J Mol Evol (1995) 40:249–259

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 1995

Recombination in AIDS Viruses

David L. Robertson,[1] Beatrice H. Hahn,[2] Paul M. Sharp[1]

[1] Department of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, United Kingdom
[2] Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Received: 15 July 1994 / Accepted: 27 September 1994

Abstract. Recombination contributes to the generation of genetic diversity in human immunodeficiency viruses (HIV) but can only occur between viruses replicating within the same cell. Since individuals have not been found to be simultaneously coinfected with multiple divergent strains of HIV-1 or HIV-2, recombination events have been thought to be restricted to the rather closely related members of the quasispecies that evolves during the course of HIV infection. Here we describe examples of both HIV-1 and HIV-2 genomes that appear to be hybrids of genetically quite divergent viruses. Phylogenetic analyses were used to examine the evolutionary relationships among multiple HIV strains. Evolutionary trees derived from different genomic regions were consistent with respect to most of the viruses investigated. However, some strains of HIV-1 and HIV-2 exhibited significantly discordant branching orders indicative of genetic exchanges during their evolutionary histories. The crossover points of these putative recombination events were mapped by examining the distribution of phylogenetically informative sites supporting alternative tree topologies. A similar example of a recombinant simian immunodeficiency virus identified in West African green monkeys has also been described recently. These results indicate that coinfection with highly divergent viral strains can occur in HIV-infected humans and SIV-infected primates and could lead to the generation of hybrid genomes with significantly altered biological properties. Thus, future characterization of primate lentiviruses should include careful phylogenetic investigation of possible genomic mosaicism.

Key words: AIDS — Human immunodeficiency virus — Simian immunodeficiency virus — Phylogenetics — Recombination — Superinfection

Introduction

Acquired immune deficiency syndrome (AIDS) is caused by two different human immunodeficiency viruses, HIV-1 and HIV-2, which are members of the lentivirus family of retroviruses. A growing number of related viruses have been discovered in nonhuman primates: these have been termed simian immunodeficiency viruses (SIV), although this name is somewhat misleading since none have yet been shown to cause immunodeficiency in their natural hosts. Both HIV and SIV have RNA genomes with features typical of retroviruses—i.e., they are about 10 kb in length, with much of this sequence taken up by *gag, pol,* and *env* genes, plus long terminal repeat regions (LTRs) which have regulatory functions. In addition, HIVs and SIVs encode a number (five or six) of shorter accessory genes (two of which, *rev* and *tat,* are each encoded by two separate exons) which are unique to lentiviruses and are believed to be responsible for these viruses' ability to maintain a chronic persistent infection even in the presence of a pronounced host immune response. (See Hahn 1994, for a recent review of the genetics of HIV.)

HIVs and SIVs exhibit extraordinary genetic diversity. This is due to their very high rate of nucleotide

Correspondence to: P.M. Sharp

sequence evolution (Coffin 1986; Hahn et al. 1986), which in turn appears to be generated by the high error rate of the viral (*pol*-encoded) reverse transcriptase (Preston et al. 1988; Roberts et al. 1988; Weber and Grosse 1989; Richetti and Buc 1990). In fact, on the microevolutionary scale, HIV and SIV each exist as "quasispecies," i.e., a population of highly related yet genetically distinct viruses that coexist within the same individual (Wain-Hobson 1993).

A second means by which genetic diversity is generated is recombination. Retroviruses are known to be highly recombinogenic (Goodrich and Duesberg 1990; Hu and Temin 1990; Zhang and Temin 1994), apparently as a consequence of their dimeric genome and a reverse transcriptase that can switch between templates during proviral DNA synthesis. Two models have been proposed to explain this process (Coffin 1979; Junghans et al. 1982), both of which involve exchange of genetic material between viral genomes copackaged within the same particle. Thus, hybrid genomes can only be generated from viruses that replicate (and copackage) within the same cell. Such recombination has been reported for closely related HIV-1 variants, i.e., quasispecies members infecting single individuals (Howell et al. 1991; Delassus et al. 1991; Vartanian et al. 1991; Groenink et al. 1992). Clearly, recombination could generate much more different viral genomes if an individual host was simultaneously coinfected with multiple divergent viral strains. However, such superinfection has not yet been documented. In fact, in those cases where genetic diversity was analysed following acute seroconversion, it was found that individuals were initially infected with only a single virus, which then evolved into a quasispecies over time (Wolfs et al. 1992; Zhang et al. 1993; Zhu et al. 1993).

Nevertheless, there have been two cases in which sequence comparisons have yielded results suggesting that recombination among divergent HIV strains can occur. First, in one of the earliest phylogenetic studies of diverse HIV-1 isolates, Li et al. (1988) found that one isolate (MAL) was relatively less divergent from another Zairean isolate (ELI) in the *env* gene than in *gag* and *pol*, as if these regions of the genome had different evolutionary histories. However, the sequences available for analysis at that time were quite limited, both in number and in length, and so it was not possible to define the putative recombination event in any detail. More recently, Gao et al. (1992) sequenced *pol* and *env* gene fragments from three newly identified HIV-2 strains and examined their phylogenetic relationships to other isolates. One of these new viruses (7312A) fell into two quite different clades within the HIV-2 radiation, depending on the gene examined. Again, the sequence data available were not sufficient to characterize the putative recombination event.

