Intersubtype Recombinant HIV-1 Sequences

David L. Robertson,[1] **Feng Gao,**[2] **Beatrice H. Hahn,**[2] **Paul M. Sharp**[3]

[1] *Laboratory of Structural and Genetic Information, CNRS-EP 91, 31 Chemin Joseph Aiguier, 13402 Marseilles Cedex 20, France;*
[2] *Departments of Medicine and Microbiology, University of Alabama at Birmingham, 702 S. 19th Street, LHRB 613, Birmingham, AL 35294;*
[3] *Division of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, U.K.*

Retroviruses are highly recombinogenic, but recombination can only occur between genomes packaged within the same virion. Until about 1994, it was generally thought that individuals do not become infected with multiple distinct HIV-1 strains, and so the possibility that recombination between divergent viruses could contribute to the evolution of HIV-1 was not widely considered. However, not long after the recognition that the considerable global diversity of the major (M) group of HIV-1 could be described in terms of discrete sequence subtypes, it was also realized that a significant fraction of HIV-1 strains are in fact intersubtype recombinants [1,2]. Analyses of viruses characterized over the last few years have confirmed that substantial numbers of HIV-1 isolates have mosaic genomes [3–16]. The finding of such hybrid viruses has a number of implications both for attempts at viral characterization (phylogenetic analyses, and tracking of the epidemic), and for concerns about the future genetic diversity of HIV-1 and its impact on vaccine development.

Hybrid viruses can be detected because their phylogenetic affinities vary depending on the region of genome analysed. A useful initial approach is to examine the extent of sequence divergence/similarity between a new sequence and a bank of reference sequences of different subtypes, for example as a diversity plot [9], or using the RIP program [17]; if the extent of relative similarity to different subtypes varies along the sequence, this may indicate that the sequence is a recombinant. However, fuller investigation must involve a phylogenetic approach, comparing trees derived by analyses of different regions of the genome, and assessing the confidence of phylogenetic clustering by a statistical approach such as the bootstrap. A thorough analysis would involve taking a window of sequence of a certain size, and moving this window along the genome in steps of a defined size, generating perhaps hundreds of trees for visual examination in the process! There are at least two short cuts. One is to analyse only a few windows, defining selected regions according to the output of the diversity analysis. Another is to not examine the entire phylogenetic tree of all subtypes, but to focus on one particular phylogenetic question. Thus, if the initial analyses suggest that a sequence may be a recombinant between two particular subtypes, it is possible to ask simply what is the bootstrap value for the clustering of the new sequence with one or another particular subtype, and plot these values as a function of position along the genome; this is the basis of the "bootscanning" approach [18]. Once the subtypes putatively involved in the recombination event have been identified, and the crossover points have been approximately localized, more precisely defined breakpoints can be determined, and their statistical significance assessed, using informative site analysis [2,6,19]. (For a fuller discussion of recombination analysis methods, see Ref.16.)

Among the full-length HIV-1 sequences available currently (i.e., in December 1997), there are 12 that appear to be intersubtype recombinants (Table 1, Figure 1). Clearly, the delineation of recombinant genomes, and localization of the crossover points, relies on the availability of nonrecombinant reference sequences. Since such sequences are, as yet, available only for subtypes A, B, C, D, F, and H (see Leitner et al., pp. III.19-III.24, this volume), the breakpoints between regions of sequence of different subtype origin can be mapped only tentatively in those viruses derived from recombination events involving subtypes E, G, I or J; this includes the majority of the full length genome hybrids—in fact, all but the first three in the accompanying Figure and Table. In several viruses there are regions (shown in white) which have not yet been assigned to any subtype; the origin of these segments may be clarified as more reference sequences become available, but it is also possible that some of these regions come from subtypes that are no longer extant.

