JVI Journals.ASM.org

Identification of Recombination in the Envelope Gene of Simian Foamy Virus Serotype 2 Isolated from *Macaca cyclopis*

Teresa A. Galvin, Imran A. Ahmed,* Muhammad Shahabuddin,* Theodore Bryan,* Arifa S. Khan

Laboratory of Retroviruses, Division of Viral Products, Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, Bethesda, Maryland, USA

The full-length sequence of simian foamy virus serotype 2 (SFVmcy-2), isolated from a Taiwanese macaque, was determined. SFVmcy-2 was highly related to SFV serotype 1 (SFVmcy-1), an isolate from the same species, except in the putative receptor binding domain (RBD) in *env*, which contained novel sequences related to SFV serotype 3 (SFVagm-3), isolated from an African green monkey. The results identify a potential region of neutralization in SFVs and demonstrate recombination between genetically divergent foamy viruses.

Simian foamy viruses (SFVs) belong in the *Spumavirus* genus of the *Spumaretrovirinae* subfamily of *Retroviridae* and are widespread in all nonhuman primates (NHPs) (1, 2). SFVs have been isolated from various tissues of different NHPs and were originally designated based upon neutralization serotyping (3). Original foamy virus isolates were designated SFV serotype 1, SFV serotype 2, and SFV serotype 3 (SFV-1, SFV-2, and SFV-3, respectively). These original monkey isolates were renamed to indicate the species of isolation (4): SFV-1, which was isolated from a Taiwanese macaque (Formosan Rock macaque or *Macaca cyclopis* [mcy]) (5, 6), was designated SFVmac, and SFV-3, which was isolated from an African green monkey (AGM; *Chlorocebus aethiops*) (7), was designated SFVagm. To distinguish SFVmac/SFV-1 from SFV-2, which was also isolated from *M. cyclopis* (5), in this paper we have designated these SFVmac viruses SFVmcy-1 and SFVmcy-2, respectively, and the SFVagm virus SFVagm-3 to distinguish it from other AGM isolates. Early results of virus isolation or antibody detection based upon neutralization serotyping showed that, although macaques generally harbor SFVs of serotype 1, viruses with serotype 2 were also present in some cases. In fact, SFVs of both serotype 1 and serotype 2 were isolated from different organs within the same monkey as well as from a single organ (8). AGMs were found to be typically infected with SFVs of serotype 2 and serotype 3, but SFV of serotype 1 was also seen (7–9). More recent molecular studies have also shown that although SFVs generally circulate within a host species, interspecies transmission can occur (10, 11).

The biological properties of SFVmcy-1 and SFVagm-3 have been well studied, and whole-genome sequences for both viruses have been determined (6, 12, 13). Our previous studies using a variety of cell lines from different species showed that SFVmcy-2 has a broad host range but different kinetics of replication in different cell lines (14). In this study, we have determined the complete nucleotide sequence of SFVmcy-2 and have determined its genetic relatedness to SFVmcy-1 and SFVagm-3 as well as to other SFVs.

The genomic sequence and structure of SFVmcy-2 was determined using virus acquired from the American Type Culture Collection (Manassas, VA) (SFV type 2, catalogue number VR 277, FV-34, lot 4D, 91-10). Full-length SFVmcy-2 sequences were obtained by assembly of cloned DNAs obtained by restriction enzyme digestion (DNA 447) or PCR amplication (DNA 448 and DNA 549) of high-molecular-weight DNA prepared from virus-infected *Mus dunni* cells (14) (Fig. 1). Nucleotide sequence comparison of SFVmcy-2 sequences with the full-length SFVmcy-1 sequence (GenBank accession number NC_010819) indicated two deletions in DNA 448; the first corresponded to the spliced-out internal promoter leader intron, and the second corresponded to the *bet* intron (15, 16). Open reading frames were identified using PlotOrf (http://emboss.bioinformatics.nl/cgi-bin/emboss/plotorf), and their order was assigned based upon similarity with SFVmcy-1. Open reading frames were found for Gag, Pol, Env, and Tas; however, the second exon (*orf-2*) of Bet contained a potential stop codon in DNA 448 and DNA 549 DNAs (indicated by an asterisk in Fig. 1). This stop codon was also found in 2 other cloned DNAs, which were obtained from virus-infected *M. dunni* cell DNA and from total nucleic acid prepared from the original ATCC virus stock (data not shown).