Here we utilize more recently determined sequences

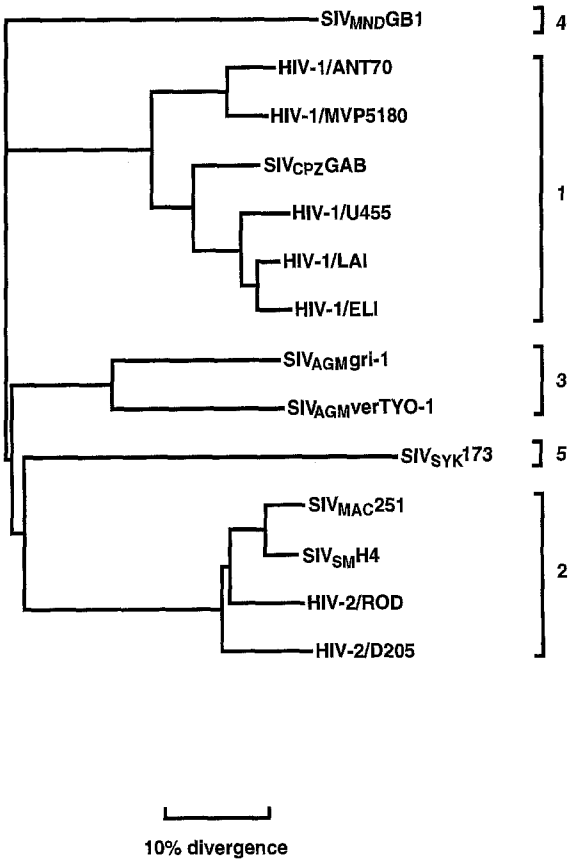


Fig. 1. Phylogenetic relationships of representative primate lentiviruses, derived from *pol* protein sequences. *Numbered brackets* at the right indicate the five major lineages. Horizontal branch lengths are drawn to scale; the *bar* indicates 0.10 amino acid replacements per site. The approximate position of the root of the tree (at the left) was determined from analyses using nonprimate lentiviruses as outgroups (Sharp et al. 1994). The precise order of branching of the five major lineages (near the root) is unclear, but bootstrap values for all other nodes (with the exception of the branching order of HIV-2_{D205} and HIV-2_{ROD}) are in the range 99–100%. Accession numbers for the sequences used are given in Appendix 1.

to demonstrate that MAL and 7312A do indeed represent hybrid viruses, and we define the presumed crossover points of the recombination events. The implications of these findings are discussed, both in terms of HIV biology and of the analysis of evolutionary relationships among these viruses.

Phylogeny of Primate Lentiviruses

First, it is necessary to review briefly the current state of knowledge concerning the phylogenetic relationships among primate lentiviruses, to provide the context in which recombinant viral genomes can be identified. (See Sharp et al. 1994 for more details.) The characterized primate lentiviruses form five distinct, approximately equidistant, lineages (Fig. 1). One lineage contains viruses isolated from humans (HIV-1) and the common

chimpanzee, *Pan troglodytes* (SIV_{CPZ}). A second includes the other human virus (HIV-2), as well as viruses from sooty mangabeys, *Cercocebus atys* (SIV_{SM}), and a number of macaques of the genus *Macaca* (SIV_{MAC} and related viruses). A third contains viruses isolated from the four different species of African green monkeys, *Cercopithecus aethiops*. The two remaining lineages are known from single isolates from the mandrill, *Mandrillus sphinx* (SIV_{MND}), and Sykes' monkey, *Cercopithecus mitis* (SIV_{SYK}).

Within each of the three lineages represented by multiple viruses there is considerable genetic diversity. For example, there are two distinct groups of HIV-1. All of the earliest-known HIV-1 isolates fall into one of these clusters, now called group M (represented by LAI, ELI, and U455 in Fig. 1). Phylogenetic analyses of viruses collected from all over the world have revealed at least eight subgroups, termed sequence subtypes A–H, within this cluster (Myers et al. 1993; Louwagie et al. 1993; Sharp et al. 1994). These eight subtypes are approximately equidistantly related, although subtypes B and D seem a little closer to each other than to the other subtypes; U455, LAI, and ELI represent subtypes A, B, and D, respectively. The second major cluster, called group O (represented by ANT70 and MVP5180 in Fig. 1), has been identified only recently (Vanden Haesevelde et al. 1994; Gurtler et al. 1994) and is known only from a small number of isolates originating from Cameroon. Importantly, an isolate from a chimpanzee (SIV_{CPZ}GAB) is more closely related to HIV-1 group M than is HIV-1 group O (Fig. 1). It seems clear that HIV-1 has arisen through cross-species transmissions of primate lentiviruses, but because only few SIV-infected chimpanzees have been found in the wild (Peeters et al. 1992) it is not yet known whether chimpanzees are the primary reservoir for HIV-1 or whether both humans and chimps have become infected from a third, as yet unidentified, species (Johnson et al. 1991; Sharp et al. 1994). Note that, under either scenario, humans must have been infected on at least two separate occasions.

Multiple (so far five) subtypes of HIV-2 have also been defined (Gao et al. 1994), but as yet only subtypes A and B (represented in Fig. 1 by ROD and D205, respectively) comprise more than one isolate. Again, the primary host of the viruses within this lineage appears to be a nonhuman primate, in this case the sooty mangabey (Hirsch et al. 1989; Myers et al. 1992; Gao et al. 1994). HIV-2 is common only in West Africa (the natural habitat of the sooty mangabey), and macaques have only been found to be infected with SIV in captivity. Moreover, phylogenetic analyses suggest that cross-species transmissions, both from mangabeys to humans in West Africa and from mangabeys to macaques in captivity, have occurred on multiple occasions (Gao et al. 1992, 1994).

Finally, within the SIV_{AGM} lineage, viruses isolated from vervet, grivet, sabaeus, and tantalus monkeys (the four species of *C. aethiops*) form species-specific clusters (only the vervet and grivet groups are represented in Fig. 1; the other two groups are approximately equidistant; Jin et al. 1994). The four species inhabit different, though partially overlapping, ranges covering most of sub-Saharan African. The host-dependent pattern of evolutionary relationships among their lentiviruses suggests that cross-species transmissions have been rare. This in turn implies that infection of African green monkeys by SIV may be a relatively ancient event.

Sequence Analyses

Recombination events can be identified most clearly in the context of phylogenetic analyses. Therefore, as a first step, we obtained multiple alignments of protein sequences using the CLUSTAL algorithm (Higgins and Sharp 1988, 1989). DNA sequence alignments were derived from the alignments of their products. For comparison of more divergent viruses, distances between protein sequences were calculated as the number of amino acid replacements per site, estimated by the empirical method of Kimura (equation 4.8 in Kimura 1983). For more closely related viruses, distances among DNA sequences were calculated as the number of nucleotide substitutions per site by the two-parameter method of Kimura, which allows for different rates of transitions and transversions (equation 4.14 in Kimura 1983). All distances were computed excluding any site which had a gap in any of the aligned sequences. Unrooted phylogenetic trees were derived using the neighbor-joining method (Saitou and Nei 1987) applied to matrices of these estimates of pairwise distances among viruses. The stability of the phylogenetic trees obtained was tested using the bootstrap approach (Felsenstein 1985); 1,000 replicates were used. These methods were implemented using CLUSTAL V (Higgins et al. 1992).

We have also used maximum parsimony and dynamically weighted parsimony methods (Williams and Fitch 1990) to examine the phylogenetic relationships among the same viruses. Evolutionary trees obtained by these methods have not been found to differ in any significant features from those calculated by the neighbor-joining method.

