Mosaic genome structure of simian immunodeficiency virus from West African green monkeys

Mojun J.Jin, Huxiong Hui,
David L.Robertson[1], Michaela C.Müller[2],
Françoise Barré-Sinoussi[2], Vanessa M.Hirsch[3],
Jonathan S.Allan[4], George M.Shaw,
Paul M.Sharp[1] and Beatrice H.Hahn[5]

Departments of Medicine and Microbiology, LHRB 613, 701 South 19th Street, University of Alabama at Birmingham, Birmingham, AL 35294, [3]Immunodeficiency Viruses Section, Laboratory of Infectious Diseases, NIAID/NIH, Twinbrook II, 12441 Parklawn Drive, Rockville, MD 20852, [4]Departments of Virology and Immunology, Southwest Foundation for Biomedical Research, 7620 Southwest Loop 410, San Antonio, TX 78228, USA, [1]Department of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK and [2]Unite de Biologie des Retrovirus, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris, France

[5]Corresponding author

Communicated by P.M.Sharp

Elucidation of the phylogenetic origins of simian and human immunodeficiency viruses (SIV and HIV) is fundamental to the understanding of HIV pathogenesis and the spread of AIDS worldwide. In this study, we molecularly characterized multiple SIV$_{AGM}$ isolates from four different African green monkey species (vervet, grivet, sabaeus and tantalus monkeys). Phylogenetic analysis of partial (1 kb) env sequences indicated that all SIV$_{AGM}$ strains cluster together, and that they fall into four distinct sequence sub-groups according to their species of origin. However, alignment of long terminal repeat sequences revealed that SIVs from West African sabaeus monkeys contain a structural feature (a duplication of the transactivation response element) thus far only found in otherwise highly divergent lentiviruses infecting sooty mangabeys (SIV$_{SM}$) and humans (HIV-2). To determine whether there were additional similarities with the SIV$_{SM}$/HIV-2 group, a full-length replication competent sabaeus provirus was cloned and sequenced. In phylogenetic trees derived from the central and 3' coding regions, the sabaeus virus clustered with SIV$_{AGM}$ isolates from other African green monkey species. However, in trees derived from the 3' half of gag and the adjacent 5' region of pol, the sabaeus virus grouped with the SIV$_{SM}$/HIV-2 lineage. These results indicated that the sabaeus virus comprised a mosaic genome which must have resulted from recombination of divergent lentiviruses in the distant past. A second, independent sabaeus isolate exhibited similar phylogenetic relationships, suggesting that all West African green monkey viruses share this complex evolutionary history. Taken together, these results indicate that African green monkeys have been infected with SIV$_{AGM}$ for very long periods of time, and that recombination and cross-species transmission in the wild have contributed to the genetic complexity of primate lentiviruses.

Key words: AIDS/cross-species transmission/lentiviral recombination/sabaeus monkeys/TAR sequence

Introduction

Lentiviruses similar to human immunodeficiency viruses (HIVs) have been identified in a wide range of African primates including mangabeys (*Cercocebus*), guenons (*Cercopithecus*), mandrills (*Papio*) and chimpanzees (*Pan*) (reviewed in Johnson et al., 1991). Although related to HIV in their physical structure, genetic composition and replicative properties, these simian immunodeficiency viruses (SIVs) differ from the human AIDS viruses in one fundamental aspect of their biology: they fail to induce clinical immunodeficiency in their natural hosts (reviewed in Cichutek and Norley, 1993). Understanding the molecular biology of these viruses, their lack of pathogenicity despite persistent replication and the processes responsible for their adaptation to a natural host may thus be important for achieving a better understanding of the virulence of HIV in man and the mechanisms underlying AIDS pathogenesis (Kurth and Norley, 1994).

One approach to gain insight into the biological differences among primate lentiviruses is to examine their phylogenetic origins in different host species. Analyses of known SIV sequences from African non-human primates indicate that there are five major viral lineages (SIV$_{CPZ}$ from chimpanzees; SIV$_{SM}$ from sooty mangabeys; SIV$_{SYK}$ from Sykes' monkeys; SIV$_{MND}$ from mandrill; and SIV$_{AGM}$ from African green monkeys) which are roughly equidistant from each other (Figure 1). The human viruses do not form lineages separate from these; rather HIV-1 clusters with SIV$_{CPZ}$, while HIV-2 clusters with SIV$_{SM}$ (reviewed in
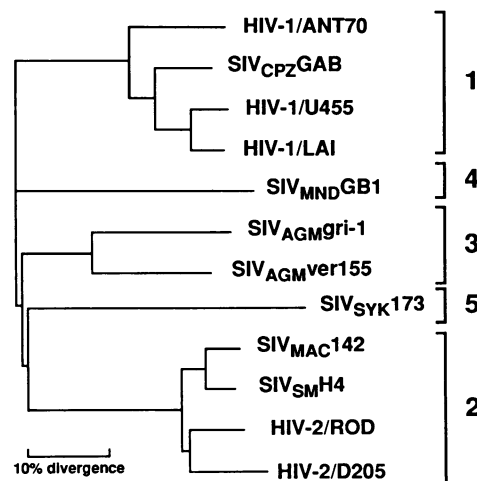

Fig. 1. Primate lentivirus phylogeny, based on pol protein sequences (adapted from Sharp et al., 1994).

Sharp *et al.*, 1994). These findings, along with epidemiological and virological data arguing for a simian origin of HIV (Hirsch *et al.*, 1989a; Huet *et al.*, 1990; Marx *et al.*, 1991; Gao *et al.*, 1992; Peeters *et al.*, 1992a), suggest that zoonotic transmissions of primate lentiviruses from naturally infected monkeys to man have generated two new human pathogens (Myers *et al.*, 1992; Sharp *et al.*, 1994).

Cross-species transmission of primate lentiviruses has also been responsible for outbreaks of epidemic disease among primates in captivity. SIV_{MAC} and SIV_{STM} cause immunodeficiency indistinguishable from human AIDS in rhesus and stump-tail macaques (Daniel *et al.*, 1985; Lowenstine *et al.*, 1988). Both of these infections are believed to have resulted from housing naturally infected sooty mangabeys with uninfected Asian macaques in USA primate centers or from inadvertent injections of animals with contaminated blood (Hirsch *et al.*, 1989a; Khan *et al.*, 1992; Novembre *et al.*, 1992). Moreover, experimental inoculation of macaques with SIV_{SM} or SIV_{AGM} can result in pathogenic infections (Murphy-Corb *et al.*, 1986; Baskin *et al.*, 1988; McClure *et al.*, 1989; Johnson *et al.*, 1990a; Hirsch *et al.*, 1994), and repeated *in vivo* passage of 'non-pathogenic' viruses can generate variants with increased virulence for both natural and unnatural hosts (Fultz *et al.*, 1989; Hirsch *et al.*, 1994). Naturally occurring primate lentiviruses are thus not inherently non-pathogenic, nor are their host species intrinsically resistant to SIV disease. Rather, lentiviral virulence appears to be dictated by a complex interplay between viral and host determinants, with

introduction of virus to a new host species (zoonosis), or serial viral passage through animals of the same species, representing contributing factors to disease expression.

Among primates known to be naturally infected with SIV, African green monkeys are the most numerous, most geographically dispersed and most commonly infected in the wild. African green monkeys have been classified as a superspecies (*Cercopithecus aethiops*) which is further subdivided into four species (Lernould 1988): *C.aethiops, sensu stricto* (common name, grivet); *C.pygerythrus* (vervet); *C.tantalus* (tantalus); and *C.sabaeus* (sabaeus). These four species are distinguishable on the basis of phenotypic markers and have different geographic ranges: grivet monkeys reside in Ethiopia and the Sudan, vervet monkeys can be found from East to South Africa, tantalus monkeys are prevalent in Central Africa and sabaeus monkeys are restricted to West Africa (since some interbreeding has been observed where these ranges overlap, some authors consider the four monkey groups as subspecies; reviewed in Lernould, 1988). Importantly, all four African green monkey species are infected with SIV_{AGM} in the wild, with seroprevalence rates as high as 30−50% in adult animals (Hendry *et al.*, 1986; Lowenstine *et al.*, 1986; Ohta *et al.*, 1988; Allan *et al.*, 1990, 1991; Hirsch *et al.*, 1993a; Müller *et al.*, 1993). African green monkeys thus comprise a large reservoir of primate lentiviruses, and may have served as a source of infection for other primate species within their natural habitat (Kodama *et al.*, 1989a).

Prior studies have shown that the spectrum of genetic

Table I. SIV_{AGM} isolates from four African green monkey species

Species	Virus isolate	Strain designation	Origin (ref.)	PCR derived sequence	
				LTR	*env* (1 kb)
Vervet	$SIV_{AGM}TYO-1$	verTYO-1	Kenya (Fukasawa *et al.*, 1988)	database[a]	database[a]
	$SIV_{AGM}155$	ver155	Kenya (Johnson *et al.*, 1990b)	database[a]	database[a]
	$SIV_{AGM}3$	ver3	Ethiopia (Baier *et al.*, 1989)	database[a]	database[a]
	$SIV_{AGM}90$	ver90	Kenya (Johnson *et al.*, 1990b)	database[b]	n/a
	$SIV_{AGM}ver-1$	ver-1 (ver692)	East Africa (Johnson *et al.*, 1990b)	database[b]	+
	$SIV_{AGM}ver-2$	ver-2	East Africa (Allan *et al.*, 1990)	+	+
Grivet	$SIV_{AGM}gri-1$	gri-1 (gri677)	Ethiopia (Allan *et al.*, 1990)	database[a]	database[a]
	$SIV_{AGM}gri-2$	gri-2	Ethiopia (Allan *et al.*, 1990)	+	+
	$SIV_{AGM}gri-3$	gri-3	Ethiopia (Allan *et al.*, 1990)	+	+
Sabaeus	$SIV_{AGM}sab-1$	sab-1	Senegal (Allan *et al.*, 1991)	+	+
	$SIV_{AGM}sab-2$	sab-2	Senegal (Allan *et al.*, 1991)	+	+
	$SIV_{AGM}sab-3$	sab-3	Senegal (Allan *et al.*, 1991)	+	+
	$SIV_{AGM}sab-4$	sab-4	Senegal (Allan *et al.*, 1991)	+	+
	$SIV_{AGM}D29$	sabD29	Senegal (Müller *et al.*, 1993)	n/a	database[b]
	$SIV_{AGM}D30$	sabD30	Senegal (Müller *et al.*, 1993)	n/a	database[b]
	$SIV_{AGM}D37$	sabD37	Senegal (Müller *et al.*, 1993)	n/a	database[b]
	$SIV_{AGM}D45$	sabD45	Senegal (Müller *et al.*, 1993)	n/a	database[b]
Tantalus	$SIV_{AGM}tan-1$	tan-1	Uganda (J.S.Allan, unpublished)	+	+
	$SIV_{AGM}tan17$	tan17	Uganda (Hirsch *et al.*, 1993a)	n/a	+
	$SIV_{AGM}tan40$	tan40	Uganda (Hirsch *et al.*, 1993a)	n/a	+
	$SIV_{AGM}tan49$	tan49	Uganda (Hirsch *et al.*, 1993a)	n/a	+
	$SIV_{AGM}B5$	tanB5	CAR (Müller *et al.*, 1993)	n/a	database[b]
	$SIV_{AGM}B14$	tanB14	CAR (Müller *et al.*, 1993)	n/a	database[b]
	$SIV_{AGM}B30$	tanB30	CAR (Müller *et al.*, 1993)	n/a	database[b]
	$SIV_{AGM}B53$	tanB53	CAR (Müller *et al.*, 1993)	n/a	database[b]

LTR and *env* sequences from different SIV_{AGM} isolates were obtained experimentally by PCR amplification of cell culture DNA (+) or from the database. Isolates are listed as originally published; strain designations have been added to indicate AGM species derivation. n/a, not available.
[a]Complete proviral sequence available
[b]Partial sequence available

diversity of SIV$_{AGM}$ exceeds that of other primate lentiviruses (Fukasawa *et al.*, 1988; Baier *et al.*, 1989; Li *et al.*, 1989; Johnson *et al.*, 1990b; Fomsgaard *et al.*, 1991). Moreover, species-specific relatedness has been observed among independent virus isolates from grivet, vervet, tantalus and sabaeus monkeys, suggesting that viral evolution may have occurred in concert with host speciation (Allan *et al.*, 1990, 1991; Hirsch *et al.*, 1993a; Müller *et al.*,

