<td>ABM55472 (Bovine foamy virus)</td>
<td>37%</td>
<td>2e-85</td>
</tr>
</tbody>
</table>