The existence of a mosaic viral genome, i.e., one that has been generated by a recombination event at some point in the past, is indicated when phylogenetic relationships for different parts of the genome are discordant. Initially, we derived evolutionary trees from complete gene (or protein) sequences. Obviously, however, recombination need not necessarily occur in intergenic regions. Therefore, we went on to estimate phylogenies for partial gene sequences, searching for discrepancies between trees based on different parts of genes.

Finally, to localize more precisely the crossover points we inspected the distribution of phylogenetically informative sites supporting alternative tree topologies. This is most simply performed by considering four sequences at a time. These should be the putative recombinant sequence, one representative of each of the two "parental" lineages thought to be involved in the recombination event, and a known outgroup. In a four-sequence alignment, an informative site is one at which two sequences share one nucleotide (or amino acid), and the other two share another. There are three possible configurations of such informative sites, two of which support the clustering of the putative recombinant with one parental sequence or the other. The distribution of these two types of sites can be tested by asking whether a break placed at any point along the alignment produces a significant difference in the ratio of the two types on each side of that cut, as assessed by a chi-square value; the optimum position of the breakpoint can be found by maximizing this value. (See, e.g., Maynard Smith 1992.)

Recombination in HIV-1

The MAL strain of HIV-1 was one of the first of African origin to be characterized; MAL was isolated from a boy with AIDS-related complex in Zaire who was thought to have been infected by a blood transfusion (Alizon et al. 1986). Li et al. (1988) performed a phylogenetic analysis of *env* gene sequences from the relatively limited number of American and African isolates of HIV-1 available at that time. MAL was found to be more closely related to other African isolates (ELI and Z6, members of what is now known as sequence subtype D) than to the American isolates (members of subtype B). However, it was also noted that when other gene sequences were used, MAL was more divergent, and appeared to be an outgroup to a cluster including ELI, Z6, and the American isolates (Li et al. 1988). Since then, the sequences of a much larger number of HIV-1 isolates have been determined. Phylogenetic analyses of *env* sequences have identified six sequence subtypes of HIV-1 with MAL falling in subtype D (Myers et al. 1993; Sharp et al. 1994). Parallel analyses of *gag* sequences have identified eight sequence subtypes, with MAL clustering with subtype A viruses (Louwagie et al. 1993; Myers et al. 1993; Sharp et al. 1994). Here we analyze this discrepancy in more detail and attempt to locate the points at which the transitions between subtype A–like and subtype D–like sequences occur.

We have investigated the phylogenetic relationship of MAL to representative members of sequence subtypes A (U455), B (OYI and LAI), and D (NDK and ELI), using $SIV_{CPZ}GAB$ as an outgroup. Subtype B sequences were included because they appear to be more closely related to subtype D than are other subtypes. Analyses of *gag*, *pol*, and *nef* genes showed MAL to be most closely related to U455. Analysis of *env* gene sequences indicated that this part of the MAL genome was more closely related to the subtype D viruses NDK and ELI, and thus confirmed that different regions of the genome had different phylogenetic histories, i.e., that MAL, or an ancestor of MAL, had resulted from recombination.

These findings suggested that there had been a recombination event with crossover points in the regions between *pol* and *env* (a distance of approximately 1,100 nucleotides) and around the junction of *env* and *nef* (which are just one base apart). The region between *pol* and *env* encodes five small accessory genes (Fig. 2A). Phylogenetic analyses of each of the four more 3′ genes all suggested that MAL contained subtype D–like sequences (not shown), although the results were less conclusive than for *env*, because these sequences are so short. However, analysis of *vif* sequences showed MAL not to be particularly close to subtype A or D, suggesting that this gene might be mosaic.

To map this possible breakpoint more precisely, we examined the distribution of phylogenetically informative sites along the *vif* sequence in an alignment including MAL, one representative from each of subtypes A (U455) and D (ELI), and an outgroup (again SIV_{CPZ}-GAB). Each informative site can support one of three possible trees (Fig. 3A): (1) in which MAL clusters with U455, as seen in the *pol* tree; (2) in which MAL clusters with ELI, as seen in the *env* tree; (3) in which MAL is the outgroup to U455 and ELI. If the mosaic nature of the MAL genome has arisen through recombination of subtype A and D viruses (only), and assuming that the more recent ancestors of U455 and ELI have not been involved in recombination, then sites supporting tree 3 should only occur through parallel substitutions on different branches of the tree. Such sites should be relatively few in number (or else the sequences are too divergent for analysis), and should be scattered along the alignment. Indeed, this is what was observed: among the 27 informative sites there were just five supporting tree 3, and they were distributed across the gene (Fig. 3B). In contrast, the 11 sites supporting each of trees 1 and 2 were predominantly clustered in the 5′ and 3′ regions of the *vif* gene, respectively (Fig. 3B). If a breakpoint is placed somewhere between positions 278 and 327 (corresponding to positions 275 and 324 in the alignment of Myers et al. 1993; pages I-A-99–103), the ratios of sites supporting each of trees 1 and 2, i.e., 10:2 in the 5′ region, and 1:9 in the 3′ region, differ in a highly significant fashion (chi square = 11.7 with 1 degree of freedom; $P < 0.001$). Analyses using other subtype A (IFA86) and subtype D (Z2Z6, NDK) isolates produced similar results and did not define the breakpoint any more precisely. Analyses of the region around the junction of *env* and *nef* genes suggest that the 3′ crossover point lies just inside the *env* gene (Fig. 2A). A breakpoint between positions 2436 and