1993). Together, these findings suggest that African green monkeys have been infected with SIV$_{AGM}$ for long periods of time, possibly even before their speciation and divergence from a common ancestor (Allan *et al.*, 1991). However, a full appreciation of the phylogenetic relationships that exist among SIV strains from different African green monkey species has been restricted by the limited numbers of independent isolates studied, a limited geographic
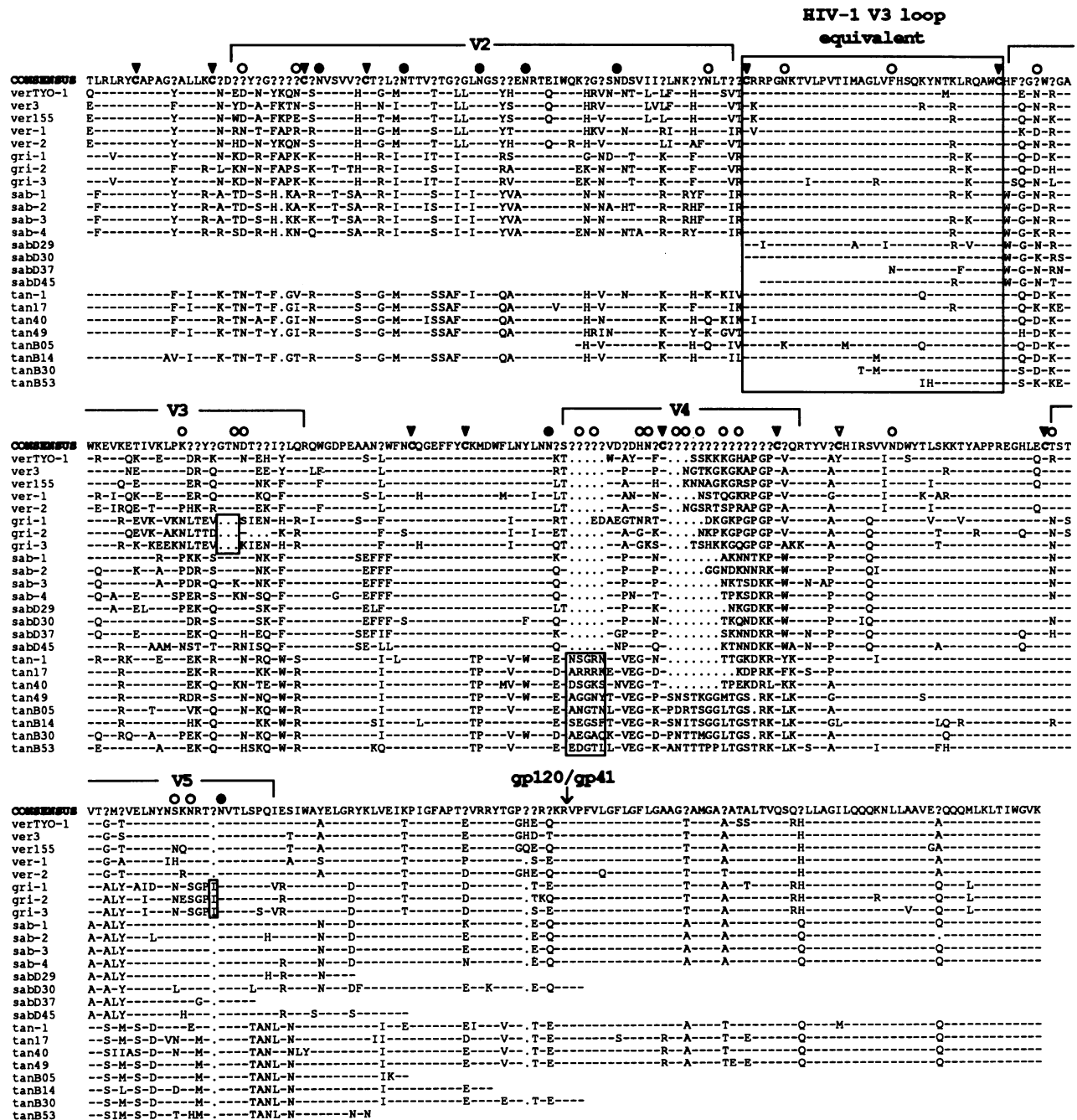
Fig. 2. Alignment of partial envelope protein sequences from multiple SIV$_{AGM}$ isolates. PCR derived *env* nucleotide sequences from vervet, grivet, tantalus and sabaeus viruses (listed in Table I) were translated, aligned with previously reported SIV$_{AGM}$ *env* sequences (Myers *et al.*, 1993) and compared with a consensus sequence generated by MASE (Faulkner and Jurkat, 1988). Dashes denote sequence identity with the consensus sequence, while dots represent gaps introduced to optimize the alignment. Question marks in the consensus sequence indicate sites in which <50% of viruses share the same amino acid residue. Triangles and circles denote cysteine residues and N-linked glycosylation sites, respectively (solid signs indicate sequence identity, while open signs indicate sequence variation). V2, V3, V4 and V5 designate hypervariable SIV$_{AGM}$ *env* domains as previously described (Müller *et al.*, 1993). A highly conserved region corresponding to the HIV-1 V3 loop, as well as species-specific codon insertions and deletions, are boxed. The envelope precursor cleavage site is indicated by an arrow above the consensus sequence.
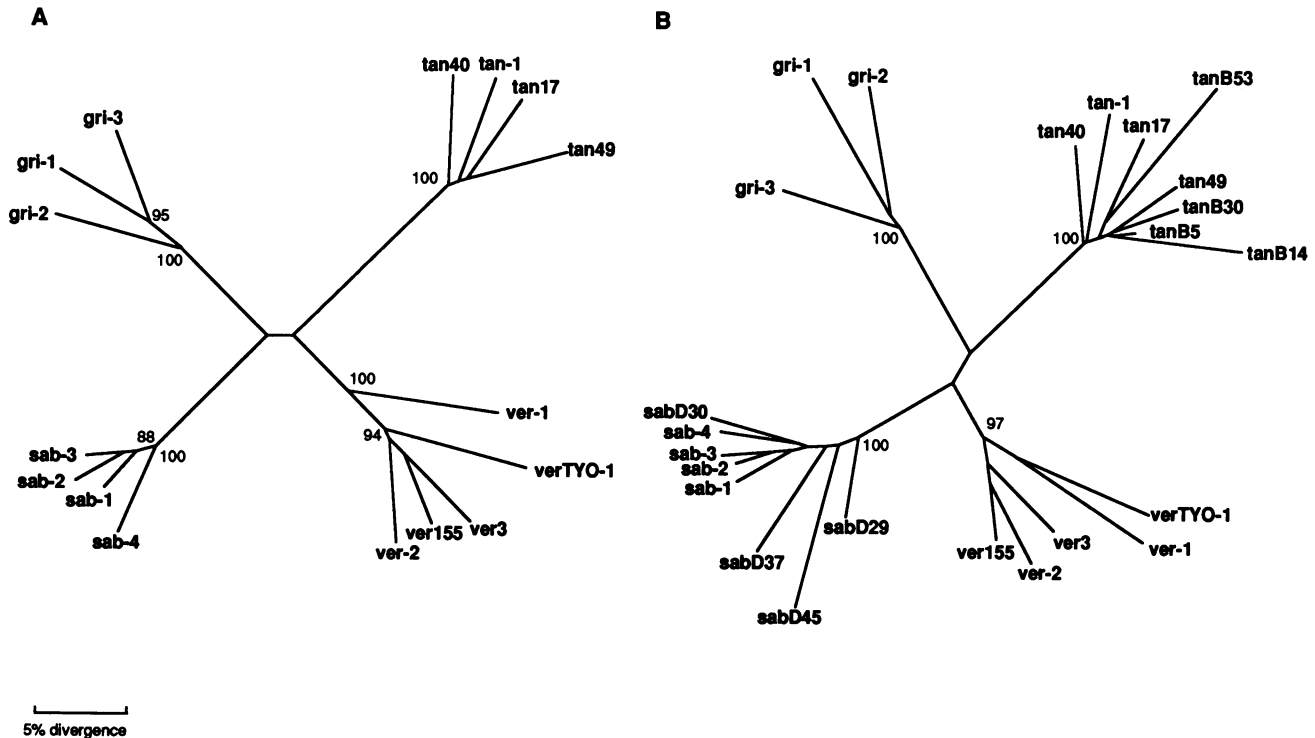
Fig. 3. Unrooted phylogenetic trees of SIV$_{AGM}$ isolates from grivet, vervet, sabaeus and tantalus monkeys. Phylogenetic relationships were estimated from comparisons of partial *env* protein sequences by the neighbor-joining method. Clusters found in >80% of 1000 bootstraps are indicated. (A) 16 isolates: 12 viruses characterized in this study compared with four previously published. The alignment consensus length is 383 residues. (B) 24 isolates: eight viruses for which shorter sequences have been recently published (Müller *et al.*, 1993) were added to the 16 in (A). The alignment consensus length is 169 residues. Virus isolates are described in Table I.

representation of these virus strains, and frequently by the analysis of only short (<500 bp) regions of the viral genome. In this study, we PCR amplified, cloned and sequenced eight complete long terminal repeat elements (LTRs) and 12 partial (1 kb) *env* regions from multiple vervet, grivet, tantalus and sabaeus viruses. We also cloned and sequenced a full-length, replication competent provirus from a sabaeus isolate thus allowing, for the first time, a complete genetic study of SIV$_{AGM}$ from West African green monkeys. Analysis of these sequences confirmed that the each of the four major African green monkey species harbors a phylogenetically distinct SIV$_{AGM}$ sub-type. However, the genome of sabaeus monkey viruses from West Africa was found to represent a mosaic of sequences derived from highly divergent lentiviruses infecting African green monkeys and sooty mangabeys. These results indicate a complex evolutionary history for SIV$_{AGM}$, involving recombination and cross-species transmission in the wild.

Results

SIVs from African green monkeys fall into phylogenetic sub-groups according to their species of origin

To determine the phylogenetic relationships of SIV$_{AGM}$ from different African green monkey populations, we PCR amplified, cloned and sequenced a 1 kb envelope region from a total of 12 geographically-diverse vervet, grivet, tantalus and sabaeus isolates. PCR primers were designed to amplify sub-genomic fragments of sufficient length for phylogenetic analyses, while at the same time allowing comparison with

previously determined *env* sequences of other SIV$_{AGM}$ isolates (Müller *et al.*, 1993). Using this approach, we were able to compare 24 independent SIV$_{AGM}$ isolates from widely varying geographic locales representing all four African green monkey species. Importantly, most vervet and grivet viruses as well as all sabaeus and tantalus isolates were obtained from wild-caught animals. Table I summarizes the designations and geographic origins of these isolates, along with references describing their molecular and biological characterization.

All SIV$_{AGM}$ envelope sequences were amplified by single round PCR from tissue culture propagated viruses. Amplified fragments contained between 1101 and 1137 bp of nucleotide sequence (excluding primer sequences) and included portions of both the surface (SU) envelope glycoprotein (beginning at the second variable region) and the transmembrane (TM) domain (ending 62 amino acids downstream of the precursor cleavage site). Alignment of their deduced amino acid sequences with those of published SIV$_{AGM}$ strains identified hypervariable (V2−V5) as well as conserved envelope domains as reported previously (Müller *et al.*, 1993). Regions corresponding to HIV-1 envelope domains of known function (reviewed in McKeating and Wiley, 1989), including the envelope glycoprotein precursor cleavage site, the CD4 binding domain and the viral fusion peptide (N-terminus of gp41) were all highly conserved (Figure 2). Also conserved were the majority of cysteine residues (11 of 12) and, as noted previously (Allan, 1993; Müller *et al.*, 1993), there was very little sequence variability in the region that corresponds to the hypervariable V3 loop of HIV-1.