Comparative sequence analysis indicated that SFVmcy-2 and SFVmcy-1 had high (93% to 98%) nucleotide and amino acid identity in the long terminal repeat (LTR), *gag*, *pol*, *tas*, and *bet* (see details in Table 1). Although SFVs are known to be more divergent in *gag* than in *env*, SFVmcy-2 and SFVmcy-1 showed greater similarity in *gag* than in *env* (94.2% versus 85.9%, respectively). Interestingly, there was more similarity in *env* between SFVmcy-2 and SFVagm-3 (76.2% nucleotide and 78.7% amino acid identity) than between SFVmcy-1 and SFVagm-3 (73.2% nucleotide and 73% amino acid identity) (Table 1). Further analysis in the *env* region showed that the majority of the nucleotide differences between SFVmcy-2 and SFVmcy-1 were in the surface glycoprotein (SU) portion, representing 79.6% of the total nucleotide changes in *env*. Based upon percent identity of the amino acids, SFVmcy-2 was more similar to SFVagm-3 (76.8%) in SU than to SFVmcy-1 (71.8%), whereas SFVmcy-2 and SFVmcy-1

Received 27 December 2012 **Accepted** 14 May 2013

Published ahead of print 22 May 2013

Address correspondence to Arifa S. Khan, arifa.khan@fda.hhs.gov.

* Present address: Imran A. Ahmed, University of Maryland, College Park, Maryland, USA; Muhammad Shahabuddin, Office of Compliance and Biologics Quality, DBSQC, CBER, FDA, Kensington, Maryland, USA; Theodore Bryan, 11104 Coppermine Rd., Woodsboro, Maryland, USA.

doi:10.1128/JVI.03555-12

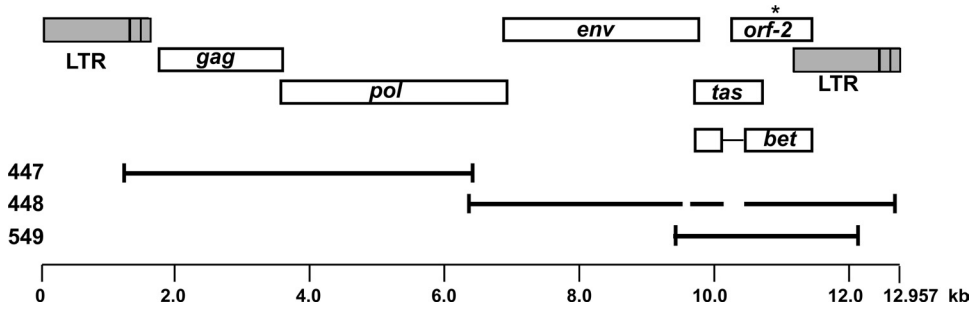

FIG 1 Genomic structure of SFVmcy-2. Predicted open reading frames (ORFs) are indicated along with the spliced Bet ORF. The position of the predicted in-frame stop codon in Bet is indicated with an asterisk. Breaks in the line of the cloned DNAs indicate deletions relative to the predicted full-length SFVmcy-2 sequence. Lengths are shown in kilobase pairs (kb).

were highly related in the LP (98.4%) and TM (97.3%) regions of *env* (Table 1). A graphical representation of the differences in *env* between SFVmcy-1 and SFVmcy-2 was generated by aligning sequences using ClustalW followed by analysis using Highlighter (www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter.html) (Fig. 2A): the vast majority of the differences between the two isolates were in the region that encodes the surface glycoprotein (SU) of the Env protein, which contains the putative receptor binding domain (RBD).

To further analyze the relatedness of SFVmcy-2 *env* to other FVs, a BLASTN search of the GenBank nr/nt database was done. SFVmcy-2 was found to have higher nucleotide and amino acid identity in the SU region to SFV-R289HybAGM, an isolate from a rhesus macaque (*Macaca mulatta* [mmu]), than to SFVmcy-1 but was more similar to SFVmcy-1 in the LP and TM regions (Table 1). Additionally, sequence analysis of other regions indicated that SFVmcy-2 was more closely related to SFVmcy-1 in LTR, *gag*, *pol*, *tas*, and *bet* (93% to 98% nucleotide and amino acid identity) than to SFV-R289HybAGM (78% to 95% nucleotide and amino acid identity).