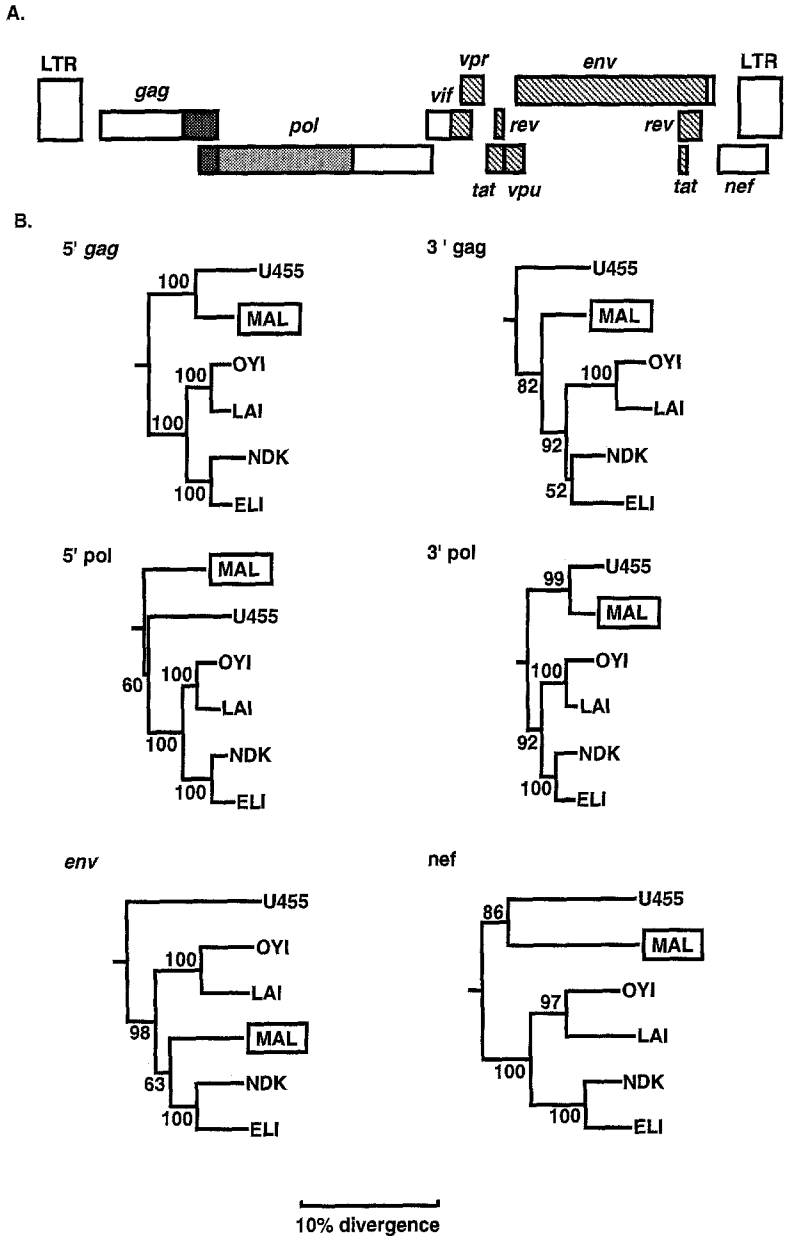


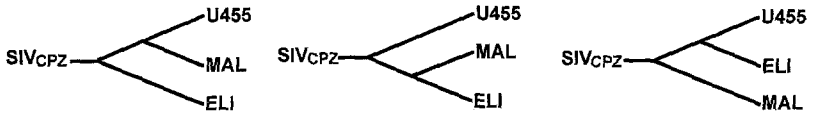
Fig. 2. The mosaic genome of HIV-1_{MAL}. **A** Diagrammatic representation of the position of genetic elements along the HIV-1 genome. *Open and hatched boxes* denote regions in which the MAL sequence is more closely related to viruses of subtypes A and D, respectively; the *stippled regions* are not closely related to subtype A or D. **B** Phylogenetic relationships of MAL *gag, pol, env,* and *nef* DNA sequences to viruses of sequence subtypes A (U455), B (OYI and LAI), and D (NDK and ELI); note that the 5′ *pol* tree did not include the region of overlap with *gag,* and the *env* tree was based on the *hatched* region of *env* only. *Horizontal branch lengths* are drawn to scale; the *bar* indicates 0.10 substitutions per site. Values at nodes indicate the percentage of bootstraps in which the cluster to the right was found. The tree was rooted using $SIV_{CPZ}GAB$ as an outgroup (Fig. 1). Accession numbers for the sequences used are given in Appendix 1.

2481 of the *env* alignment yields ratios of 25:76 in the 5′ region and 6:1 in the 3′ region for sites supporting each of trees 1 and 2, respectively: the chi square is = 11.9 with 1 degree of freedom ($P < 0.001$).

Similar phylogenetic and informative site distribution analyses of partial *gag* and *pol* sequences gave some indication that the evolutionary history of the MAL genome may be even more complex (Fig. 2B). The major part of *gag* and the 3′ region of *pol* from MAL are clearly subtype A. However, the sequence between these regions, including the region of *gag-pol* overlap, is not particularly similar to subtype A or D (or B). The breakpoint in *gag* mapped between positions 1068–1100, while that in *pol* mapped to a point between positions 2100 and 2123 (again, numbered according to Myers et al. 1993). Separate analyses based on the 3′ region of *gag* and the 5′ region of *pol* (the latter excluding the region of overlap) gave somewhat different trees, but it was not possible to define a clear breakpoint between these regions. This may indicate that an ancestor of MAL was also involved in recombination with another divergent strain, but sequences from this region of the HIV-1 genome are not available for representatives of the other subtypes, and so at this time it is not possible to resolve this matter.

Other strains of HIV-1 also appear to be recombinant. For example, CM238 and CM243 were isolated in Thailand (Louwagie et al. 1993), and are closely related to each other. In *gag*-derived trees these Thai strains cluster with sequence subtype A viruses, including MAL and U455 (Louwagie et al. 1993). However, in *env* trees they form a separate lineage, as distinct from subtype A as are

A.



B.

site	U455 A	MAL ?	ELI D	SIV_{CPZ}	tree
36	G	G	A	A	1
56	G	G	A	A	1
109	G	A	A	G	2
139	A	A	C	C	1
140	G	G	C	C	1
180	G	G	A	A	1
183	A	T	A	T	3
233	A	A	G	G	1
243	C	C	T	T	1
255	G	G	A	A	1
275	T	A	A	T	2
276	G	G	A	A	1
278	A	A	G	G	1
321	A	G	A	G	3
327	C	T	T	C	2
331	C	T	T	C	2
368	G	A	A	G	2
435	T	T	C	C	1
436	C	T	C	T	3
455	A	C	C	A	2
478	G	A	A	G	2
482	A	G	A	G	3
507	A	G	G	A	2
508	T	C	C	T	2
516	G	A	A	G	2
538	A	C	C	A	2
545	G	A	G	A	3

Fig. 3. Location of the recombination breakpoint in the *vif* gene of HIV-1_{MAL}. **A** Three possible relationships among HIV-1_{MAL}, HIV-1_{U455}, and HIV-1_{ELI}, with SIV_{CPZ}GAB as the outgroup. **B** Phylogenetically informative sites in the *vif* gene alignment, and the tree (from part A) that they support.

subtypes B and D (Sharp et al. 1994); this lineage has been termed *env* sequence subtype E. No other representatives of subtype E have yet been described, and the putative A/E recombinants have not been completely sequenced. Thus, at the present time, it is not possible to define which viruses may have been involved in the putative recombination event.