Table 1 Full-length Intersubtype Recombinant HIV-1 Sequences

Sequence	Acc. No.	Subtypes	Origin	Year	References
ZAM184	U86780	A,C	Zambia	1990	[12]
92RW009.6	U88823	A,C	Rwanda	1992	[16]
93BR029.4	AF005495	B,F	Brazil	1993	[16]
90CF402.1	U51188	A,E	C.A.R.	1990	[9]
93TH253.3	U51189	A,E	Thailand	1993	[9]
CM240	U54771	A,E	Thailand	1990	[8]
92NG083.2	U88826	A,G	Nigeria	1992	[16]
92NG003.2	U88825	A,G	Nigeria	1992	[16]
IbNg	L39106	A,G	Nigeria	1991	[33]
MAL	K03456	A,D,I,?	Zaire	1985	[20]
Z321B	U76035	A,G,I,?	Zaire	1976	[14]
94CY032.3	AF049337	A,G,I,?	Cyprus	1994	[25]

The references indicate the sequence publication. IbNg [9] and MAL [19,22] were only subsequently recognized as recombinants. Utilizing the latest set of reference sequences for the various subtypes (see Leitner et al., pp. III.19–III.24), we have subjected all of these sequences to further analysis. The locations of the segments of different subtype origins, revised according to these latest analyses (Robertson, Gao, Sharp and Hahn; in preparation), are shown in Figure 1.

Most of the recombinant viruses originate from Africa (Table 1), and a substantial majority (11 of the 12 full length hybrids) include segments of subtype A (Figure 1). This is not surprising, since it appears that the various subtypes of HIV-1 first arose in central Africa and have co-circulated there for many years. Furthermore, in that region, subtype A is the most common. Among the hybrids of African origin, MAL is interesting because it was one of the first African (and, in retrospect, nonsubtype B) viruses to be characterized [20]. Li et al. first suggested that MAL may be a recombinant virus in 1987 [21,22], but the realization of the widespread occurrence of recombination among HIV-1 subtypes did not occur for about another 7 years. While MAL includes regions that are clearly of subtype A and D origin, there is a long region encompassing the 5′ half of the *pol* gene that has defied characterization [19]. Now it appears that part of that region may be subtype I (see below), but the origin of the remainder has still not been determined (Figure 1). Thus, the ancestry of MAL may have involved successive recombination events of sequences derived from three or four different subtypes. Z321B, which also appears to include sequence blocks from three (or four?) different subtypes, is the oldest of the known hybrid viruses, having been isolated in 1976. Clearly, by that date, the group M subtypes had diverged sufficiently that the various segments recombined within Z312B are detectable as having distinct subtype origins. This implies that the subtypes must have begun to diverge substantially earlier than the 1970s, and is consistent with the recent finding that a virus from 1959 appears, from phylogenetic analyses, to have postdated the group M radiation [23].

Among the hybrid viruses isolated outside of Africa are two subtype "E" viruses (CM240 and 93TH253.3) from Thailand. As shown in Figure 1, a large fraction of the genome of these subtype "E" viruses, including the *gag* and *pol* genes, appears to be of subtype A origin [8,9]. Since all Thai "E" viruses appear to be closely related, and since the mosaic structure of their genomes appears to be identical to that of the other full length subtype "E" virus (90CF402.1) from the Central African Republic, it is most likely that a single recombination event occurred in Africa, and that the subtype "E" viruses in Thailand are due to the introduction of a descendant of that recombination event [9]. While the subtype "E" viruses clearly indicate that a subtype E arose in the past, it is possible that full length subtype E viruses no longer exist. In contrast, the recombination event between subtypes B and F to generate 93BR029.4 most likely occurred in Brazil, where both of these subtypes are common.