Pairwise amino acid sequence comparisons were

performed to examine the extent of SIV_{AGM} genetic diversity in the PCR amplified envelope fragment. As reported previously, SIV_{AGM} isolates were more closely related to each other than they were to other primate lentiviruses (data not shown). Moreover, the extent of sequence variation differed depending on whether isolates from the same or different African green monkey species were compared. Amino acid sequence differences between viruses from the same species (e.g. between vervet viruses) ranged between 5 and 17%, while interspecies variation (e.g. between vervet and grivet viruses) was higher and ranged between 21 and 31% (data not shown). Sequence alignments also revealed species-specific codon insertions and deletions (Figure 2). For example, all eight tantalus viruses contained a five codon insertion in the V4 region, while all three grivet viruses were characterized by a three codon deletion in V3 and a single codon insertion in V5. In agreement with earlier reports (Hirsch et al., 1993a; Müller et al., 1993), these results confirm host-specific differentiation and molecular signatures among a larger number of SIV_{AGM} isolates.

To determine the evolutionary relationships among the various African green monkey viruses, phylogenetic trees were constructed using the neighbor-joining algorithm (Saitou and Nei, 1987). First, 16 SIV_{AGM} strains were compared across a protein alignment of consensus length 383 residues. The results (Figure 3A) demonstrated that vervet, grivet, sabaeus and tantalus monkey viruses each form distinct monophyletic groups, as indicated by the long branches leading to each viral sub-group and the 100% bootstrap values outside each sub-group. Second, we included eight additional strains for which only shorter env sequences could be compared (consensus alignment length 169 residues). This analysis yielded very similar results (Figure 3B). Taken together, the trees in Figure 3 demonstrate that the evolutionary relationships among SIV_{AGM} isolates are correlated with host species rather than geographical origin. This is particularly apparent in the tantalus cluster which comprises viruses from monkeys captured more than 1000 miles apart, in the Central African Republic (CAR) (Müller et al., 1993) and Uganda (Hirsch et al., 1993a), but is also evident in the separate grouping of vervet and grivet viruses which form distinct phylogenetic clusters despite their common geographic origin (Kenya/Ethiopia). These results thus indicate that vervet, grivet, sabaeus and tantalus monkeys each harbor their own SIV_{AGM} sub-types which have likely evolved within their respective hosts for long periods of time.

Sabaeus viruses contain an LTR structure unique to the HIV-2/SIV_SM group of viruses

LTRs are responsible for the transcriptional regulation of primate lentiviruses and are thus important for viral replication and pathogenesis (reviewed in Gaynor, 1992). To determine whether viruses from the different African green monkeys exhibited species-specific differences in these regulatory regions, we PCR amplified, cloned and sequenced complete LTRs from eight additional vervet, grivet, tantalus and sabaeus viruses (see Table I). Together with published sequences, this allowed a comprehensive analysis of 14 complete SIV_{AGM} LTRs.

Comparisons among these LTR sequences confirmed species-specific diversity in a second genomic region.

Individual vervet, grivet and sabaeus LTRs differed in 10—28% of their nucleotide sequence, while LTRs from viruses representing different species exhibited between 22 and 42% divergence. There were also considerable length differences among the various SIV_{AGM} LTRs. ver-1 contained the shortest LTR (634 bp), while LTRs from sabaeus viruses were considerably longer (767—776 bp). To identify which regions were affected by these length differences, functional LTR domains surrounding the core enhancer sequences were aligned (Figure 4). This analysis revealed that known regulatory elements, including NFκB sites, SP1 sites, the TATA region, the transactivation response element (TAR) and polyadenylation [poly(A)] sites were highly conserved among all SIV_{AGM} LTRs. Length variation, however, resulted from differences in copy numbers of these regulatory elements. All vervet and grivet LTRs, except for ver-1, contained two NFκB binding sites; in contrast, all sabaeus and tantalus LTRs contained only one (Figure 4). Similarly, most SIV_{AGM} LTRs contained three potential (non-consensus) SP1 binding sites, except for ver-1 which contained one and gri-1 which contained four. The most striking difference, however, involved the TAR element, which constitutes the binding site for the viral tat protein and cellular factor(s) important for viral transactivation (for review see Gaynor, 1992). In contrast to all other SIV_{AGM} strains, the LTRs from all four sabaeus viruses contained two copies of this TAR element. Twenty-four out of 26 base pairs, including two sequence motifs of essential function (a 2 bp bulge and 6 bp terminal loop), were perfectly duplicated. Moreover, these duplicated sequences were surrounded by additional nucleotide insertions, which accounted for most of the length differences between sabaeus and other SIV_{AGM} LTRs.

Duplicated TAR sequences have previously only been reported for the LTRs of the sooty mangabey group of viruses (including HIV-2 and SIV_{MAC}), where they fold into complicated RNA structures (Fenrick et al., 1989; Berkhout, 1992). Figure 5 depicts secondary structure predictions for vervet, grivet, tantalus and sabaeus TAR RNA sequences, along with TAR structures experimentally determined for HIV-1 and HIV-2 (SIV_{SM} and SIV_{MAC} TAR structures are not shown because they are almost identical to that of HIV-2; Fenrick et al., 1989; Berkhout, 1992). The results revealed duplicated TAR hairpin loops in the sabaeus LTR, each containing a 2 bp bulge (binding site for tat), a 4 bp stem and a 6 bp terminal loop with the consensus sequence 3'-CUGGGX-5' [recognition site for a cellular factor(s) which cooperates with tat in transactivation of HIV gene expression; reviewed in Gaynor, 1992]. In contrast, vervet, grivet and tantalus LTRs each exhibited only one TAR stem—loop structure with these sequence motifs. Thus, the sabaeus viruses alone share this LTR structural feature with otherwise apparently highly divergent lentiviruses infecting humans and sooty mangabeys in West Africa.