While this paper was in preparation, we learned of two more instances of recombinant HIV-1 isolates. Sabino et al. (1994) have determined the *env* gene sequences of isolates from two epidemiologically linked individuals from Brazil and shown that these genes appear to be the result of a recombination between viruses of sequence subtypes B and F. Diaz et al. (1994) have detected a recombinant *env* sequence in an infant who was transfused with packed red blood cells from two different HIV-1–positive donors on the same day. In this case, both of the viruses involved were sequence subtype B, but despite this close relationship the recombination was discernible because both parental genomes were available for analysis.

Recombination in HIV-2

HIV-2 strain 7312A was isolated from a male student from Abidjan (Cote d'Ivoire) who presented with generalized lymphadenopathy. Initial phylogenetic analyses were performed on partial *pol* and *env* sequences amplified from uncultured PBMC DNA by nested polymerase chain reaction (Gao et al. 1992). The results showed discordant branching orders: the *pol* fragment of 7312A was found to be most closely related to D205, the prototypic isolate of what is now known as HIV-2 subtype

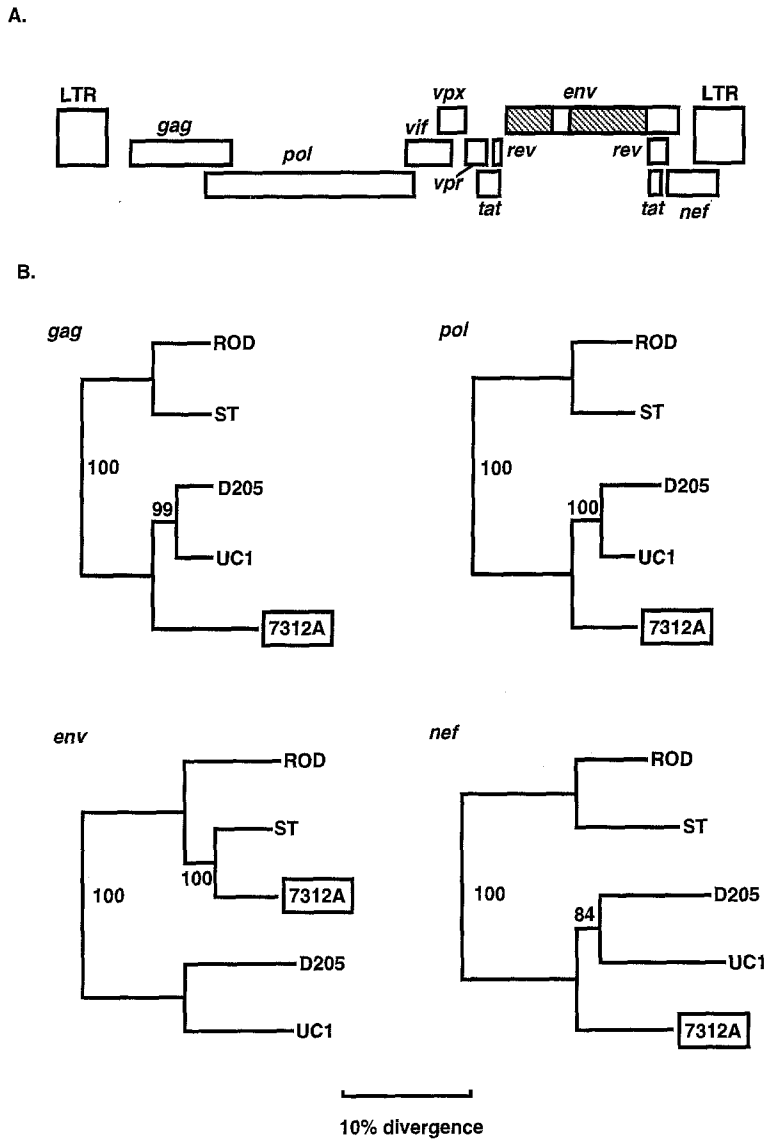


Fig. 4. The mosaic genome of HIV-2_{7312A}. **A** Diagrammatic representation of the position of genetic elements along the HIV-2 genome. *Open and hatched boxes* denote regions in which the 7312A sequence is more closely related to viruses of subtypes B and A, respectively. **B** Phylogenetic relationships of 7312A *gag, pol, env,* and *nef* DNA sequences to viruses of sequence subtypes A (ROD and ST) and B (D205 and UC1); note that the *env* tree was based only on the *hatched* regions of *env. Horizontal branch lengths* are drawn to scale; the *bar* indicates 0.10 substitutions per site. The approximate position of the root of the tree (at the left) was determined from analyses using other primate lentiviruses as outgroups (Fig. 1). Values at nodes indicate the percentage of bootstraps in which the cluster to the right was found. Accession numbers for the sequences used are given in Appendix 1.

B, but the *env* fragment clustered with ST, a representative of subtype A (Gao et al. 1992). These results suggested that there was a crossover in the region between these two fragments, although it was also possible that individual 7312A was simultaneously coinfected with two different strains of HIV-2, which had been differentially amplified in the polymerase chain reaction (PCR) procedure used to obtain DNA fragments for sequencing.

Subsequently, a complete replication-competent provirus of 7312A has been cloned and sequenced (Hui et al. 1994). Here we use this newly determined sequence to examine whether 7312A is indeed a recombinant. Phylogenetic analysis of individual gene sequences revealed that most of the genome of 7312A is subtype B-like; only the *env* gene is subtype A-like (Fig. 4). This confirms a mosaic 7312A genome and rules out double infection. More detailed analysis of informative site configurations (as performed for HIV-1_{MAL} above) revealed four putative recombination breakpoints within the *env* gene (Table 1). The first mapped to the beginning of the *env* gene. Two more breakpoints delimit a region of approximately 250–300 bp of subtype B–like sequence within the predominantly subtype A–like *env* gene. The final breakpoint is localized to a site near the end of *env* in the region where the second exons of *rev* and *tat* begin.

We have suggested (Sharp et al. 1994; Gao et al. 1994) that HIV-2 sequence subtypes A and B likely represent independent transmissions of sooty mangabey viruses to humans. The comparatively small amount of divergence between 7312A and ST in the major part of *env* indicates that the recombination event was relatively recent, and occurred in a human host rather than in a mangabey. However, it is not possible to determine whether the recombination actually occurred in 7312A or in an earlier individual.