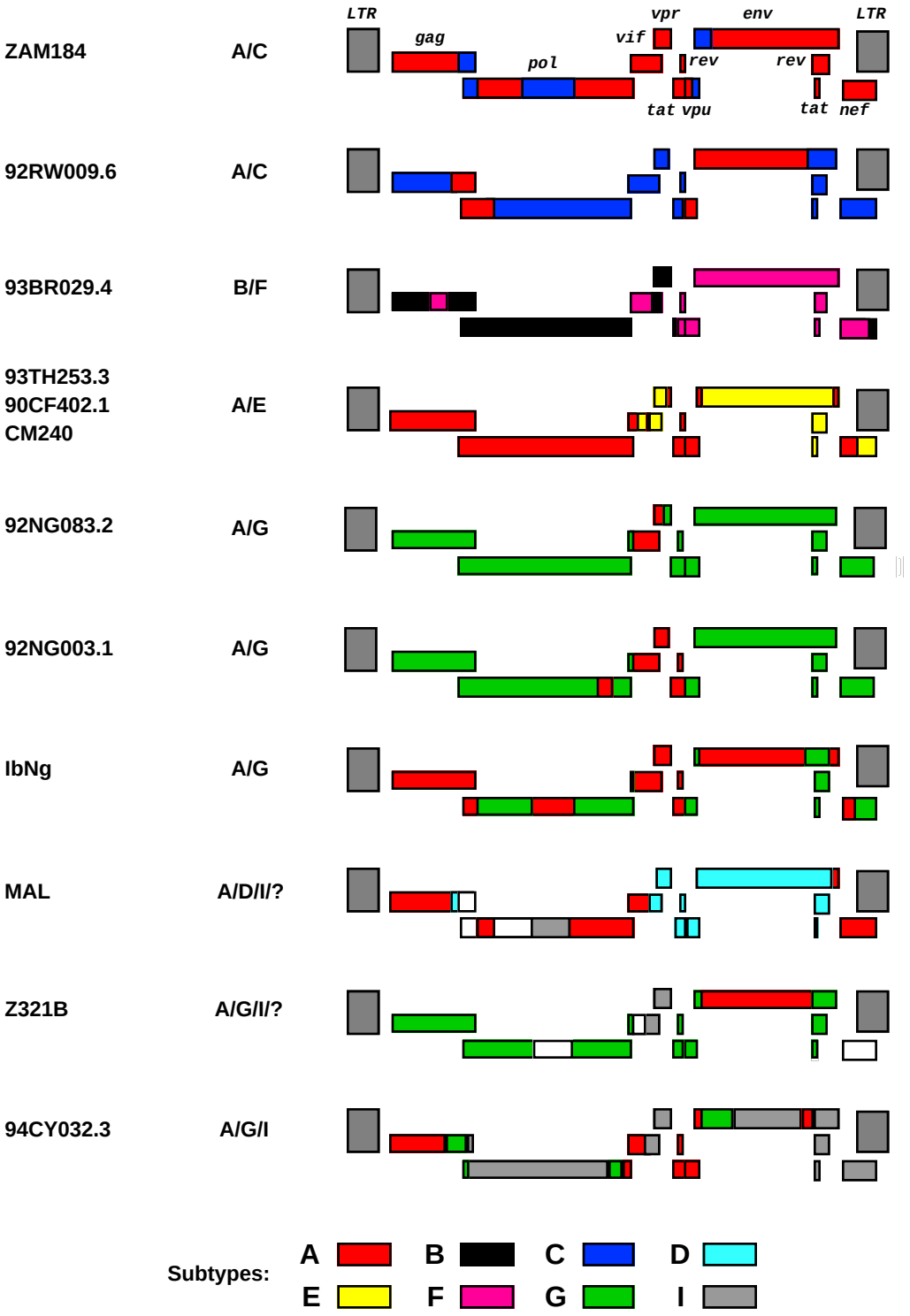

Figure 1. Full-length intersubtype recombinant HIV-1 sequences. The color code indicates the subtype origin of different segments of the genomes; regions in white are sequences whose origin has not been determined. The LTRs are hatched because they have not been analysed in detail. See Table 1 for more details.

At least three other viruses from Brazil, for each of which only partial sequences are available, appear to represent independently generated B/F hybrids [3,6,18]. The other non-African example, 94CY032.3, was isolated in Cyprus. Initially, partial sequences from this virus formed the basis of the proposition of subtype I [24]. Analysis of a recently determined full length sequence [25] suggests that it is a recombinant of at least three subtypes, A, G and I. Since 94CY032.3 is the prototypic subtype I virus, it is most parsimonious to assume that all segments which cannot be assigned to any other subtype are of a common (i.e., subtype I) origin. Two of the recombinant viruses from Africa, MAL and Z321B, contain short segments of sequence closely related to 94CY032.3, but not similar to other subtypes; thus, these regions have also been designated as being of subtype I origin (Figure 1). It is not clear whether the multiple recombination events in the ancestry of 94CY032.3 occurred in Africa, or in Cyprus, where a number of different subtypes are also found [24].