To investigate the possibility of recombination in SFVmcy-2, phylogenetic trees were generated using MEGA5.10 (Molecular Evolutionary Genetics Analysis; www.megasoftware.net [17]). Nucleotide sequences (listed in Table 2) were aligned in MEGA5 using ClustalW, and the evolutionary history was inferred using the maximum-likelihood method based on the general time-reversible model (18), chosen because it had the lowest Bayesian information criteria score in a model test performed in MEGA5. The bootstrap consensus tree inferred from 1,000 replicates was taken to represent the evolutionary history of the taxa analyzed (19). The initial tree(s) for the heuristic search was obtained automatically by applying neighbor-joining and BioNJ algorithms to a matrix of pairwise distances estimated using the maximum-composite-likelihood (MCL) approach and then selecting the topology with the superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites. All positions containing gaps and missing data were eliminated. Highly congruent trees were generated using the neighbor-joining method and the Kimura-2 parameter substitution model (data not shown). Similar results were seen with trees constructed with LTR, *gag*, *pol*, and *tas* nucleotide sequences, showing that SFVmcy-1 and SFVmcy-2 branched together and clustered with SFV-R289HybAGM and SFVagm-3 (data not shown) whereas, in the *env* tree, SFVmcy-2 and SFV-R289HybAGM were on the same branch (Fig. 2B). This analysis also included *env* sequences of two naturally occurring SFVs isolated from rhesus macaques (designated SFVmmu-A4W and

TABLE 1 Sequence analysis of full-length SFV genomes from different monkey species

Genomes compared	Sequence	% identity[a]								
		LTR	Gag	Pol	Tas	Bet	Env	LP	SU	TM
SFVmcy-2 vs SFVmcy-1	Nucleotide	94.9	94.2	95.4	95.3	93.2	85.9	94.4	74.6	94.6
	Amino acid		96.8	98.1	96.4	93.6	86.3	98.4	72.1	97.4
SFVmcy-2 vs SFVagm-3	Nucleotide	64.6	68.3	81.3	64.5	62.0	76.2	73.3	75.8	77.4
	Amino acid		64.8	85.4	54.2	53.3	78.7	72.2	77.0	82.2
SFVmcy-1 vs SFVagm-3	Nucleotide	64.7	67.8	81.4	65.2	61.6	73.2	73.5	68.3	77.9
	Amino acid		63.8	85.7	54.2	53.7	73.0	72.2	64.8	83.2
SFVmcy-2 vs SFV-R289	Nucleotide	86.5	85.9	89.3	84.4	78.8	86.5	86.0	83.8	89.4
	Amino acid		88.9	95.3	81.8	78.2	94.2	95.2	92.7	95.0
SFVmcy-1 vs SFV-R289	Nucleotide	87.6	84.4	89.7	84.5	79.6	81.8	85.4	73.1	89.9
	Amino acid		87.6	95.3	82.1	78.9	83.9	93.7	72.1	94.5
SFVagm-3 vs SFV-R289	Nucleotide	64.7	68.1	81.3	66.0	62.3	76.3	74.9	74.6	78.4
	Amino acid		65.7	85.8	54.2	53.1	79.3	74.6	77.2	83.2

[a] Numbers are percent identity as determined using the ClustalW alignment option in MegAlign (Lasergene; DNASTAR, Madison, WI).