Due to sequence insertions surrounding the second TAR element, the computer predicted two additional stem—loops in the sabaeus TAR region downstream of the putative tat binding domains. These structures differ in primary sequence and overall length from a similarly located third stem—loop in HIV-2 and SIV_{AGM} LTRs (Sakuragi et al., 1991; Berkhout, 1992). Whether the two sabaeus stem—loops are actually formed in native viral RNA and whether they
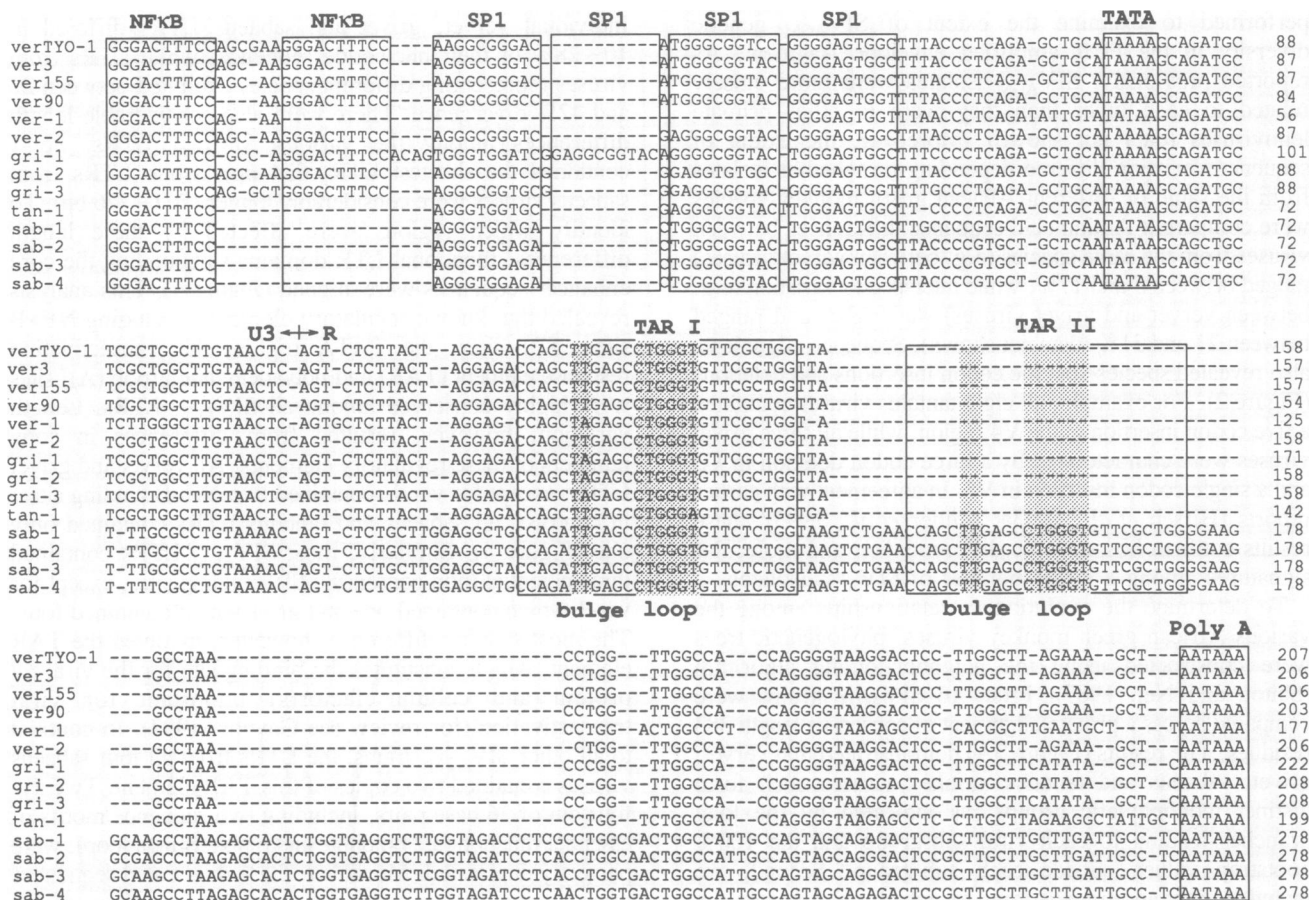
```
              NFκB           NFκB         SP1       SP1        SP1       SP1              TATA
verTYO-1  GGGACTTTCCAGCGAAGGGACTTTCC----AAGGCGGGAC----------ATGGGCGGTCC-GGGGAGTGGCTTTACCCTCAGA-GCTGCATAAAAGCAGATGC  88
ver3      GGGACTTTCCAGC-AAGGGACTTTCC----AGGGCGGGTC----------ATGGGCGGTAC-GGGGAGTGGCTTTACCCTCAGA-GCTGCATAAAAGCAGATGC  87
ver155    GGGACTTTCCAGC-ACGGGACTTTCC----AAGGCGGGAC----------ATGGGCGGTAC-GGGGAGTGGCTTTACCCTCAGA-GCTGCATAAAAGCAGATGC  87
ver90     GGGACTTTCC----AAGGGACTTTCC----AGGGCGGGCC----------ATGGGCGGTAC-GGGGAGTGGTTTTACCCTCAGA-GCTGCATAAAAGCAGATGC  84
ver-1     GGGACTTTCCAG--AA-------------------------------------GGGGAGTGGTTTAACCCTCAGATATTGTATATAAGCAGATGC            56
ver-2     GGGACTTTCCAGC-AAGGGACTTTCC----AGGGCGGGTC----------GAGGGCGGTAC-GGGGAGTGGCTTTACCCTCAGA-GCTGCATAAAAGCAGATGC  87
gri-1     GGGACTTTCC-GCC-AGGGACTTTCCACAGTGGGTGGATCGGAGGCGGTACAGGGGCGGTAC-TGGGAGTGGCTTTCCCCTCAGA-GCTGCATAAAAGCAGATGC 101
gri-2     GGGACTTTCCAGC-AAGGGACTTTCC----AGGGCGGTCCG----------GGAGGTGTGGC-GGGGAGTGGCTTTACCCTCAGA-GCTGCATAAAAGCAGATGC  88
gri-3     GGGACTTTCCAG-GCTGGGGCTTTCC----AGGGCGGTGCG----------GGAGGCGGTGGTTTTGCCCTCAGA-GCTGCATAAAAGCAGATGC            88
tan-1     GGGACTTTCC--------------------AGGGTGGTGC----------GAGGGCGGTACTTGGGAGTGGCTT-CCCCTCAGA-GCTGCATAAAAGCAGATGC   72
sab-1     GGGACTTTCC--------------------AGGGTGGAGA----------CTGGGCGGTAC-TGGGAGTGGCTTGCCCCGTGCT-GCTCAATATAAGCAGCTGC   72
sab-2     GGGACTTTCC--------------------AGGGTGGAGA----------CTGGGCGGTAC-TGGGAGTGGCTTACCCCGTGCT-GCTCAATATAAGCAGCTGC   72
sab-3     GGGACTTTCC--------------------AGGGTGGAGA----------CTGGGCGGTAC-TGGGAGTGGCTTACCCCGTGCT-GCTCAATATAAGCAGCTGC   72
sab-4     GGGACTTTCC--------------------AGGGTGGAGA----------CTGGGCGGTAC-TGGGAGTGGCTTACCCCGTGCT-GCTCAATATAAGCAGCTGC   72

               U3 ←+→ R                                 TAR I                          TAR II
verTYO-1  TCGCTGGCTTGTAACTC-AGT-CTCTTACT--AGGAGACCAGCTTGAGCCTGGGTGTTCGCTGGTTA----------------------------------  158
ver3      TCGCTGGCTTGTAACTC-AGT-CTCTTACT--AGGAGACCAGCTTGAGCCTGGGTGTTCGCTGGTTA----------------------------------  157
ver155    TCGCTGGCTTGTAACTC-AGT-CTCTTACT--AGGAGACCAGCTTGAGCCTGGGTGTTCGCTGGTTA----------------------------------  157
ver90     TCGCTGGCTTGTAACTC-AGT-CTCTTACT--AGGAGACCAGCTTGAGCCTGGGTGTTCGCTGGTTA----------------------------------  154
ver-1     TCTTGGGCTTGTAACTC-AGTGCTCTTACT--AGGAG-CCAGCTAGAGCCTGGGTGTTCGCTGGT-A----------------------------------  125
ver-2     TCGCTGGCTTGTAACTCCAGT-CTCTTACT--AGGAGACCAGCTTGAGCCTGGGTGTTCGCTGGTTA----------------------------------  158
gri-1     TCGCTGGCTTGTAACTC-AGT-CTCTTACT--AGGAGACCAGCTAGAGCCTGGGTGTTCGCTGGTTA----------------------------------  171
gri-2     TCGCTGGCTTGTAACTC-AGT-CTCTTACT--AGGAGACCAGCTAGAGCCTGGGTGTTCGCTGGTTA----------------------------------  158
gri-3     TCGCTGGCTTGTAACTC-AGT-CTCTTACT--AGGAGACCAGCTAGAGCCTGGGTGTTCGCTGGTTA----------------------------------  158
tan-1     TCGCTGGCTTGTAACTC-AGT-CTCTTACT--AGGAGACCAGCTTGAGCCTGGGAGTTCGCTGGTGA----------------------------------  142
sab-1     T-TTGCGCCTGTAAAAC-AGT-CTCTGCTTGGAGGCTGCCAGATTGAGCCTGGGTGTTCTCTGGTAAGTCTGAACCAGCTTGAGCCTGGGTGTTCGCTGGGGAAG  178
sab-2     T-TTGCGCCTGTAAAAC-AGT-CTCTGCTTGGAGGCTGCCAGATTGAGCCTGGGTGTTCTCTGGTAAGTCTGAACCAGCTTGAGCCTGGGTGTTCGCTGGGGAAG  178
sab-3     T-TTGCGCCTGTAAAAC-AGT-CTCTGCTTGGAGGCTGCCAGATTGAGCCTGGGTGTTCTCTGGTAAGTCTGAACCAGCTTGAGCCTGGGTGTTCGCTGGGGAAG  178
sab-4     T-TTTCGCCTGTAAAAC-AGT-CTCTGTTTGGAGGCTGCCAGATTGAGCCTGGGTGTTCTCTGGTAAGTCTGAACCAGCTTGAGCCTGGGTGTTCGCTGGGGAAG  178
                                      bulge loop                              bulge loop
                                                                                                    Poly A
verTYO-1  ----GCCTAA-------------------------------------CCTGG---TTGGCCA---CCAGGGGTAAGGACTCC-TTGGCTT-AGAAA--GCT---AATAAA  207
ver3      ----GCCTAA-------------------------------------CCTGG---TTGGCCA---CCAGGGGTAAGGACTCC-TTGGCTT-AGAAA--GCT---AATAAA  206
ver155    ----GCCTAA-------------------------------------CCTGG---TTGGCCA---CCAGGGGTAAGGACTCC-TTGGCTT-AGAAA--GCT---AATAAA  206
ver90     ----GCCTAA-------------------------------------CCTGG---TTGGCCA---CCAGGGGTAAGGACTCC-TTGGCTT-GGAAA--GCT---AATAAA  203
ver-1     ----GCCTAA-------------------------------------CCTGG--ACTGGCCCT--CCAGGGGTAAGAGCCTC-CACGGCTTGAATGCT-----TAATAAA  177
ver-2     ----GCCTAA-------------------------------------CTGG---TTGGCCA---CCAGGGGTAAGGACTCC-TTGGCTT-AGAAA--GCT---AATAAA  206
gri-1     ----GCCTAA-------------------------------------CCCGG---TTGGCCA---CCGGGGGTAAGGACTCC-TTGGCTTCATATA--GCT---CAATAAA  222
gri-2     ----GCCTAA-------------------------------------CC-GG---TTGGCCA---CCGGGGGTAAGGACTCC-TTGGCTTCATATA--GCT---CAATAAA  208
gri-3     ----GCCTAA-------------------------------------CC-GG---TTGGCCA---CCGGGGGTAAGGACTCC-TTGGCTTCATATA--GCT--CAATAAA  208
tan-1     ----GCCTAA-------------------------------------CCTGG--TCTGGCCAT--CCAGGGGTAAGAGCCTC-CTTGCTTAGAAGGCTATTGCTAATAAA  199
sab-1     GCAAGCCTAAGAGCACTCTGGTGAGGTCTTGGTAGACCCTCGCCTGGCGACTGGCCATTGCCAGTAGCAGAGACTCCGCTTGCTTGCTTGATTGCC-TCAATAAA  278
sab-2     GCGAGCCTAAGAGCACTCTGGTGAGGTCTTGGTAGATCCTCACCTGGCGACTGGCCATTGCCAGTAGCAGGGACTCCGCTTGCTTGCTTGATTGCC-TCAATAAA  278
sab-3     GCAAGCCTAAGAGCACTCTGGTGAGGTCTCGGTAGATCCTCACCTGGCGACTGGCCATTGCCAGTAGCAGAGACTCCGCTTGCTTGCTTGATTGCC-TCAATAAA  278
sab-4     GCAAGCCTTAGAGCACACTGGTGAGGTCTTGGTAGATCCTCAACTGGTGACTGGCCATTGCCAGTAGCAGAGACTCCGCTTGCTTGCTTGATTGCC-TCAATAAA  278
```

Fig. 4. Alignment of SIV$_{AGM}$ LTR sequences surrounding the core enhancer region. PCR derived LTR sequences from vervet, grivet, tantalus and sabaeus viruses (listed in Table I) are compared with previously published vervet and grivet LTR sequences (Myers *et al.*, 1993). Putative transcription factor binding sites, including NFκB, SP1, TATA, TAR and poly(A), are boxed (Gaynor, 1992). Dashes indicate gaps introduced to optimize the alignments. Functional motifs within the TAR domains, including bulge and loop structures of the TAR hairpin, are highlighted (also see Figure 5). The putative U3/R boundaries are indicated by an arrow.

influence viral transcription or transactivation, remain to be determined. Because of these additional sequences, however, the sabaeus TAR region appears to be one of extraordinary complexity. Future analysis will have to determine whether this has biological relevance.

Genomic organization of SIV$_{AGM}$ from West African sabaeus monkeys is similar to that of the other SIV$_{AGM}$ strains

To examine whether there were any additional similarities between sabaeus viruses and other West African primate lentiviruses, we molecularly cloned and sequenced a complete sabaeus virus genome. To generate a complete proviral clone, genomic DNA of isolate SIV$_{AGM}$sab-1 (see Table I) was partially digested with *Mbo*I and ligated into *Bam*HI cleaved arms of λGEM12. Following screening with a sabaeus-specific LTR probe, four hybridization positive lambda phage clones were identified, one of which (gMJ8) contained a full-length proviral genome based on restriction enzyme analysis. To test the replication competence of this clone, proviral DNA was transfected into Cos-1 cells and transfected cultures were subsequently cocultivated with Molt4 (clone 8) cells. Positive RT activity and p27 antigen levels in culture supernatants, as well as syncytium formation in acutely infected cells, indicated that the sabaeus provirus was biologically active and replicated with considerable

cytopathic effect (Figure 6B and C). Moreover, cell-free transmission studies confirmed that it was able to infect susceptible target cells in a cell independent fashion. Finally, time course studies with equivalent amounts of input virus (based on RT activity) revealed that the sabaeus clone reproducibly replicated to higher titers than previously reported vervet (SA-212; Shibata *et al.*, 1990) and grivet (gri-1/λII; Fomsgaard *et al.*, 1991) reference clones (Figure 6A). Because of these properties, MJ8 was selected for nucleotide sequence analysis.

The complete sabaeus proviral sequence was 10 036 nucleotides long (data not shown; sequence available from the EMBL/GenBank DNA sequence data library under accession number U04005). Analysis revealed the presence of LTRs of 767 bp, and eight major open reading frames corresponding to *gag*, *pol*, *vif*, *vpx*, *tat*, *rev*, *env* and *nef* genes. With respect to their overall length, genomic location and the identifiable functional motifs of their putative protein products, these genes were similar to their homologs in vervet and grivet proviruses. As in other SIV$_{AGM}$ genomes, there was no *vpr* equivalent (Fukasawa *et al.*, 1988). Regulatory sequences including the primer binding site (5′-TGGCGCCCGAACAGGGAC-3′), the polypurine tract (5′-AAAAGAAAAGGGAGGA-3′), the ribosomal frameshift site (5′-TTTTTTAGG-3′), as well as major splice donor and acceptor sites, were highly conserved between
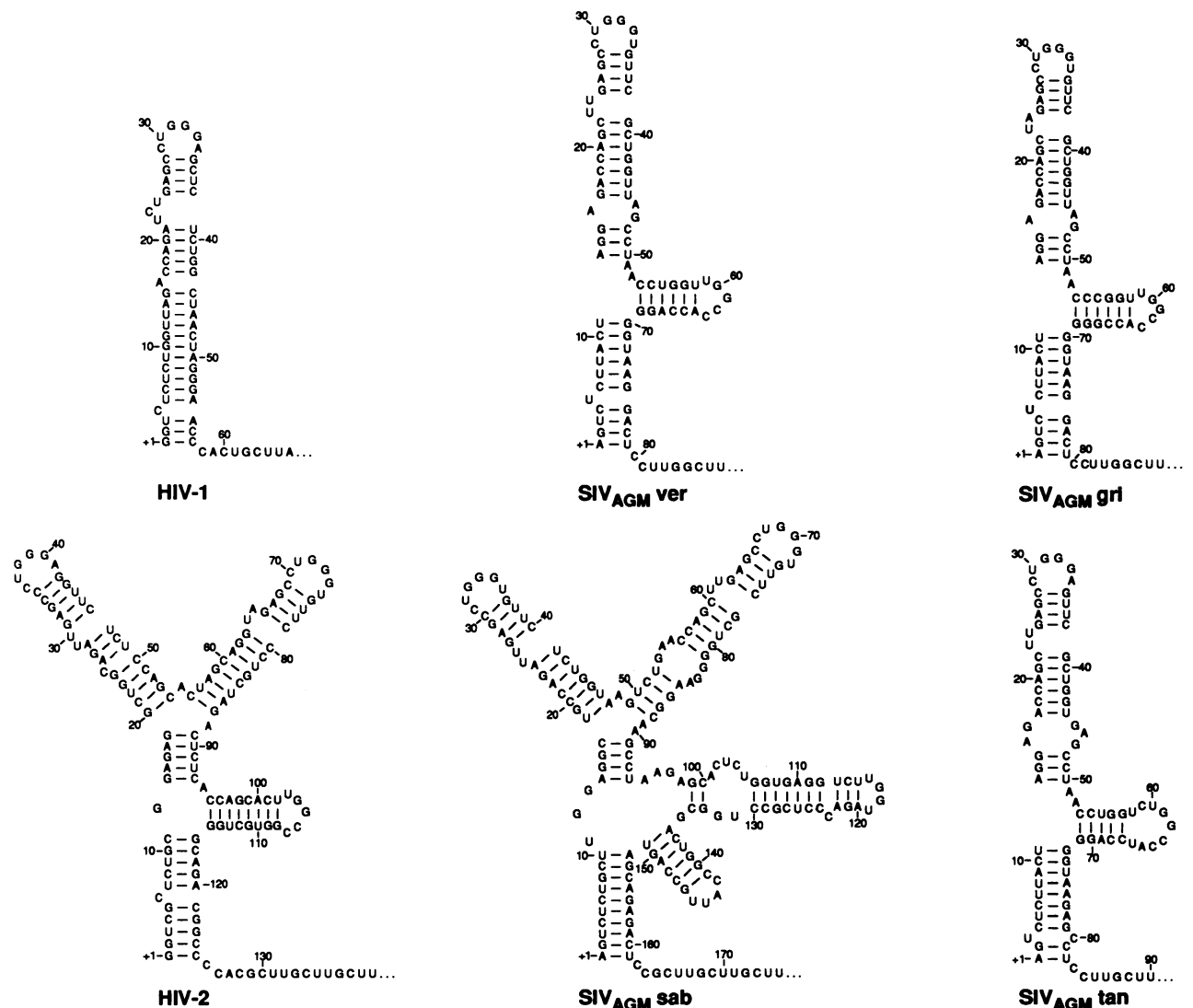
Fig. 5. Secondary structure comparison of TAR RNA sequences from SIV_AGM, HIV-1 and HIV-2. Secondary structures were predicted using the SQUIGGLES program of the UWGCG software package. The predicted free energy values for each of the structures are: −25.5 kcal/mol (HIV-1/HXB2), −56.7 kcal/mol (HIV-2/ROD), −31.3 kcal/mol (verTYO-1), −32.7 kcal/mol (gri-1), −27.2 kcal/mol (tan-1) and −55.6 kcal/mol (sab-1). The hairpin structures of HIV-1 and HIV-2 have been verified experimentally (Fenrick et al., 1989; Berkhout, 1992).