It is notable that most of the mosaic genomes have a "patchwork" appearance due to multiple cross-over events (Figure 1). This is as expected given the known properties of reverse transcriptase. As more recombinant viruses are characterized, it may be possible to discern whether there are some consistent features to the patterns of genetic exchange. For example, certain sites may be hotspots for recombination; a hotspot may exist because the genome adopts a secondary structure that induces stalling of the reverse transcriptase, or because a region is conserved and so has a higher level of sequence identity. Alternatively, some exchanges may be "forbidden;" numerous components of the HIV-1 genome interact, either at the level of their RNA or the encoded translation product, and certain combinations of divergent interacting components may be inviable.

From the current data it is not possible to estimate the frequency of multiple infection or of recombination events. The 12 intersubtype recombinants come from a total of 53 available full length sequences. Of the 53 full length sequences, 30 come from subtype B, which is the focus of attention in the Western world but constitutes only about 16% of the global pandemic. Among the 12 hybrids, three are similar subtype "E" viruses, selected for investigation because of their impact on the epidemic in Thailand, but which likely all arose from a single ancestral recombination event (see above). Thus, counting only one of the three subtype "E" genomes, 10/21 full length non-subtype B sequences are independently generated recombinants of different subtypes. Among the much larger numbers of viruses that have been partially characterized, there are many additional examples of intersubtype recombinants [2–6,10,11,13,15]. We have previously estimated that roughly 10% of viruses for which at least a near full-length *gag* or *env* gene sequence is available are recombinants [2]. This figure is possibly an underestimate of the frequency of recombinant viruses, both because of the excess representation of subtype B, and because a virus for which only one gene has been sequenced may be a recombinant but not have a crossover within that gene; for example, if judged only on the basis of their *gag* or *env* gene sequences (see Figure 1), 92NG003.1 and 92NG083.2 appear to represent nonrecombinant subtype G viruses, while 93BR029 was originally classified as subtype F on the basis of an *env* gene phylogeny [6]. Finally, the frequency of new recombinants in any given geographic area will obviously depend on the number and prevalence of co-circulating subtypes: multiple subtypes well represented in the same population can be expected to yield a higher number of new hybrid viruses.

Recombinant viruses have also been detected among HIV-2 strains [19,26] and among SIVs [27,28]. As yet, no viruses have been described that are recombinants of the two highly divergent major groups of HIV-1, M and O, nor have recombinants of HIV-1 and HIV-2 been found. Individuals multiply infected with HIV-1 groups M and O [29], and with both HIV-1 and HIV-2 [30], have been reported. Recombination between these more divergent viruses may not occur because of the extent of sequence dissimilarity or may yield replication attenuated or defective viruses. Rather few HIV-1 intrasubtype recombinants have been described [7,31,32], but in this case this is probably merely a problem of detection.

References

[1] Hahn, B.H., Robertson, D.L., McCutchan, F.E. and Sharp, P.M. Recombination and diversity of HIV: implications for vaccine development. Pp. 87–94 in: *Retroviruses of Human AIDS and Related Diseases*. Eds. Girard, M. and Dodet, B. Neuvieme Colloque des Cent Gardes, 1994.

[2] Robertson, D.L., Sharp, P.M., McCutchan, F.E. and Hahn, B.H. Recombination in HIV-1. *Nature* **374**:124–126, 1995.

[3] Sabino, E.C., Shpaer, E.G., Morgado, M.G., Korber, B.T., Diaz, R.S., Bongertz, V., Cavalcante, S., Galvao-Castro, B., Mullins, J.I. and Mayer, A. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J. Virol*. **68**:6340–6346, 1994.