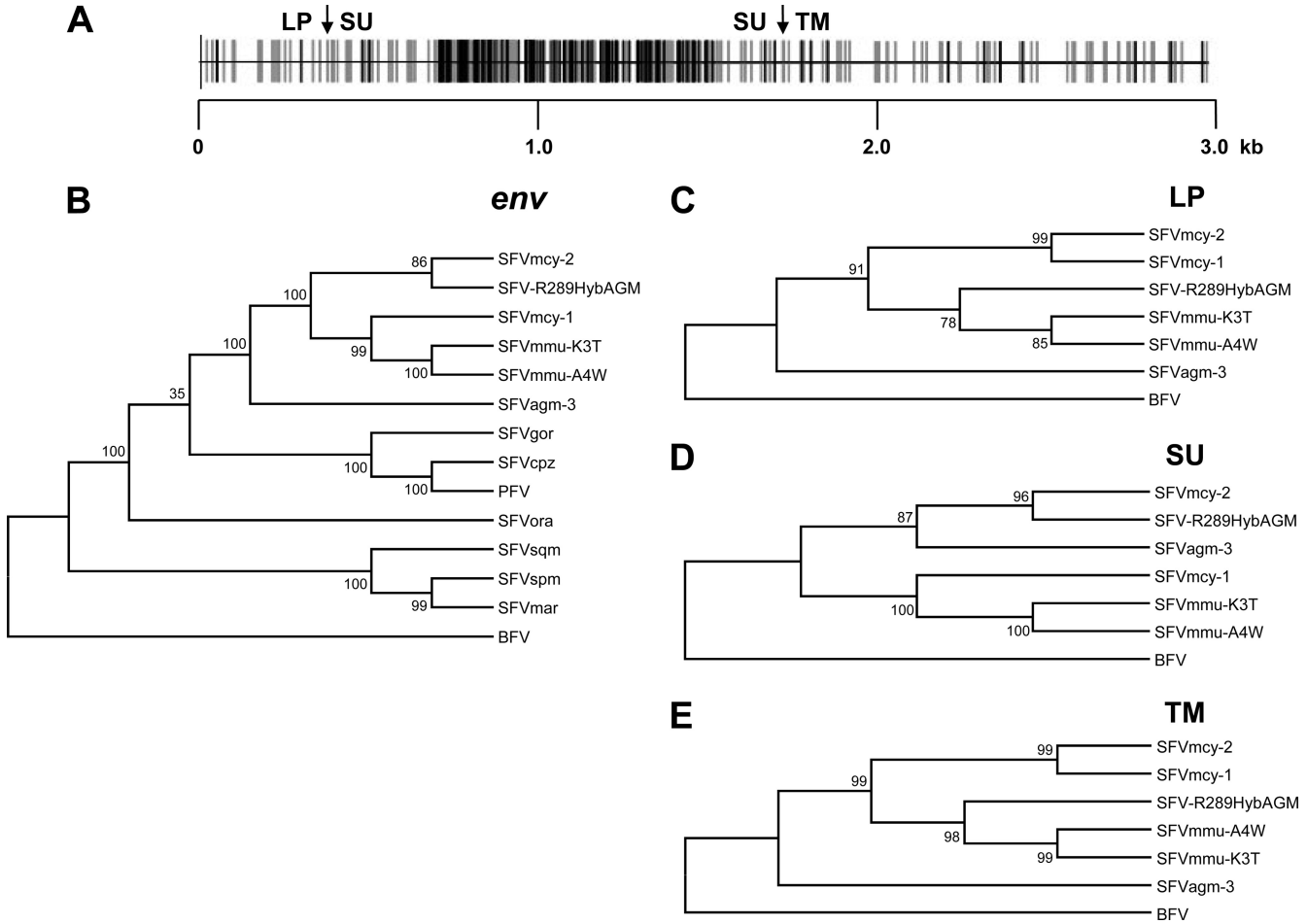

FIG 2 Alignment and phylogenetic analysis of SFV *env* sequences. (A) Schematic representation of aligned nucleotide sequences of SFVmcy-1 and SFVmcy-2. Gray vertical lines indicate silent mutations, and black vertical lines indicate nonsilent mutations. LP, SU, and TM junctions are shown by arrows. Length is shown in kb. (B to E) Results of phylogenetic analysis of full-length *env*, using a total of 2,855 positions in the final data set (B), LP, using a total of 369 positions in the final data set (C), SU, using a total a total of 1,303 positions in the final data set (D), and TM, using a total of 1,244 positions in the final data set (E), are shown. BFV was used as an outgroup for all of the trees. The accession numbers of the sequences used for the comparisons are listed in Table 2.

SFVmmu-K3T) and cloned from virus-infected *M. dunni* cell DNA. The results showed that SFVmmu branched with SFVmcy-1. Further phylogenetic analysis in the LP, SU, and TM regions of SFVmcy-2 indicated that SFVmcy-2 and SFVmcy-1 branched together based upon LP and TM (Fig. 2C and E, respectively) but SFVmcy-2 branched with SFV-R289HybAGM and SFVagm-3 based upon SU (Fig. 2C). It was noted that in LP and TM, SFV-R289HybAGM clustered with the SFVmmu viruses.