the sabaeus and other African green monkey proviruses. Similarly conserved were functional protein domains, including the myristylation signal in p16 and nef (N-terminal glycine), the metal binding domain of the gag nucleocapsid (C-X2-C-X4-H-X4-C), the active sites in protease (DTG) and reverse transcriptase (YMDD), as well as cysteine-rich and basic regions in tat and rev. Amino acid sequence alignments, however, also identified differences that were unique to the sabaeus provirus. The sabaeus genome contained a vpx gene that was considerably longer (20 amino acids) than that of vervet and grivet viruses. Moreover, its env gene was prematurely terminated by an in-frame stop codon at amino acid position 732. This created a very short cytoplasmic anchor of ~17 amino acid residues. Finally, the sabaeus tat coding region also contained an in-frame stop codon which resulted from a nucleotide point mutation immediately preceding the splice junction between the first and second tat exons. The CA^A (Gln) codon present in all vervet and grivet proviruses was changed to a premature TA^A (stop) codon (with ^ indicating the splice junction).

Thus, unless alternative splice sites are used, tat expression in MJ8 is restricted to the first coding exon. The observed env and tat sequence changes, however, may not be representative of sabaeus proviruses as they exist in vivo. Premature truncations of SIV env genes are frequently observed following in vitro propagation and are believed to reflect viral adaptation to human cell lines (Hirsch et al., 1989b; Kodama et al., 1989b). Similarly, truncation of tat may also be a by-product of growth in tissue culture, since expression of the first tat exon is sufficient for in vitro viral replication (reviewed in Cullen, 1992).

Phylogenetic analysis of SIV from West African green monkeys reveals a mosaic genome

To determine the evolutionary relationships of the sabaeus provirus to other primate lentiviruses, phylogenetic trees were constructed from full-length gag, pol and env protein sequences. The env derived tree indicated, as expected, that the sabaeus virus clustered with the other African green monkey viruses (vervet and grivet isolates, in this case, since
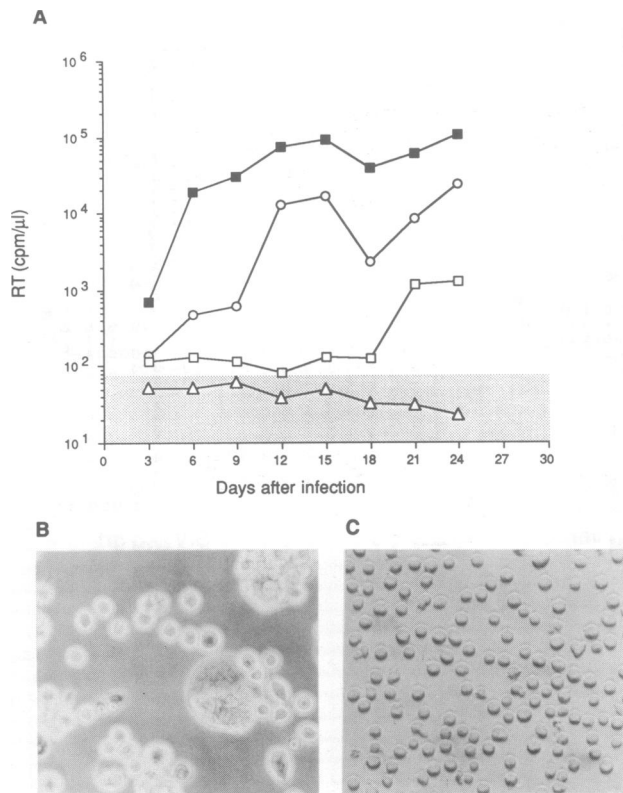
Fig. 6. Biological characterization of the sabaeus provirus. (A) Replication characteristics of the sab-1 clone MJ8 and two other SIV$_{AGM}$ reference clones, the grivet clone gri-1/λII (Foomsgaard *et al.*, 1991) and the vervet (TYO-1 isolate) clone SA-212 (Shibata *et al.*, 1990), assessed in Molt4 (clone 8) cells. Molt4 (clone 8) cells (1×10^6) were infected with equal amounts of the respective viruses (660 000 c.p.m. RT) derived from Cos-1 cell transfection of proviral DNA. Virus production was monitored by measuring supernatant RT activity in 3 day intervals as described in Li *et al.* (1991). The shaded area at the bottom represents the background range for the RT measurements. (B) Syncytium induction in Molt4 (clone 8) cells acutely infected with the sab-1 clone MJ8. (C) Uninfected Molt4 (clone 8) cells. ■, MJ8; ○, gri-1/λII; □, SA-212; △, mock.

no full-length tantalus virus sequence is available). However, *gag* and *pol* derived trees had a different topology. In both cases, the sabaeus virus formed a sixth distinct lineage, approximately equidistant from each of the other five lineages (as depicted in Figure 1), though clustering (in both cases) non-significantly with the SIV$_{SM}$/HIV-2/SIV$_{MAC}$ group of viruses. Such discrepant phylogenetic positions in trees based on analyses of different coding regions of the viral genome have been seen in HIV-1 (Li *et al.*, 1988) and HIV-2 (Gao *et al.*, 1992), and were taken to indicate that an ancestor of the virus under study was a recombinant. To examine this possibility, additional phylogenetic analyses were performed using shorter fragments of each of the three major proteins, as well as some of the accessory proteins (*vif, vpx* and *nef*). These results are summarized in Figure 7. In trees derived from protein sequences encoded by the central and 3' regions of the genome, i.e. the C-terminal part of *pol*, or *vif, vpx, env* or *nef*, the sabaeus virus clustered with other African green monkey isolates. This clustering was highly significant for the longer sequences (i.e. 3' *pol* and *env*), as judged by bootstrap values of 99% or 100%. In contrast, in trees derived from the C-terminal half of *gag* or the N-terminal region of *pol* (whose coding sequences overlap

somewhat), the sabaeus virus grouped with, though at some distance from, the SIV$_{SM}$/HIV-2 lineage. These results strongly suggest that these two regions of the sabaeus genome have different evolutionary histories. Finally, in trees derived from the remaining part of *gag*, i.e. the N-terminal half, a third result emerged: the sabaeus virus did not cluster with either the vervet/grivet SIV$_{AGM}$ lineage or the SIV$_{SM}$/HIV-2 lineage, but formed a distinct lineage equidistant from both of these.

Additional phylogenetic analyses were performed, dividing the *gag* and *pol* proteins at different sites in an attempt to map the positions of the putative recombination events more precisely. We also surveyed the distribution of phylogenetically informative sites supporting the alternative locations of the sabaeus lineage within the primate lentiviral tree (Figure 7). These analyses allowed approximate location of the breakpoints to the regions around codon 310 of the *gag* gene, and around 710 of the *pol* gene (where numbers refer to the position in the vervet consensus sequence in the alignment of Myers *et al.*, 1993; pages II-C-2 and II-C-6).

To investigate whether the sab-1 isolate was unusual in having a mosaic genome, we examined a second sabaeus virus, termed sabD37, which was originally isolated by investigators at the Institut Pasteur in Dakar, Senegal (Müller *et al.*, 1993). DNA from tissue culture cells infected with sabD37 was subjected to PCR amplification in the 5' part of the *pol* gene, cloned and sequenced. A phylogenetic tree constructed from the deduced amino acid sequence showed that sabD37 clustered with sab-1. Importantly, both viruses also clustered more closely with the SIV$_{SM}$/HIV-2 lineage than with grivet or vervet viruses (Figure 8). Although both sab-1 and sabD37 were derived from monkeys captured in Senegal (Allan *et al.*, 1991; Müller *et al.*, 1993), it is unlikely that the two viruses, which were independently studied by investigators in France and USA, were derived from the same monkey troop, or were otherwise epidemiologically linked. Indeed, the two viruses exhibited considerable sequence divergence in both *pol* (Figure 8) and *env* (Figure 2) regions, indicating that they are not unusually closely related among sabaeus viruses. These results thus suggest that the mosaic genome structure is a general feature of SIVs from West African green monkeys.