[4] Hahn, B.H., Robertson, D.L. and Sharp, P.M. Intersubtype recombination in HIV-1 and HIV-2. Pp. III.22–III.29 in: *Human Retroviruses and AIDS: a Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Eds. Myers, G., Korber, B., Hahn, B.H., Jeang, K.-T., Mellors, J.W., McCutchan, F.E., Henderson, L.E. and Pavlakis, G.N. Los Alamos National Laboratory, Los Alamos, New Mexico, 1995.

[5] Leitner, T., Escanilla, D., Marquina, S., Wahlberg, J., Brotstrom, C., Hansson, H.B., Uhlen, M. and Albert, J. Biological and molecular characterization of subtype D, subtype G, and subtype A/D recombinant HIV-1 transmissions in Sweden. *Virology* **209**:136–146, 1995.

[6] Gao, F., Morrison, S.G., Robertson, D.L., Thornton, C.L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B., von Briesen, H., Beddows, S., Weber, J., Sharp, P.M., Shaw, G.M., Hahn, B.H., and the WHO and NIAID Networks for HIV Isolation and Characterization. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. *J. Virol*. **70**:1651–1667, 1996.

[7] Diaz, R.S., Sabino, E.C., Mayer, A., Mosley, J.W., Busch, M.P. and The Transfusion Study Group. Dual human immunodeficiency virus type-1 infection and recombination in a dually exposed transfusion recipient. *J. Virol*. **69**: 3273–3281, 1995.

[8] Carr, J.K., Salminen, M.O., Koch, C., Gotte, D., Artenstein, A.W., Hegerich, P.A., St. Louis, D., Burke, D.S. and McCutchan, F.E. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol*. **70**:5935–5943, 1996.

[9] Gao, F., Robertson, D.L., Morrison, S.G., Hui, H., Craig, S., Decker, J., Fultz, P.N., Girard, M., Shaw, G.M., Hahn, B.H. and Sharp, P.M. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol*. **70**: 7013–7029, 1996.

[10] Cornelissen, M., Kampinga, G., Zorgdrager, F., Goudsmit, J. and UNAIDS Network for HIV Isolation and Characterization. Human immunodeficiency virus type 1 subtypes defined by *env* show high frequency of recombinant *gag* genes. *J. Virol*. **70**:8209–8212, 1996.

[11] Bobkov, A., Cheingsong-Popov, R., Salminen, M., McCutchan, F., Louwagie, J., Ariyoshi, K., Whittle, H. and Weber, J. Complex mosaic structure of the partial envelope sequence from a Gambian HIV Type 1 isolate. *AIDS Res. Hum. Retrovirus*. **12**:169–171, 1996.

[12] Salminen, M.O., Carr, J.K., Robertson, D.L., Hegerich, P., Gotte, D., Koch, C., Sanders-Buell, E., Gao, F., Sharp, P.M., Hahn, B.H., Burke, D.S. and McCutchan, F.E. Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. *J. Virol*. **71**: 2647–2655, 1997.

[13] Wieland, U., Seelhoff, A., Hofmann, A., Kuhn, J.E., Eggers, H.J., Mugyenyi, P. and Schwander, S. Diversity of the vif gene of human immunodeficiency virus type 1 in Uganda. *J. Gen. Virol*. **78**:393–400, 1997.

[14] Choi, D.J., Dube, S., Spicer, T.P., Slade, H.B., Jensen, F.C. and Poiesz, B.J. HIV type 1 isolate Z321, the strain used to make a therapeutic HIV type 1 immunogen, is intersubtype recombinant. *AIDS Res. Hum. Retrovirus*. **13**: 357–361, 1997.

[15] Takehisa, J., Oseikwasi, M., Ayisi, N.K., Hishida, O., Miura, T., Igarashi, T., Brandful, J., Ampofo, W., Netty, V.B.A., Mensah, M., Yamashita, M., Ido, E. and Hayami, M. Phylogenetic analysis of human immunodeficiency virus 1 in Ghana. *Acta. Virology* **41**:51–54, 1997.