To identify the regions of recombination in *env*, the nucleotide sequences encompassing the complete SU region plus 200 bp of adjacent upstream and downstream sequence of different SFVs were subjected to recombination analysis using similarity plot analysis (Simplot) and BootScan analysis in SimPlot 3.5.1 (20) (http://sray.med.som.jhmi.edu/SCRoftware/simplot) with default parameters except that BootScan repetitions were set to 1,000 and the Kimura 2-parameter option was used (Fig. 3A and B). SFVmcy-2 was used as the reference sequence in these analyses, and BFV was used as an outgroup. SFV sequences used for the analysis are listed in Table 2 and were aligned using ClustalW in MEGA. The Simplot showed high similarity to SFVagm-3 in the central portion and to SFVmcy-1 in the adjacent sequences (Fig. 3A). Bootscan analysis found evidence of recombination between SFVmcy-1 (green line) and an SFVagm-3-related virus (red line) and defined the possible recombination breakpoints as being between nucleotides 7700 and 7746 for the 5′ breakpoint and between nucleotides 8593 and 8659 for the 3′ breakpoint (SFVmcy-2 accession number KF026286). A similar analysis was done using SFV-R289HybAGM as the reference sequence. The results indicated that SFV-R289HybAGM was generated by recombination between an SFVmmu-related virus and an SFVagm-3-related virus (blue line and red line, respectively; Fig. 3C). Interestingly, the recombination breakpoint regions were similar to those seen with SFVmcy-2: the 5′ breakpoint was between nucleotides 7681 and 7733, and the 3′ breakpoint was between nucleotides 8606 and 8655 (SFV-R289HybAGM; accession number JN801175) (Fig. 3C). Alignments of the amino acids in the SU regions of the SFVmcy-1, SFVmcy-2, SFV-R289HybAGM, SFVmmu-K3T, and SFVagm-3 with the translated predicted recombination breakpoints are shown in Fig. 4. The possibility that recombination in SFVmcy-2 occurred during passage in our laboratory was excluded by determining the full-length *env* sequence amplified directly from nucleic acid prepared from an aliquot of