Discussion

Previous phylogenetic analyses have suggested that SIV$_{AGM}$ isolates cluster according to their African green monkey species of origin (Allan *et al.*, 1991; Hirsch *et al.*, 1993a; Müller *et al.*, 1993). These studies, however, were based on comparisons of only small numbers of isolates and short sub-genomic regions, and were further limited by the restricted geographic representation of the animal hosts. Thus, the possibility remained that the observed sequence relationships merely reflected a recent spread of virus within animals from a particular geographic region. To determine whether this was the case, we extended our analyses of SIV$_{AGM}$ genetic variation to include a larger number of isolates from more widespread geographic locales. We also analyzed multiple independent isolates from animals representing each of the four African green monkey species and compared these viral strains in two different genomic regions. We now report that host species, and not geography, correlate with SIV$_{AGM}$ evolutionary relationships. This is
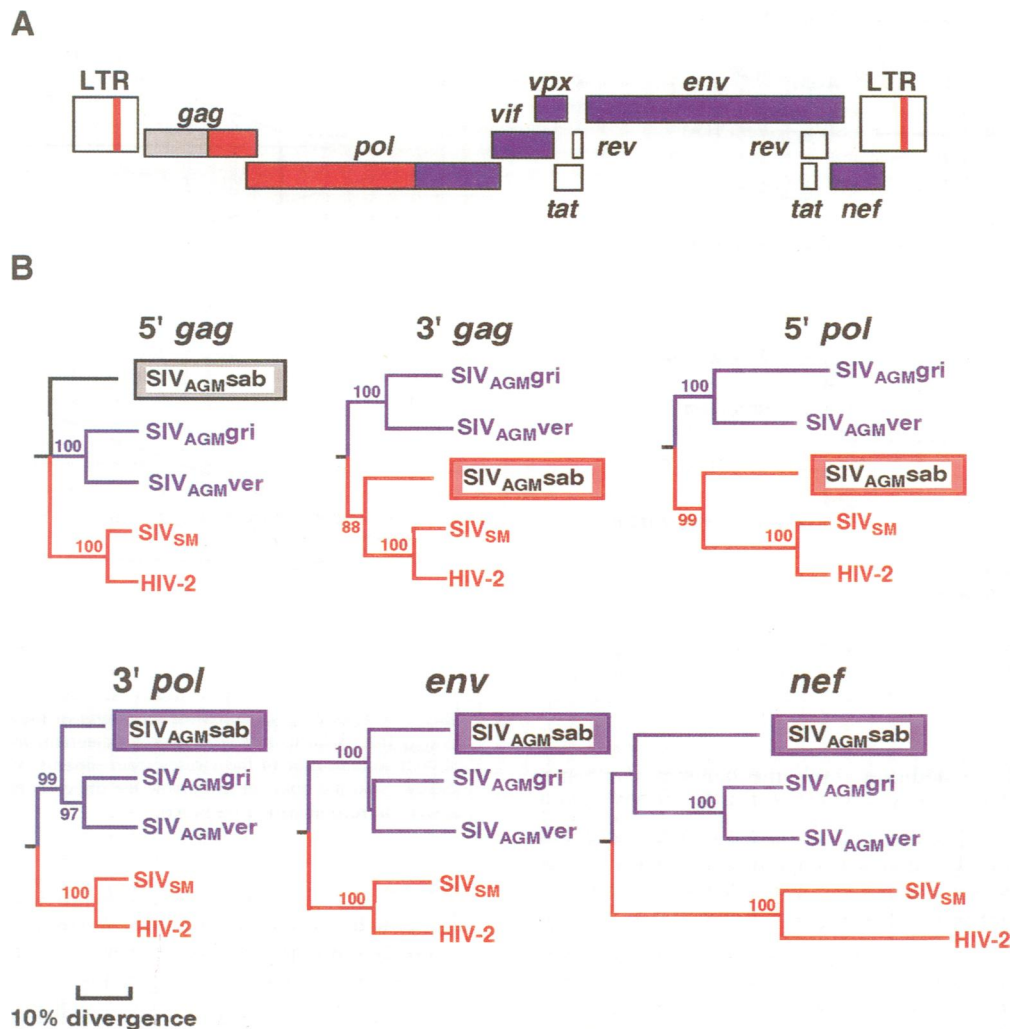
Fig. 7. Mosaic evolutionary history of the SIV$_{AGM}$sab-1 genome. (A) Schematic diagram of the genomic organization of sab-1. Genomic regions are colored according to whether the sab-1 sequence is more closely related to other SIV$_{AGM}$ strains (blue), to the SIV$_{SM}$/HIV-2 lineage (red) or to both virus strains equally (gray), based on trees in (B); the red band in the LTR represents the TAR duplication (see Figures 3 and 4). (B) Phylogenetic relationships of regions of the sab-1 genome to viruses from other African green monkeys (grivet virus, gri-1, and vervet virus, TYO-1), and to viruses from sooty mangabeys (SIV$_{SM}$H4, Hirsch *et al.*, 1989a) and humans (HIV-2/ROD; Guyader *et al.*, 1987). Trees were estimated from comparison with protein sequences by the neighbor-joining method, using SIV from mandrill (SIV$_{MND}$GB1; Tsujimoto *et al.*, 1989) as the outgroup. The scale bar indicates 10% divergence (or 0.10 amino acid replacements per site). Clusters found in >80% of 1000 bootstraps are indicated.

best exemplified in the phylogenetic trees depicted in Figure 4 which show that viruses from tantalus monkeys captured more than 1000 miles apart (Uganda and CAR) cluster together, whereas viruses infecting vervet and grivet monkeys in the same geographic region (Ethiopia and Kenya) do not. Moreover, we have identified several signature sequences which independently confirm host-specific differentiation. These include the arrangement and copy number of transcription factor binding sites in the LTR (NFκB, SP1, TAR) as well as species-specific insertions and deletions in the viral *env* gene. Taken together, these results show that vervet, grivet, tantalus and sabaeus monkeys each harbor their own SIV$_{AGM}$ sub-type which supports the hypothesis that African green monkeys have been infected with SIV$_{AGM}$ for a very long time—perhaps since their divergence from a common ancestor (Allan *et al.*, 1991).

We have also determined the first complete proviral sequence for SIV$_{AGM}$ from a sabaeus monkey. Phylogenetic analyses of sequences from different regions of the sabaeus genome yielded significantly discordant results, indicating

that these regions have different evolutionary histories; such an evidently mosaic genome most likely resulted from a recombination event involving quite divergent primate lentiviruses. Retroviruses are known to undergo frequent recombination (Hu and Temin, 1990), and this propensity appears to be a consequence of their dimeric genome, and a reverse transcriptase enzyme that can switch templates during proviral DNA synthesis (reviewed in Coffin, 1992). Two models for the recombination process have been proposed, both of which are likely operative during reverse transcription. One, the copy-choice model (Coffin, 1979), postulates a transfer of reverse transcriptase to the second RNA molecule during minus strand synthesis. The other, the strand displacement/assimilation model (Junghans *et al.*, 1982), proposes recombination via displacement of the elongating DNA plus strand. Importantly, both mechanisms involve exchange of genetic material between RNA genomes packaged within the same virus particle. Thus, a mosaic genome like that of the sabaeus virus can only be generated if two divergent lentiviruses replicate (and co-package) within
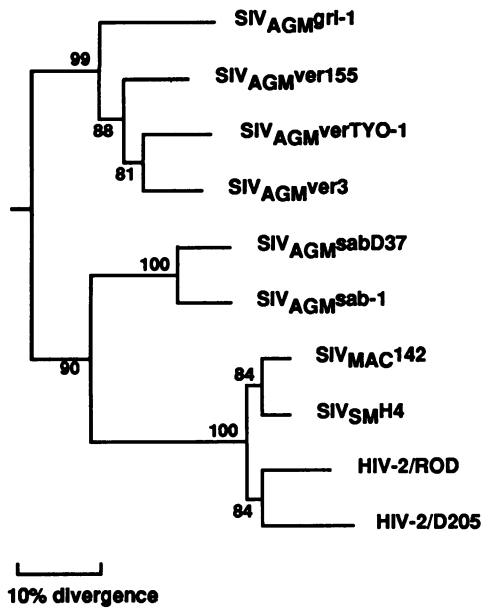
Fig. 8. Phylogenetic relationships of partial (N-terminal) *pol* protein sequences from a second sabaeus isolate, SIV_{AGM}sabD37. The alignment consensus length is 241 residues. Methodology as for Figure 7.
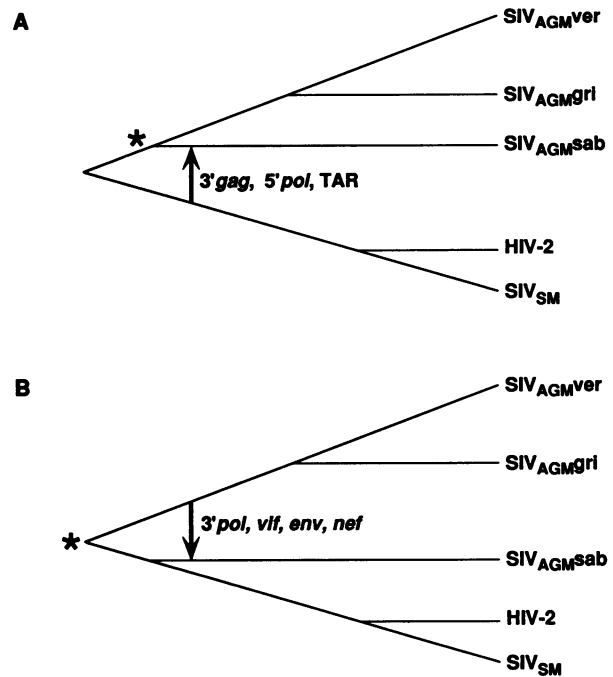

Fig. 9. Two possible scenarios for the origin of the mosaic genome of SIV_{AGM} from sabaeus monkeys. In each diagram, the * indicates the last common ancestor of the viruses from sabaeus, vervet and grivet monkeys, and the bold vertical arrow the direction of horizontal transfer via recombination (see text).

the same cell. In addition, both mechanisms can cause multiple template switches (Hu and Temin, 1992), which may provide an explanation for the rather complex structure of the sabaeus LTR which contains three SP1 sites located between one NFxB site and two TAR elements (Figure 4). This arrangment of regulatory elements may have resulted from multiple cross-overs during the recombination event, since most SIV_{AGM} isolates contain two NFxB sites, three SP1 sites and a single TAR element, whereas most SIV_{SM}/HIV-2 isolates contain a single NFxB site, four SP1 sites and a duplicated TAR region (Dewhurst *et al.*, 1990; Hirsch *et al.*, 1993b). Similarly, the unusual phylogenetic position of the sabaeus virus in the tree derived from the 5' *gag* region (Figure 7) may also be the result of a series of template switches. That is, this region may comprise a complex mosaic of sequence segments derived from both vervet/grivet and SIV_{SM}/HIV-2 lineages, which alternate but are too small (or too complex) to be detectable. Phylogenetic analysis of such a mosaic sequence would be expected to yield an apparently independent lineage as is seen in the 5' *gag* tree (Figure 7), and as was also seen in a tree based on the entire (chimeric) *pol* protein (discussed above).

Until now, recombination involving lentiviruses from different primate species has not been reported. Examples of mosaic genomes have been identified for both HIV-1 (the MAL isolate; Alizon *et al.*, 1986) and HIV-2 (the 7312A isolate; Gao *et al.*, 1992). However, these represent recombinants between virus strains within a single species, i.e. man (Li *et al.*, 1988; Gao *et al.*, 1992; Sharp *et al.*, 1994). Multiple cases of HIV-1/HIV-2 co-infections have been documented (Evans *et al.*, 1988; Rayfield *et al.*, 1988; Pieniazek *et al.*, 1991; George *et al.*, 1992; Peeters *et al.*, 1992b; Grez *et al.*, 1994), yet recombinants between these two viruses have not been observed. These findings, together with *in vitro* studies of HIV/SIV chimeras (reviewed in Shibata and Adachi, 1992), suggest that naturally occurring recombinants between members of major lentiviral lineages

are extremely rare, presumably because a mixture of highly divergent genes and proteins render progeny virions defective or replication impaired. In this case, it is not clear whether the recombinant sabaeus virus genome became fixed through random genetic drift, or whether it perhaps conferred some selective advantage (possibly influencing replication potential and/or transmissibility). It will thus be interesting to determine exactly how the chimeric *gag*−*pol* proteins in the sabaeus genome function, and how they interact with the other SIV_{AGM}-like viral components in virion assembly, maturation and entry processes.

The mosaic structure of the sabaeus genome also documents the apparent functionality of a duplicated (SIV_{SM}/HIV-2-like) TAR sequence in the presence of an SIV_{AGM}-like *tat* protein. Since *tat* is known to bind directly to TAR RNA during the transactivation process, this finding may be surprising at first. Recent data by Sakuragi *et al.* (1991), however, appear to provide an explanation. Comparing *tat* proteins from different lentiviral lineages for their ability to transactivate homologous and heterologous LTRs, these investigators found that SIV_{AGM} (vervet), HIV-2/SIV_{SM} and SIV_{MND} *tat* cross-transactivated SIV_{AGM} (vervet), SIV_{SM}/HIV-2 and SIV_{MND} LTRs with equal efficiency, although these same *tat* proteins failed to efficiently transactivate the HIV-1 LTR (reviewed in Sakai *et al.*, 1993). These data thus suggest that SIV_{AGM} and SIV_{SM} HIV-2 *tat* proteins are interchangeable with respect to their transactivation potential of SIV_{AGM} or SIV_{SM}/HIV-2 LTRs. These findings are also consistent with recent results by Tao and Frankel (1993), which indicate that the number of TAR hairpins does not determine the relative specificities of HIV and SIV *tat* proteins.