[16] Gao, F., Robertson, D.L., Carruthers, C.D., Morrison, S.G., Jian, B., Chen, Y., Barre-Sinoussi, F., Girard, M., Srinivasan, A., Abimiku, A.G., Shaw, G.M., Sharp, P.M., and Hahn, B.H. A comprehensive panel of near full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol*. (Submitted).

[17] Siepel, A.C., Halpern, A.L., Macken, C. and Korber, B.T.M. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res. Hum. Retrovirus*. **11**:1413–1416, 1995.

[18] Salminen, M., Carr, J.K., Burke, D.S. and McCutchan, F.E. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retrovirus*. **11**:1423–1425, 1995.

[19] Robertson, D.L., Hahn, B.H. and Sharp, P.M. Recombination in AIDS viruses. *J. Mol. Evol.* **40**:249–259, 1995.

[20] Alizon, M., Wain-Hobson, S., Gluckman, J.-C. and Sonigo, P. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* **46**:63–74, 1986.

[21] Li, W.-H., Tanimura, M. and Sharp, P.M. Rates and dates of divergence between AIDS virus nucleotide sequences. *Genetics* **116**:s44, 1987.

[22] Li, W.-H., Tanimura, M. and Sharp, P.M. Rates and dates of divergence between AIDS virus nucleotide sequences. *Molec. Biol. Evol*. **5**:313–330, 1988.

[23] Zhu, T., Korber, B.T., Nahmias, A.J., Hooper, E., Sharp, P.M. and Ho, D.D. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**:594–597, 1998.

[24] Kostrikis, L.G., Bagdades, E., Cao, Y., Zhang, L., Dimitriou, D. and Ho, D.D. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J. Virol*. **69**:6122–6130, 1995.

[25] Gao, F., Robertson, D.L., Carruthers, C.D., Salminen, M., Kostrikis, L., Ho, D.D., Sharp P.M., and Hahn B.H. Unpublished.

[26] Gao, F., Yue, L., Robertson, D.L., Hill, S.C., Hui, H., Biggar, R.J., Neequaye, A.E., Whelan, T.M., Ho, D.D., Shaw, G.M., Sharp, P.M. and Hahn, B.H. Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J. Virol*. **68**: 7433–7447, 1994.

[27] Jin, M.J., Hui, H., Robertson, D.L., Muller, M.C., Barre-Sinoussi, F., Hirsch, V.M., Allan, J.S., Shaw, G.M., Sharp, P.M. and Hahn, B.H. Mosaic genome structure of simian immunodeficiency virus from West African green monkeys. *EMBO J*. **13**:2935–2947, 1994.

[28] Chen, Z., Telfer, P., Gettie, A., Reed, P., Zhang, L., Ho, D.D. and Marz, P.A. Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J. Virol*. **70**:3617–3627, 1996.

[29] Takehisa, J., Zekeng, L., Miura, T., Ido, E., Yamashita, M., Mboudjeka, I., Gurtler, L.G., Hayami, M. and Kaptue, L. Triple HIV-1 infection with group O and Group M of different clades in a single Cameroonian AIDS patient. *J. AIDS Human Retrovirol*. **14**:81–2, 1997.

[30] Grez, M., Dietrich, U., Balfe, P., von Briesen, H., Maniar, J.K., Mahambre, G., Delwart, E.L., Mullins, J.I., Rubsamen-Waigmann, H. 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, 1994.

[31] Liu, S.-L., Mulvania, T., He, X., Kosloff, B., Rodrigo, A., Corey, L. and Mullins, J.I. Human immunodeficiency virus type 1 recombination in vivo associated with selection advantage. (In preparation.)

[32] Zhu, T.F., Wang, N., Carr, A., Wolinsky, S., and Ho, D.D. Evidence for co-infection by multiple strains of Human Immunodeficiency Virus Type 1 subtype B in an acute seroconvertor. *J. Virol*. **69**:1324–1327, 1995

[33] Howard, T.M. and Rasheed, S. Genomic structure and nucleotide sequence analysis of a new HIV type 1 subtype A strain from Nigeria. *AIDS Res. Hum. Retrovirus*. **12**:1413–1425, 1996.