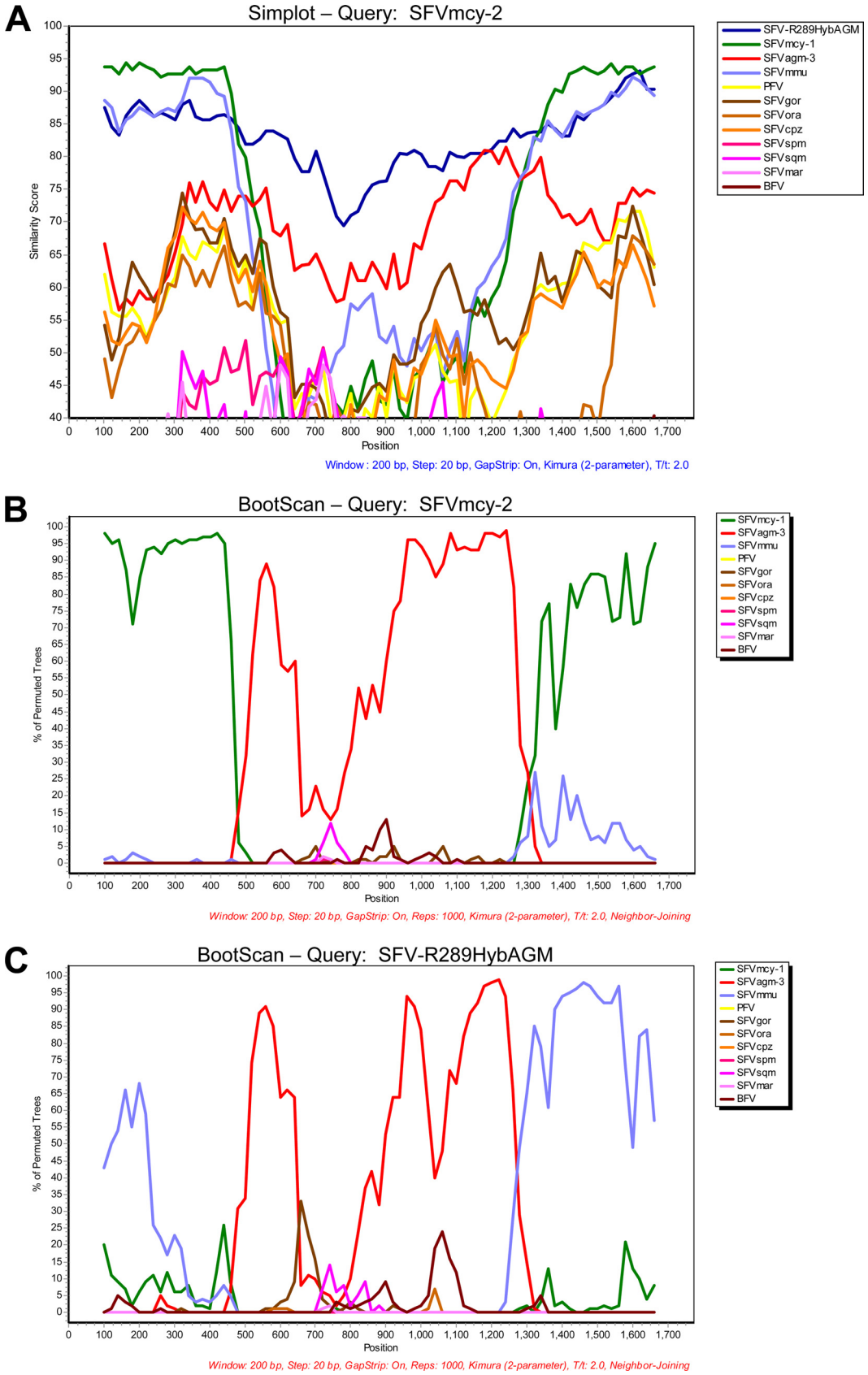

FIG 3 Recombination analysis of SFVmcy-2 SU region. (A and B) Simplot analysis (A) and BootScan analysis (B) comparing SFVmcy-2 SU and other SFV isolates. SFVmcy-2 SU was used as the reference (query) sequence in both analyses. (C) BootScan analysis comparing SFV-R289HybAGM SU and other SFV isolates. SFV-R289HybAGM SU was used as the reference (query) sequence.

TABLE 2 GenBank accession numbers for SFV isolates used in the study

Isolate	GenBank accession no.
SFVmcy-2	KF026286
SFVmcy-1	NC_010819
SFVagm-3	NC_010820
PFV (primate foamy virus)	NC_001736
SFVcpz (chimpanzee)	NC_001364
SFVgor (gorilla)	JQ867465
SFVora (orangutan)	AJ544579
SFVspm (spider monkey)	EU010385
SFVsqm (squirrel monkey)	GU356394
SFVmar (marmoset)	GU356395
SFV-R289HybAGM	JN801175
SFVmmu-K3T	KF026287
SFVmmu-A4W	KF026288
BFV (bovine foamy virus)	NC_001831

the original ATCC virus stock, which was found to have >99% identity with the SFVmcy-2 *env* described in this paper (there were 12 nucleotide differences in a total of 2,949 bases).

Coinfection of NHPs with SFVs that are genetically diverse is known. PCR analysis of SFV RNAs from fecal samples found recombination in the *pol* region due to superinfection of chimpanzees by circulating SFV strains (11). Our study indicates that SFVmcy-2 was generated by recombination involving an SFVmcy-1-like virus and a novel virus, serotypically distinct but genetically related to SFVagm-3. Since all of the sequences in SFVmcy-2 and SFVmcy-1 were highly similar in the entire genome except in SU, this region most likely contains the epitopes for the serotypic differences between the viruses. Since SFVmcy-2 and SFVmcy-1 were isolated from the same monkey species (Taiwanese macaque), it seems that parent viruses with two different serotypes may be cocirculating in the species and were involved in recombination due to coinfection. Evidence of recombination between genetically distinct viruses was also seen in rhesus macaques based upon an analysis of SFV-R289HybAGM that showed that a similar recombination event had occurred between an SFVmmu-related virus of rhesus macaques and an SFVagm-3-related virus. These results suggest the cocirculation of genetically distinct SFVs in different macaque species that can recombine to generate novel viruses. It is interesting to note that the greatest sequence diversity between the feline foamy virus (FFV) F17/951-like serotype and the FUV7-like virus serotype resides in a similar region of SU, which is also thought to be evidence of recombination (21, 22). Since the recombination regions in SFVmcy-2 and SFV-