Which viruses and primate hosts were involved in the

putative recombination event that generated the mosaic sabaeus genome structure? In Figure 9, we outline two possible scenarios. In both cases we assume that the C.aethiops superspecies was ancestrally infected with SIV, and that the divergence among the SIV$_{AGM}$ sub-groups coincided with the speciation events within the superspecies (in both parts of Figure 9 the asterisk indicates the divergence of sabaeus viruses from the ancestor of vervet and grivet viruses). In the first scenario (Figure 9A), we assume that the SIV$_{SM}$/HIV-2 lineage branched off before the divergence of the sabaeus virus from vervet and grivet viruses, and thus that subsequent recombination brought the SIV$_{SM}$/HIV-2-like regions (3'gag, 5'pol and TAR) into the sabaeus lineage. This suggests that a sabaeus monkey was simultaneously co-infected with SIV$_{AGM}$ and an ancestor of SIV$_{SM}$ acquired by simian−simian cross-species transmission. Whether the donor was a sooty mangabey or another as yet unidentified African primate is not known. Alternatively (Figure 9B), the divergence of the sabaeus virus from the other African green monkey viruses may represent the most ancient point in the tree. Under this scenario, an ancestral sabaeus virus underwent recombination with a SIV$_{AGM}$ ancestral to the vervet and grivet lineages, receiving sequences covering approximately the 3' half of the genome. This scenario involves transmission among different species of African green monkeys to produce the individual co-infected monkey. It also suggests that cross-species transmission occurred separately from sabaeus monkeys to sooty mangabeys (or again some unidentified intermediate) to generate the SIV$_{SM}$ lineage. Whatever the correct scenario, it is important to emphasize that sabaeus monkeys and sooty mangabeys are both found in West Africa and that they reside in overlapping natural habitats (P.Marx, personal communication). Also, this cross-species transmission event must have occurred a very long time ago (compare branch lengths in Figures 7 and 8), thus providing evidence for an ancient recombination event.

The mosaic sabaeus genome reported here, together with the recent characterization of SIV from Sykes' monkeys (Hirsch et al., 1993b), have implications for the origins of the entire family of primate lentiviruses. Before these reports, four major lineages of primate lentiviruses were known, each represented by SIVs from a different genus of Old World primates (consider Figure 1 without the SIV$_{SYK}$ lineage). Such phylogenetic relationships could have been the result of a very old viral−host relationship, if the most ancestral point in the tree reflected the (relatively) ancient divergence of the different monkey genera. However, the identification of a fifth lineage of viruses also involving the genus Cercopithecus (i.e. SIV$_{SYK}$ from C.mitis), as well as the identification of a highly divergent SIV$_{AGM}$ with evolutionary links to another major lineage (SIV$_{SM}$), implicate Cercopithecus monkeys in a more central role in the evolutionary origins of the known primate lentiviruses. This conclusion is further supported by the finding that the majority of Old World monkey species known to harbor an SIV-like virus belong to the genus Cercopithecus (reviewed in Johnson et al., 1990b). Moreover, yellow baboons (Papio hamadryas cynocephalus) which do not belong to the Cercopithecus genus but live in the same habitat as vervet monkeys have been found to occasionally harbor antibodies that cross-react with SIV$_{AGM}$ antigens, leading to the hypothesis that these animals are also infected with an

SIV$_{AGM}$-like virus (Kodama et al., 1989a). In light of these findings, it is conceivable that the entire radiation of primate lentiviruses arose in Cercopithecus monkeys, followed by extensive cross-species transmission to other primates. Perhaps even the HIV-1 lineage, so far known only in humans and chimpanzees but not obviously naturally widespread in either of those hosts, originated in some (as yet unidentified) Cercopithecus species (reviewed in Sharp et al., 1994). Elucidation of the phylogenetic relationships of all naturally occurring SIVs will be necessary to evaluate this possibility.

Materials and methods

SIV$_{AGM}$ isolates

The designation and geographic origin of all SIV$_{AGM}$ isolates analyzed in this study are listed in Table I, along with selected references describing their molecular or biological characterization. SIV$_{AGM}$90, SIV$_{AGM}$155 and SIV$_{AGM}$3 were derived from vervet monkeys (V.Hirsch and R.Kurth, personal communication). The African green monkey hosts of each of the other SIV$_{AGM}$ isolates have been published (Allan et al., 1990; Sakuragi et al., 1991; Hirsch et al., 1993a; Müller et al., 1993) and are indicated in the respective strain names. Isolates ver-1 and gri-1 have also been referred to as ver692 and gri677 (Johnson et al., 1990b; Myers et al., 1993). Viruses for which sequence information was obtained from EMBL/GenBank are indicated. The SIV$_{AGM}$tan-1 isolate was obtained from a female offspring of a group of tantalus monkeys imported from Uganda in 1979. Virus was isolated in 1989 by cocultivation of PHA-stimulated peripheral blood mononuclear cells with the immortalized human T-cell line Molt4 (clone 8) as described (Allan et al., 1990).

Polymerase chain reaction

For PCR analysis, high molecular weight DNA was extracted from T-cell lines productively infected with SIV$_{AGM}$ as described (Li et al., 1991). PCRs were carried out in a total volume of 100 μl, containing 1 μg of genomic DNA, 10 mM Tris−HCl (pH 8.3), 50 mM KCl, 20 pmol of each primer, 200 μM of each of the four dNTPs, 1.5 mM MgCl$_2$, 0.01% w/v gelatin and 2.5 U Taq polymerase. Samples were over-layered with 100 μl of mineral oil to avoid evaporation and then subjected to 40 amplification cycles consisting of a denaturing step (94°C, 1 min), a primer-annealing step (55°C, 1 min) and a primer extension step (72°C, 1 min). Oligonucleotide primers were designed according to HIV/SIV consensus sequences (Myers et al., 1993), except for LTR primers 406 and 407 which were published previously (Johnson et al., 1990b). Primer designations, nucleotide sequences, genomic location, and amplification specificities are listed in Table II. PCR products were analyzed by agarose gel electrophoresis, purified using Geneclean and subsequently sub-cloned into M13mp18 utilizing restriction enzyme sites or into pCRII by T/A overhang (Invitrogen, San Diego, CA).

Lambda phage cloning of SIV$_{AGM}$sab-1

To obtain a full-length sabaeus provirus, a genomic lambda phage library was constructed by methods previously described (Maniatis et al., 1982). Total genomic DNA of Molt4 (clone 8) cells persistently infected with the SIV$_{AGM}$sab-1 isolate (see Table I) was partially digested with MboI, fractionated by sucrose gradient centrifugation to obtain DNA fragments of 9−20 kb length, and ligated into BamHI cleaved arms of λGem12 (Promega, Madison, WI). Ligation products were then packaged in vitro, titered and plated on bacteria (KW251). A total of 1.5×10⁶ phage plaques were screened with a radiolabeled LTR probe specific for sab-1, and four hybridization positive phage clones were identified. All positive phage recombinants were plaque purified and their restriction enzyme maps determined. This analysis identified one clone (λMJ8) that contained a full-length proviral genome which was subsequently shown to be replication competent by transfection into Cos-1 cells. For sequence analysis, sub-genomic fragments were generated by digestion of λMJ8 with BamHI and HindIII, respectively, followed by shotgun sub-cloning into M13mp18/19. The complete sabaeus provirus including flanking cellular sequences (a 18 kb NotI fragment) was also sub-cloned into pBluescript KSII (generating psab-1) to facilitate subsequent transfection studies.

DNA transfection and viral infectivity studies

Biological characterization of the sabaeus (MJ8) proviral clone was performed by comparison with two biologically active SIV$_{AGM}$ reference clones

Table II. PCR primer pairs used for amplification of SIV$_{AGM}$ sub-genomic fragments

Primer pair	Nucleotide sequence	Location	Specificity	Fragment size (bp)	Genomic region
LTRA	5'-gaggaattccTGGATGGGATTTATTACTCCGAAAG-3'	8547–8571	sab-1, sab-2	771–821	LTR
LTRB	5'-gaggaattccACTCAAGTCCCTGTTCGGGCGCCA-3'	220–243	sab-3, sab-4 ver-2		
LTR406	5'-atgcgagctcTGGATGGGATTTATTACTCC-3'	8547–8566	gri-2, gri-3	763–766	LTR
LTR407	5'-aagtgaattccACTCAAGTCCCTGTTCGGG-3'	225–243			
LTRA*	5'-TTTTTAAAAGAAAAGGGGGGACTGG-3'	8525–8543	tan-1	762	LTR
LTR407	5'aagtgaattccACTCAAGTCCCTGTTCGGG-3'	225–243			
envA	5'-GAAGCTTGTGATAAAACATATTGGGAT-3'	6415–6441	sab-1, sab-2	1152–1188	*env*
envB	5'-AGAGCTGTGACGCGGGCATTGAGG-3'	7566–7589	sab-3, sab-4 ver-1, ver-2 gri-2, gri-3 tan-1, tan17 tan40, tan49		
polA	5'-GCTCTATTAGACACAGGAGCAGATGAC-3'	2024–2050	sabD37	705	*pol*
polB	5'-GGGTGAGCCTTTCCACCCCTGTGGGAG-3'	2714–2740			

Primer pairs were designed according to HIV/SIV consensus sequences (Meyers *et al.*, 1993). Sequences are numbered according to the published sequence of TYO-1 (Fukasawa *et al.*, 1988). Bases in upper case represent SIV$_{AGM}$-specific sequences; bases in lower case were added to facilitate sub-cloning (restriction enzyme sites are in bold); LTR 406/407 primers were published previously (Johnson *et al.*, 1990b).

representing vervet (SA-212; Shibata *et al.*, 1990) and grivet (gri-1/ΛII, Fomsgaard *et al.*, 1991) viruses. Viral stocks were prepared by transfection of 10 μg of plasmid DNA into Cos-1 cells (2×10^5) using the Stratagene DNA transfection kit (La Jolla, CA). Twenty-four hours post-transfection, 2×10^6 Molt4 (clone8) cells were added to the Cos-1 monolayer and co-cultivated for an additional 48 h. Non-adherent cells were then transferred to a new flask and monitored for RT activity. To compare viral replication characteristics, culture supernatants of transfected Cos-1 cells were normalized to contain equal numbers of virions (660 000 RT units), filtered through a 0.2 μm filter and used to infect 2×10^6 Molt4 (clone8) cells in a volume of 3.5 ml. Following a 24 h incubation, cultures were washed extensively with Hanks' balanced salt solution, and supernatants were collected in 3 day intervals to test RT activity (Li *et al.*, 1991). The appearance of multi-nucleated giant cells was monitored on a daily basis.

DNA sequence analysis
All PCR derived SIV$_{AGM}$ fragments (*env*, *pol* and LTR) as well as the complete sab-1 (MJ8) provirus were sequenced using the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). Both single-stranded (M13 based sub-clones) and double-stranded DNA templates (pCRII based sub-clones) were sequenced using universal as well as SIV$_{AGM}$-specific oligonucleotide primers. Generally only one clone per amplification product was sequenced, except for sabaeus LTRs for which two clones were sequenced. Nucleotide sequence analysis was performed using PC/Gene (Intelligenetics), the University of Wisconsin Genetics Computer Group (UWGCG) software package and MASE (Faulkner and Jurka, 1988).

Phylogenetic analysis
Phylogenetic relationships were estimated from comparisons of predicted protein sequences. Sequences were aligned using CLUSTAL (Higgins and Sharp, 1988, 1989). Evolutionary distances between all pairs of sequences were computed using Kimura's empirical method (Kimura, 1983; eqn. 4.8) to estimate the number of superimposed amino acid replacements; sites at which there was a gap in any sequence in the alignment were excluded from all comparisons. Phylogenetic relationships were estimated from these distances by the neighbor-joining method (Saitou and Nei, 1987). The reliability of branching orders was estimated by the bootstrap approach (Felsenstein, 1985). These methods were implemented using CLUSTAL V (Higgins *et al.*, 1992).