Another possible example of a recombinant HIV-2 has recently been identified (Gao et al. 1994). Strain FA was isolated from a female AIDS patient hospitalized with end-stage disease in Accra (Ghana). Phylogenetic analyses of partial *gag, pol,* and *env* sequences all

Table 1. Mapping recombination breakpoints in HIV-2_{7312A}[a]

Gene	Region[b]	Phylogeny[c] 1	2	3	Chi square[d]
tat	All	2	16	6	
					42.9
env	1–599	58	7	13	
					42.9
env	669–841	1	14	2	
					52.3
env	887–2022	77	8	7	
					64.7
env	2059–2424	10	37	6	

[a] The values given are the number of phylogenetically informative sites supporting each of the three possible relationships of HIV-2_{7312A} to HIV-2_{ST} (subtype A), HIV-2_{D205} (subtype B), and SIV_{SM} (the outgroup). See Fig. 3 for a more detailed example of this method and Fig. 4A for a diagrammatic representation of these results

[b] Region of the aligned sequence. Positions are according to the alignment of Myers et al. (1993; pages I-B-49–75)

[c] Phylogenies: 1 places HIV-2_{7312A} with HIV-2_{ST} in subtype A; 2 places HIV-2_{7312A} with HIV-2_{D205} in subtype B; 3 places HIV-2_{7312A} outside a cluster of HIV-2_{ST} and HIV-2_{D205}

[d] Chi-square test for heterogeneity between adjacent regions of the distribution of informative sites supporting trees 1 and 2. All values have 1 degree of freedom and have probabilities $\ll 0.001$

showed FA to be a subtype A virus. However, in the *gag* and *pol* trees FA clustered with one particular subset of subtype A viruses, while in the *env* tree FA fell in a different group; in each case, the clustering was significant (as assessed by bootstraps). Interestingly, based on *env* sequences, FA was most closely related to 7312A, suggesting that this lineage has been involved in two successive recombinations. Since these were again PCR-amplified sequences, it cannot yet be ruled out that individual FA is in fact doubly infected. If the sequences from FA do reflect recombination, it must have again been a relatively recent event, occurring substantially after the radiation of viruses in HIV-2 subtype A.

Conclusions

In this paper, we have described examples of HIV-1 and HIV-2 which have mosaic genomes, sequences from different regions of which have different phylogenetic histories. Elsewhere (Jin et al. 1994), we have recently described analogous results indicating that sabaeus monkeys (one of the four *C. aethiops* species) are infected with a virus (SIV_{AGM}sab) that resulted from a recombination event involving the ancestors of SIV_{AGM} and SIV_{SM} (from sooty mangabeys). In one sense, these observations are not surprising since retroviruses are known to be highly recombinogenic. However, these results are very interesting because they imply that individual hosts must have been simultaneously coinfected with rather divergent viruses. This is the first, albeit indirect, evidence for such superinfection.

Laboratory investigations of recombination in retroviruses have indicated two points of particular relevance here. First, recombination may involve multiple crossovers along the viral genome (Hu and Temin 1990). Indeed, in the detailed analyses of both HIV-1_{MAL} and HIV-2_{7312A} described above, we found evidence for such multiple crossovers. This may also explain why the ancestry of the 5′ end of the SIV_{AGM}sab genome was hard to elucidate (Jin et al. 1994), since that region may be highly mosaic. Second, the efficiency of recombination depends on the length of sequence identity shared between the parental strains (Zhang and Temin 1994). The MAL and 7312A viruses have arisen from recombination of different subtypes of HIV-1 and HIV-2, respectively; i.e., these events did not involve very highly divergent viruses. Even the event generating the SIV_{AGM}sab genome did not involve extremely divergent strains (Jin et al. 1994). That is, while the viruses involved were ancestors of strains that today appear in different primate lentivirus lineages, at the time in the past when the event occurred they were possibly little more divergent than subtypes of, for example, HIV-1 are today.

These examples probably reflect only a fraction of the recombination events that have occurred in the past. First, the experience over the last few years has been that ever-more genetic diversity is revealed by the characterization of new isolates. Second, it is quite likely that a number of naturally occurring recombination events would generate nonfunctional viruses, or at least viruses with reduced fitness, since the HIV life cycle involves complex interactions between different genetic elements along the genome. Obviously, defective viruses would not be expected to persist to be picked up for sequence analysis. In this light it is interesting to compare the breakpoints found in the three recombinant lentiviruses. The analyses of MAL (above) and SIV_{AGM}sab (Jin et al. 1994) suggest that hybrid *gag* and *pol* proteins can be quite functional. In contrast, in all three recombinant lentiviruses, both exons of *rev* and *tat* come from only one of the parental viruses. This may indicate that hybrids of these genes or their products are not fully functional.

Although superinfection with multiple strains of different subtypes of HIV-1 or HIV-2 had not previously been reported, individuals who are simultaneously coinfected with HIV-1 and HIV-2 have recently been found (Grez et al. 1994). However, HIV-1/HIV-2 recombinants have not yet been identified. This may be because the extent of sequence divergence between these two different lineages is much greater than within either, such that the efficiency of recombination may be much reduced. However, it seems more likely that HIV-1/HIV-2 recombinants have not been detected because they are functionally impaired (Ranganathan and Srinivasan 1993).

In recent years there have been many analyses of the phylogenetic relationships within and among the various types of primate lentiviruses. These efforts are aimed both at elucidating the complex evolutionary origins of these viruses and at subtype classification of new isolates from around the world, and are important to gaining a better understanding of the epidemiology of AIDS. For example, given the genetic diversity apparent within HIV-1, if vaccines are successfully developed against one subtype they might not be effective against another. In that case it is vital to know which subtypes are circulating in particular communities. Subtype classification has been largely performed by phylogenetic analysis of short genomic sequences. The existence of recombinant viruses indicates that caution must be exercised when extrapolating from such analyses. For example, one of the seminal papers demonstrating the existence of multiple diverse genetic subtypes of HIV-1 utilized partial *gag* sequences (Louwagie et al. 1993). In this region, the isolate MAL falls in subtype A. However, if a subtype A–specific vaccine were targeted at *env* peptides, MAL might escape since its *env* region is subtype D.