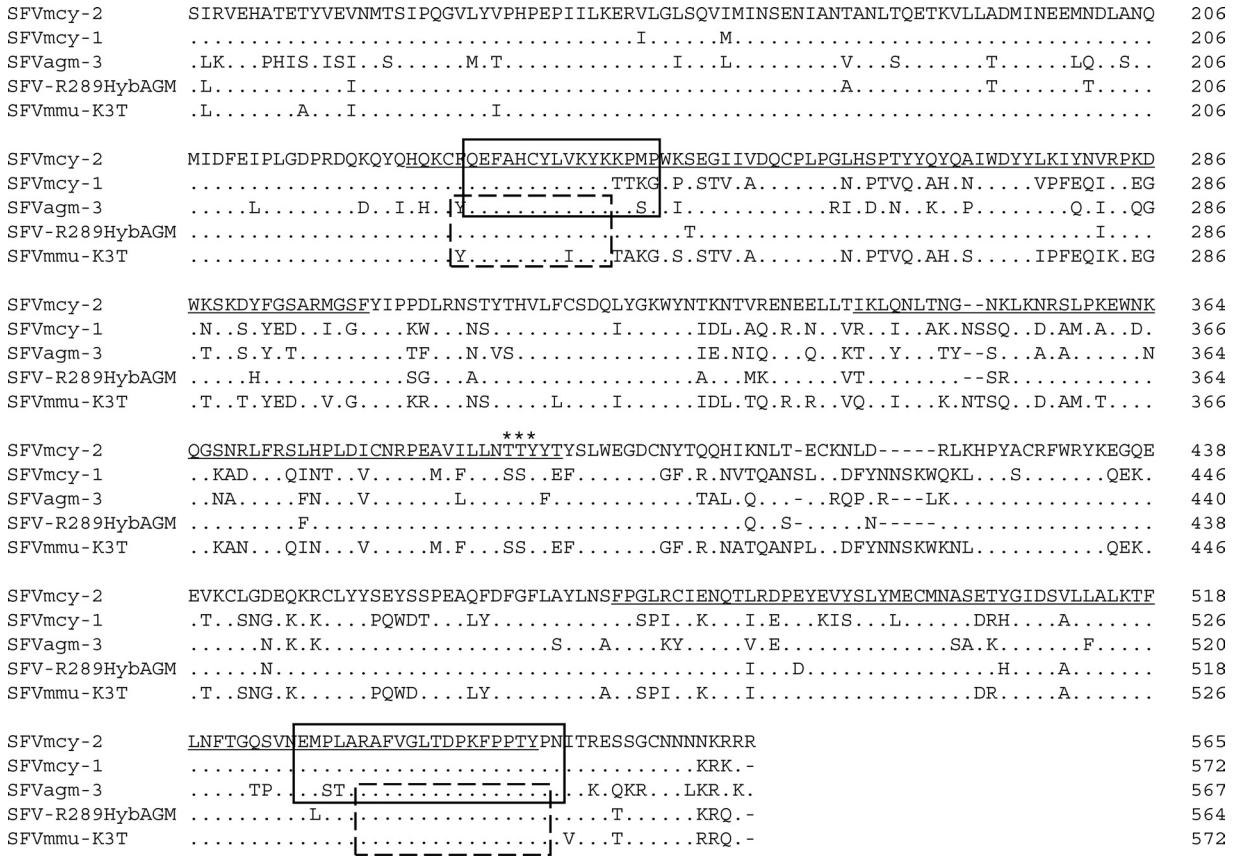

FIG 4 Alignment of SFV SU regions. Amino acid sequences of SFVmcy-2, SFVmcy-1, SFVagm-3, SFV-R289HybAGM, and SFVmmu-K3T were aligned in MegAlign using ClustalW with the BLOSUM protein weight matrix. Dots indicate identity, and dashes indicate gaps. The glycosylation site found to be required for folding in the prototype foamy virus (PFV; originally designated human foamy virus or HFV) and preserved in all SFVs is labeled with asterisks. Residues predicted to be included in the three essential domains of the bipartite receptor binding domain are underlined. The amino acids that correspond to the nucleotides predicted for the recombination breakpoint sequences by SimPlot 3.5.1 are boxed as follows: solid line, SFVmcy-2; broken line, SFV-R289HybAGM.

R289HybAGM, which independently arose in two different monkey species, were similar and corresponded to a putative recombination region in FFV, it seems that foamy viruses may have a "hot spot" for recombination in SU.

Nucleotide sequence accession numbers. The GenBank accession numbers for the sequences reported in this paper are as follows: for SFVmcy-2, KF026286; for SFVmmu-K3T, KF026287; and for SFVmmu-A4W, KF026288.