Nucleotide accession numbers
All sequences were submitted to GenBank and are available under accession numbers U03994–U04018.

References

Alizon,M., Wain-Hobson,S., Montagnier,L. and Sonigo,P. (1986) *Cell*, **46**, 63–74.
Allan,J.S. (1993) *AIDS*, **7** (suppl. 1), S43–S50.
Allan,J.S., Kanda,P., Kennedy,R.C., Cobb,E.K., Anthony,M. and Eichberg,J.W. (1990) *AIDS Res. Hum. Retrovir.*, **6**, 275–285.
Allan,J.S., Short,M., Taylor,M.E., Su,S., Hirsch,V.M., Johnson,P.R., Shaw,G.M. and Hahn,B.H. (1991) *J. Virol.*, **65**, 2816–2828.
Baier,M., Werner,A., Cichutek,K., Garber,C., Müller,C., Kraus,G., Ferdinand,F.J., Hartung,S., Papas,T.S. and Kurth,R. (1989) *J. Virol.*, **63**, 5119–5123.
Baskin,G.B., Murphey-Corb,M., Watson,E.A. and Martin,L.M. (1988) *Vet. Pathol.*, **25**, 456–467.
Berkhout,B. (1992) *Nucleic Acids Res.*, **20**, 27–31.
Cichutek,K. and Norley,S. (1993) *AIDS*, **7** (suppl. 1), S25–S35.
Coffin,J. (1979) *J. Gen. Virol.*, **42**, 1–26.
Coffin,J.M. (1992) *Curr. Top. Microbiol. Immunol.*, **176**, 143–164.
Cullen,B.R. (1992) *Microbiol. Rev.*, **56**, 375–394.
Daniel,M.D., Letvin,N.L., King,N.W., Kannagi,M., Sehgal,P.K., Hunt,R.D., Kanki,P.J. and Essex,M. (1985) *Science*, **228**, 1201–1204.
Dewhurst,S., Embretson,J.E., Anderson,D.C., Mullins,J.I. and Fultz,P.N. (1990) *Nature*, **345**, 636–640.
Evans,L.A., Odehouri,K., Thomson-Honnebier,G., Barboza,A., Moreau,J.,

Seto,D., Legg,H., Cheng-Mayer,C. and Levy,J.A. (1988) *Lancet*, 2, 1389–1391.

Faulkner,D.M. and Jurka,J. (1988) *Trends Biochem. Sci.*, 13, 321–322.

Felsenstein,J. (1985) *Evolution*, 39, 783–791.

Fenrick,R., Malim,M.H., Hauber,J., Le,S.-Y., Maizel,J. and Cullen,B.R. (1989) *J. Virol.*, 63, 5006–5012.

Fomsgaard,A., Hirsch,V.M., Allan,J.S. and Johnson,P.R. (1991) *Virology*, 182, 397–402.

Fukasawa,M., Miura,T., Hasegawa,A., Masegawa,A., Morikawa,S., Tsujimoto,H., Miki,K., Kitamura,T. and Hayami,M. (1988) *Nature*, 333, 457–461.

Fultz,P.N., McClure,H.M., Anderson,D.C. and Switzer,W.M. (1989) *AIDS Res. Hum. Retrovir.*, 5, 397–409.

Gao,F., Yue,L., White,A.T., Pappas,P.G., Barchue,J., Hanson,A.P., Greene,B.M., Sharp,P.M., Shaw,G.M. and Hahn,B.H. (1992) *Nature*, 358, 495–499.

Gaynor,R. (1992) *AIDS*, 6, 347–363.

George,J.R., Ou,C.Y., Parekh,B., Brattegaard,K., Brown,V., Boateng,E. and de Cock,K.M. (1992) *Lancet*, 340, 337–339.

Grez,M., Dietrich,U., Balfe,P., von Briesen,H., Maniar,J.K., Mahambre,G., Delwart,E.L., Mullins,J.I. and Rübsamen-Waigmann,H. (1994) *J. Virol.*, 68, 2161–2168.

Guyader,M., Emerman,M., Sonigo,P., Clavel,F., Montagnier,L. and Alizon,M. (1987) *Nature*, 326, 662–669.

Hendry,R.M., Wells,M.A., Phelan,M.A., Schneider,A.L., Epstein,J.S. and Quinnan,G.V. (1986) *Lancet*, ii, 1957–1962.

Higgins,D.G. and Sharp,P.M. (1988) *Gene*, 73, 237–244.

Higgins,D.G. and Sharp,P.M. (1989) *Comp. Applic. Biosci.*, 5, 151–153.

Higgins,D.G., Bleasby,A.J. and Fuchs,R. (1992) *Comp. Applic. Biosci.*, 8, 189–191.

Hirsch,V.M. and Johnson,P.R. (1994) *Virus Res.*, in press.

Hirsch,V.M., Olmsted,R.A., Murphey-Corb,M., Purcell,R.H. and Johnson,P.R. (1989a) *Nature*, 339, 389–391.

Hirsch,V.M., Edmondson,P., Murphey-Corb,M., Arbeille,B., Johnson,P.R. and Mullins,J.I. (1989b) *Nature*, 341, 573–574.

Hirsch,V.M., McGann,C., Dapolito,G., Goldstein,S., Ogen-Odoi,A., Biryawaho,B., Lakwo,T. and Johnson,P.R. (1993a) *Virology*, 197, 426–430.

Hirsch,V.M., Dapolito,G.A., Goldstein,S., McClure,H., Emau,P., Fultz,P.N., Isahakia,M., Lenroot,R., Myers,G. and Johnson,P.R. (1993b) *J. Virol.*, 67, 1517–1528.

Hu,W.-S. and Temin,H.M. (1990) *Science*, 250, 1227–1233.

Hu,W.-S. and Temin,H.M. (1992) *J. Virol.*, 66, 4457–4463.

Huet,T., Cheynier,R., Meyerhans,A., Roelants,G. and Wain-Hobson,S. (1990) *Nature*, 345, 356–359.

Johnson,P.R. and Hirsch,V.M. (1991) In Koff,W. (ed.), *Annual Review of AIDS Research*. Marcel Dekker, New York, Vol. I, pp. 47–62.

Johnson,P.R., Goldstein,S., London,W.T., Fomsgaard,A. and Hirsch,V.M. (1990a) *J. Med. Primatol.*, 19, 279–286.

Johnson,P.R., Fomsgaard,A., Allan,J., Gravell,M., London,W.T., Olmsted,R.A. and Hirsch,V.M. (1990b) *J. Virol.*, 64, 1086–1092.

Junghans,R.P., Boone,L.R. and Skalka,A.M. (1982) *Cell*, 30, 53–62.

Khan,A.S., Galvin,T.A., Lowenstine,L.J., Jennings,M.B., Gardner,M.B. and Buckler,C.E. (1991) *J. Virol.*, 65, 7061–7065.

Kimura,M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Kodama,T., Silva,D.P., Daniel,M.D., Phillips-Conroy,J.E., Jolly,C.J., Rogers,J. and Desrosiers,R.C. (1989a) *AIDS Res. Hum. Retrovir.*, 5, 337–343.

Kodama,T., Wooley,D.P., Naidu,Y.M., Kestler,H.W., III, Daniel,M.D., Li,Y. and Desrosiers,R.C. (1989b) *J. Virol.*, 63, 4709–4714.

Kurth,R. and Norley,S. (1994) *Curr. Top. Microbiol. Immunol.*, in press.

Lernould,J.M. (1988) In Gautier-Hion,A., Bourliere,F. and Gautier,J.-P. (eds), *A Primate Radiation: Evolutionary Biology of African Guenons*. Cambridge University Press, Cambridge, pp. 54–78.

Li,W.-H., Tanimura,M. and Sharp,P.M. (1988) *Mol. Biol. Evol.*, 5, 313–330.

Li,Y., Naidu,Y.M., Daniel,M.D. and Desrosiers,R.C. (1989) *J. Virol.*, 63, 1800–1802.

Li,Y., Kappes,J.C., Conway,J.A., Price,R.W., Shaw,G.M. and Hahn,B.H. (1991) *J. Virol.*, 65, 3973–3985.

Lowenstine,L.J., Pederson,N.C. and Higgins,J. (1986) *Int. J. Cancer*, 38, 563–573.

Lowenstine,L.J., Lerche,N.W., Marx,P.A., Gardner,M.B. and Pedersen,N.C. (1988) In Girard,M. and Valette,L. (eds), *Retroviruses of Human AIDS and Related Animal Viruses*. Pasteur Vaccines, Paris, pp. 174–176.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 269–295.

Marx,P.A. *et al.* (1991) *J. Virol.*, 65, 4480–4485.

McClure,H.M., Anderson,D.C., Fultz,P.N., Ansari,A.A., Lockwood,A.E. and Brodie,A. (1989) *Vet. Immunol. Immunopathol.*, 21, 13–24.

McKeating,J.A. and Wiley,R.L. (1989) *AIDS*, 3 (suppl. 1), S35–S41.

Müller,M.C. *et al.* (1993) *J. Virol.*, 67, 1227–1235.

Murphey-Corb,M., Martin,L.N., Rangan,S.R.S., Baskin,G.B., Gormus,B.J., Wolf,R.H., Andes,W.A., West,M. and Montelaro,R.C. (1986) *Nature*, 321, 435–437.

Myers,G., MacInnes,K. and Korber,B. (1992) *AIDS Res. Hum. Retrovir.*, 8, 373–386.

Myers,G., Korber,B., Berzofsky,J.A. and Smith,R.F. (1993) *Human Retroviruses and AIDS 1993*. Theoretical Biology and Biophysics, Los Alamos, NM.

Novembre,F.J., Hirsch,V.M., McClure,H.M., Fultz,P.N. and Johnson,P.R. (1992) *Virology*, 186, 783–787.

Ohta,Y., Masuda,T., Tsujimoto,H., Ishikawa,K., Kodama,T., Morikawa,S., Nakai,M., Honjo,S. and Hayami,M. (1988) *Int. J. Cancer*, 41, 115–122.

Peeters,M., Fransen,K., Delaporte,E., Van den Haesevelde,M., Gershy-Damet,G.-M., Kestens,L., van der Groen,G. and Piot,P. (1992a) *AIDS*, 6, 447–451.

Peeters,M., Gershy-Damet,G.M., Fransen,K., Koffi,K., Coulibaly,M., Delaport,E., Piot,P. and van der Groen,G. (1992b) *Lancet*, 340, 339–340.

Pieniazek,D. *et al.* (1991) *AIDS*, 5, 1293–1299.

Rayfield,M. *et al.* (1988) *J. Infect. Dis.*, 158, 1170–1176.

Saitou,N. and Nei,M. (1987) *Mol. Biol. Evol.*, 4, 406–425.

Sakai,H., Sakuragi, J-i., Sakuragi,S., Kawamura,M. and Adachi,A. (1993) *Arch. Virol.*, 129, 1–10.

Sakuragi,J.-I., Fukasawa,M., Shibata,R., Sakai,H., Kawamura,M., Akari,H., Kiyomasu,T., Ishimoto,A., Hayami,M. and Adachi,A. (1991) *Virology*, 185, 455–459.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, 74, 5463–5467.

Sharp,P.M., Robertson,D.L., Gao,F. and Hahn,B.H. (1994) *AIDS* 1994 (suppl. 1), S27–S42.

Shibata,R. and Adachi,A. (1992) *AIDS Res. Hum. Retrovir.*, 8, 403–409.

Shibata,R., Miura,T., Hayami,M., Sakai,H., Ogawa,K., Kiyomasu,T., Ishimoto,A. and Adachi,A. (1990) *J. Virol.*, 64, 307–312.

Tao,J. and Frankel,A.D. (1993) *Proc. Natl Acad. Sci. USA*, 90, 1571–1575.

Tsujimoto,H., Hasegawa,A., Maki,N., Fukasawa,M., Miura,T., Speidel,S., Cooper,R.W., Moriyama,E.N., Gojobori,T. and Hayami,M. (1989) *Nature*, 341, 539–541.