In addition to its potential impact on current AIDS vaccine development efforts, intersubtype recombination may also influence clinically important properties of HIV, including transmissibility, virulence, and replication potential. There have been numerous reports of recombination in other groups of viruses (e.g., see Strauss and Strauss 1988), sometimes with serious consequences (Javier et al. 1986; Hahn et al. 1988), and the potential for similar events has to be considered for HIV. Although it is presently unknown how and under what circumstances the same person becomes superinfected with two different viruses, the mosaic genomes described above document that coinfection and recombination are certainly possible for both HIV-1 and HIV-2. Moreover, current data reveal a variety of different genetic combinations, including, for example, A/D, B/F, and A/E recombinants for HIV-1. Most of these viruses appear to be fully viable and pathogenic (Alizon et al. 1986; Louwagie et al. 1993), and at least one has been successfully transmitted to another individual (Sabino et al. 1994). In the light of these findings, it will be important to determine how frequently such recombinant viruses are generated and whether they differ in their natural history and pathogenicity. Such studies will be particularly important in geographic regions where multiple sequence subtypes are known to circulate in the same populations, such as Africa, Asia, and South America.

The finding of intersubtype recombinants also raises the question of *when* during the course of HIV infection superinfection can occur. It has been thought that the host's immune response may play a role in protecting against infection by a second HIV strain. It may be that superinfection is possible only if exposure to a second virus occurs during the first weeks after initial infection, before an effective immune response is mounted. Alternatively, perhaps the host's immune protection fails to extend to more divergent strains, e.g., members of different subtypes, an issue of obvious importance for vaccine development efforts.

In conclusion, the demonstration that recombinant primate lentiviruses exist has two important implications. First, individuals can be simultaneously coinfected with divergent strains of HIV or SIV. Second, it is clear that phylogenetic analyses must recognize that partial genomic sequences may not reveal the entire evolutionary history and genetic relationships of these viruses, and that even partial sequences should be critically examined to determine whether they are mosaic.

Acknowledgments. We thank Allen Mayer for communicating results prior to publication. This work has been supported by a University of Nottingham studentship (to D.L.R.), by grants from the National Institutes of Health (RO1 AI 25291; RO1 AI 28273; NO1 AI 35170; to B.H.H.), and by shared facilities of the UAB Center for AIDS Research (P30 AI 27767). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

Alizon M, Wain-Hobson S, Montagnier L, Sonigo P (1986) Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. Cell 46:63–74

Coffin JM (1979) Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J Gen Virol 42:1–26

Coffin JM (1986) Genetic variation in AIDS viruses. Cell 46:1–14

Delassus S, Cheynier R, Wain-Hobson S (1991) Evolution of human immunodeficiency virus type 1 *nef* and long terminal repeat sequences over 4 years in vivo and in vitro. J Virol 65:225–231

Diaz R, Sabino EC, Mayer A, Mosley JW, Busch MP, The Transfusion Study Group (1994) Evidence of HIV-1 dual infection and recombination in a dually exposed transfusion recipient. J Virol (submitted)

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Gao F, Yue L, White AT, Pappas PG, Barchue J, Hanson AP, Greene BM, Sharp PM, Shaw GM, Hahn BH (1992) Human infection by genetically diverse SIV_{SM}-related HIV-2 in West Africa. Nature 358:495–499

Gao F, Yue L, Robertson DL, Hill SC, Hui H, Biggar RJ, Neequaye AE, Whelan TM, Ho DD, Shaw GM, Sharp PM, Hahn BH (1994) Genetic diversity of Human Immunodeficiency Virus type 2: evidence for distinct sequence subtypes with differences in virus biology. J Virol 68:7433–7447

Goodrich DW, Duesberg PH (1990) Retroviral recombination during reverse transcription. Proc Natl Acad Sci USA 87:2052–2056

Grez M, Dietrich U, Balfe P, von Briesen H, Maniar JK, Mahambre G, Delwart EL, Mullins JI, Rubsamen-Waigmann H (1994) Genetic analysis of human immunodeficiency virus type 1 and 2 (HIV-1 and HIV-2) mixed infections in India reveals a recent spread of HIV-1 and HIV-2 from a single ancestor for each of these viruses. J Virol 68:2161–2168

Groenink M, Andeweg AC, Fouchier RAM, Broersen S, van der Jagt

RCM, Schuitemaker H, de Goede REY, Bosch ML, Huisman HG, Tersmette M (1992) Phenotype-associated *env* gene variation among eight related human immunodeficiency virus type 1 clones: evidence for in vivo recombination and determinants of cytotropism outside the V3 domain. J Virol 66:6175–6180

Gurtler LG, Hauser PH, Eberle J, von Brunn A, Knapp S, Zekeng L, Tsague JM, Kaptue L (1994) A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. J Virol 68: 1581–1585

Hahn BH (1994) Viral genes and their products. In: Broder S, Merigan T, Bolognesi D (eds) Textbook on AIDS medicine. Williams & Williams, New York, pp 21–43

Hahn BH, Shaw GM, Taylor ME, Redfield RR, Markham PD, Salahuddin SZ, Wong-Staal F, Gallo RC, Parks ES, Parks WP (1986) Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science 232:1548–1553

Hahn CS, Lustig S, Strauss EG, Strauss JH (1988) Western equine encephalitis virus is a recombinant virus. Proc Natl Acad Sci USA 85:5997–6001

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. Comput Appl Biosci 5:151–153

Hirsch VM, Olmsted RA, Murphy-Corb M, Purcell RH, Johnson PR (1989) An African primate lentivirus (SIV_{SM}) closely related to HIV-2. Nature 339:389–392

Howell RM, Fitzgibbon JE, Noe M, Ren Z, Gocke DJ, Schwartzer TA, Dubin DT (1991) In vivo sequence variation of the human immunodeficiency virus type 1 *env* gene: evidence for recombination among variants found in a single individual. AIDS Res Hum Retroviruses 7:869–876

Hu W-S, Temin HM (1990) Retroviral recombination and reverse transcription. Science 250:1227–1233

Hui H, Kim J, Hill SC, Kappes J, Shaw GM, Hahn BH (1994) Sequence analysis of a recombinant HIV-2 provirus. (in preparation)

Javier RT, Sedarati F, Stevens JG (1986) Two avirulent herpes simplex viruses generate lethal recombinants in vivo. Science 234:746–748

Jin MJ, Hui H, Robertson DL, Muller MC, Barre-Sinoussi F, Hirsch VM, Allan JS, Shaw GM, Sharp PM, Hahn BH (1994) Mosaic genome structure of simian immunodeficiency virus from West African green monkeys. EMBO J 13:2935–2947

Johnson PR, Hirsch VM, Myers G (1991) Genetic diversity and phylogeny of nonhuman primate lentiviruses. In: Koff W (ed) Annual review of AIDS research, vol 1. Marcel Dekker, New York, pp 47–62