REFERENCES

1. **Khan AS.** 2009. Simian foamy virus infection in humans: prevalence and management. Expert Rev. Anti Infect. Ther. **7:**569–580.
2. **Rethwilm A.** 2010. Molecular biology of foamy viruses. Med. Microbiol. Immunol. **199:**197–207.
3. **Hooks JJ, Gibbs CJ, Jr.** 1975. The foamy viruses. Bacteriol. Rev. **39:**169–185.
4. **Meiering CD, Linial ML.** 2001. Historical perspective of foamy virus epidemiology and infection. Clin. Microbiol. Rev. **14:**165–176.
5. **Johnston PB.** 1961. A second immunologic type of simian foamy virus: monkey throat infections and unmasking by both types. J. Infect. Dis. **109:**1–9.
6. **Johnston PB.** 2000. Strain FV-21 of simian foamy virus type 1 was cloned and sequenced after isolation from the Taiwan monkey Macaca cyclopsis. J. Microbiol. Immunol. Infect. **33:**60–61.
7. **Stiles GE, Bittle JL, Cabasso VJ.** 1964. Comparison of simian foamy virus strains including a new serological type. Nature **201:**1350–1351.
8. **O'Brien TC, Albrecht P, Hannah JE, Tauraso NM, Robbins B, Trimmer RW.** 1972. Foamy virus serotypes 1 and 2 in rhesus monkey tissues. Arch. Gesamte Virusforsch. **38:**216–224.
9. **Stiles GE.** 1968. Serologic screening of rhesus and grivet monkeys for SV40 and the foamy viruses. Proc. Soc. Exp. Biol. Med. **127:**225–230.
10. **Leendertz FH, Zirkel F, Couacy-Hymann E, Ellerbrok H, Morozov VA, Pauli G, Hedemann C, Formenty P, Jensen SA, Boesch C, Junglen S.** 2008. Interspecies transmission of simian foamy virus in a natural predator-prey system. J. Virol. **82:**7741–7744.
11. **Liu W, Worobey M, Li Y, Keele BF, Bibollet-Ruche F, Guo Y, Goepfert PA, Santiago ML, Ndjango JB, Neel C, Clifford SL, Sanz C, Kamenya S, Wilson ML, Pusey AE, Gross-Camp N, Boesch C, Smith V, Zamma K, Huffman MA, Mitani JC, Watts DP, Peeters M, Shaw GM, Switzer WM, Sharp PM, Hahn BH.** 2008. Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. PLoS Pathog. **4:**e1000097. doi:10.1371/journal.ppat.1000097.
12. **Kupiec JJ, Kay A, Hayat M, Ravier R, Peries J, Galibert F.** 1991. Sequence analysis of the simian foamy virus type 1 genome. Gene **101:**185–194.
13. **Renne R, Friedl E, Schweizer M, Fleps U, Turek R, Neumann-Haefelin D.** 1992. Genomic organization and expression of simian foamy virus type 3 (SFV-3). Virology **186:**597–608.
14. **Khan AS, Sears JF, Muller J, Galvin TA, Shahabuddin M.** 1999. Sensitive assays for isolation and detection of simian foamy retroviruses. J. Clin. Microbiol. **37:**2678–2686.
15. **Löchelt M, Flugel RM, Aboud M.** 1994. The human foamy virus internal promoter directs the expression of the functional Bel 1 transactivator and Bet protein early after infection. J. Virol. **68:**638–645.
16. **Muranyi W, Flugel RM.** 1991. Analysis of splicing patterns of human spumaretrovirus by polymerase chain reaction reveals complex RNA structures. J. Virol. **65:**727–735.
17. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:**2731–2739.
18. **Nei M, Kumar S.** 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, NY.
19. **Felsenstein J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
20. **Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC.** 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. **73:**152–160.
21. **Winkler IG, Flugel RM, Lochelt M, Flower RL.** 1998. Detection and molecular characterisation of feline foamy virus serotypes in naturally infected cats. Virology **247:**144–151.
22. **Zemba M, Alke A, Bodem J, Winkler IG, Flower RL, Pfrepper K, Delius H, Flugel RM, Lochelt M.** 2000. Construction of infectious feline foamy virus genomes: cat antisera do not cross-neutralize feline foamy virus chimera with serotype-specific Env sequences. Virology **266:**150–156.