Junghans RP, Boone LR, Skalka AM (1982) Retroviral DNA H structures: displacement-assimilation model of recombination. Cell 30: 53–62

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Li W-H, Tanimura M, Sharp PM (1988) Rates and dates of divergence between AIDS virus nucleotide sequences. Mol Biol Evol 5:313–330

Louwagie J, McCutchan F, Peeters M, Brennan TP, Sanders-Buell E, Eddy GA, van der Groen G, Fransen K, Gershey-Damet G-M, Deleys R, Burke DS (1993) Phylogenetic analysis of *gag* genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. AIDS 7:769–780

Maynard Smith J (1992) Analysing the mosaic structure of genes. J Mol Evol 34:126–129

Myers G, Korber B, Wain-Hobson S, Smith RF, Pavlakis GN (1993) Human retroviruses and AIDS. Theoretical Biology and Biophysics, Los Alamos, NM

Myers G, MacInnes K, Korber B (1992) The emergence of simian/human immunodeficiency viruses. AIDS Res Hum Retroviruses 8:373–386

Peeters M, Fransen K, Delaporte E, Vanden Haesevelde M, Gershy-Damet G-M, Kestens L, van der Groen G, Piot P (1992) Isolation and characterization of a new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-captured chimpanzee. AIDS 6:447–451

Preston BD, Poiesz BJ, Loeb LA (1988) Fidelity of HIV-1 reverse transcriptase. Science 242:1168–1171

Ranganathan PN, Srinivasan A (1993) Recombination between human immunodeficiency viruses (HIV) type 1 and 2 results in generation of defective hybrid viruses. Biochem Biophys Res Commun 193: 311–317

Richetti M, Buc H (1990) Reverse transcriptase and genome variability: the accuracy of DNA replication is enzyme specific and sequence dependent. EMBO J 9:1583–1593

Roberts JD, Bebenek K, Kunkel TA (1988) The accuracy of reverse transcriptase from HIV-1. Science 242:1171–1173

Sabino EC, Shpaer EG, Morgado MG, Korber BTM, Diaz RS, Bongertz V, Cavalcante S, Galvao-Castro B, Mullins JI, Mayer A (1994) Identification of HIV-1 envelope genes recombinant between subtypes B and F in two epidemiologically-linked individuals from Brazil. J Virol 68:6340–6346

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sharp PM, Robertson DL, Gao F, Hahn BH (1994) Origins and diversity of human immunodeficiency viruses. AIDS 8:S27–S42

Strauss JH, Strauss EG (1988) Evolution of RNA viruses. Annu Rev Microbiol 42:657–683

Vanden Haesevelde M, Decourt RJ, De Leys RJ, Vanderborght B, van der Groen G, van Heuverswijn H, Saman E (1994) Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. J Virol 68:1586–1596

Vartanian J-P, Meyerhans A, Asjo B, Wain-Hobson S (1991) Selection, recombination, and G → A hypermutation of human immunodeficiency virus type 1 genomes. J Virol 65:1779–1788

Wain-Hobson S (1993) The fastest genome evolution ever described: HIV variation *in situ*. Curr Opin Genet Dev 3:878–883

Weber J, Grosse F (1989) Fidelity of human immunodeficiency virus type 1 reverse transcriptase in copying natural DNA. Nucleic Acids Res 17:1379–1393

Williams PL, Fitch WM (1990) Phylogeny determinations using dynamically weighted parsimony method. Methods Enzymol 183: 615–626

Wolfs TFW, Zwart G, Bakker M, Goudsmit J (1992) HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology 189:103–110

Zhang, J, Temin HM (1994) Retrovirus recombination depends on the length of sequence identity and is not error prone. J Virol 68:2409–2414

Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Leigh Brown AJ, Simmonds P (1993) Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. J Virol 67:3345–3356

Zhu T, Mo H, Wang N, Nam DS, Cao Y, Koup RA, Ho DD (1993) Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. Science 261:1179–1181

Appendix 1. Sources for sequences analyzed

Isolate	Acc. #[a]	Reference
$HIV\text{-}1_{ANT70}$	L20587	Vanden Haesewelde et al. (1994) *J. Virology* 68:1586–1596
$HIV\text{-}1_{ELI}$	K03454	Alizon et al. (1986) *Cell* 46:73–74
$HIV\text{-}1_{LAI}$	K02013	Wain-Hobson et al. (1985) *Cell* 40:9–17

Appendix 1. Continued

Isolate	Acc. #[a]	Reference
HIV-1_{MAL}	K03456	Alizon et al. (1986) *Cell* 46:63–74
HIV-$1_{MVP5180}$	L20571	Gurtler et al. (1994) *J. Virology* 68:1581–1585
HIV-1_{NDK}	M27323	Spire et al. (1989) *Gene* 81:275–284
HIV-1_{OYI}	M26727	Huet et al. (1989) *AIDS* 3:707–715
HIV-1_{U455}	M62320	Orom et al. (1990) *AIDS Res. Hum. Retro.* 6:1073–1078
HIV-2_{D205}	X61240	Dietrich et al. (1992) *AIDS Res. Hum. Retro.* 8:1619–1629
HIC-2_{ROD}	M15390	Guyader et al. (1987) *Nature* 326:662–669
HIV-2_{ST}	M31113	Kumar et al. (1990) *J. Virology* 64:890–901
HIV-2_{UC1}	L07625	Barnett et al. (1993) *J. Virology* 67:1006–1014
HIV-2_{7312A}	L36874	Hui et al. (1994) (in preparation)

Appendix 1. Continued

Isolate	Acc. #[a]	Reference
$SIV_{CPZ}GAB$	X52154	Huet et al. (1990) *Nature* 345:356–359
$SIV_{SM}H4$	X14307	Hirsch et al. (1989) *Nature* 339:389–391
$SIV_{MAC}251$	M19499	Kestler et al. (1988) *Nature* 331:619–622
$SIV_{AGM}gri\text{-}1$	M58410	Fomsgaard et al. (1991) *Virology* 182:397–402
SIV_{AGM} verTYO-1	X07805	Fukasawa et al. (1988) *Nature* 333:457–461
$SIV_{SYK}173$	L06042	Hirsch et al. (1993) *J. Virology* 67:1517–1528
$SIV_{MND}GB1$	X15781	Tsujimoto et al. (1989) *Nature* 341:539–541

[a] Acc. # is the GenBank/EMBL DNA sequence data library accession number.