

The Molecular Epidemiology of Human Immunodeficiency Virus Type 1 in Edinburgh

Edward C. Holmes,* Lin Qi Zhang,* Pamela Robertson, Alexander Cleland, Elizabeth Harvey, Peter Simmonds, and Andrew J. Leigh Brown

Centre for HIV Research, Division of Biological Sciences, and Department of Medical Microbiology, Medical School, University of Edinburgh, Scotland

Human immunodeficiency virus (HIV) type 1 sequences obtained from HIV-infected persons in different risk groups in Edinburgh were studied to determine the number and origin of virus variants and patterns of virus transmission. Phylogenetic analysis revealed that 12 of 14 hemophilic patients who had been exposed to a single common batch of factor VIII had closely related *gag* gene sequences. Sequences from intravenous drug users and patients infected through heterosexual contact formed another distinct group, and 2 other hemophiliacs formed a third group. However, epidemiologic relationships inferred from analysis of the V3 region of the *env* gene were less conclusive, especially when the V3 loop was taken in isolation. This appears to be due to the length of time since infection and the action of selection, which has favored the independent appearance of similar V3 loop variants.

Phylogenetic analysis of human immunodeficiency virus (HIV) nucleotide sequence data has been important in studies of the origin and global distribution of HIV [1]. Such analysis has also been used to establish the likelihood of HIV transmission from infected health care workers to patients [2–4]. In contrast, only recently have phylogenetic methods been used to determine the transmission networks of HIV within a single community [5, 6]. Such analysis may, however, provide information as to how many HIV variants established the epidemic in a particular community, whether different transmission groups possess characteristic variants, and how the distribution of these variants changes with time. In this study we present the results of a detailed phylogenetic analysis of HIV sequence variants in Edinburgh.

Intravenous drug users (IVDUs) are the risk group with the largest number of HIV-1-infected patients in Scotland. By the end of March 1993, the cumulative total of reported HIV-1 positive persons in Scotland was 1921, of which 942 (~49%) were thought to have become infected through intravenous drug use [7]. This proportion is much higher than that found in the rest of the United Kingdom, where on average <15% of HIV infections within any community are related to drug use. The difference is particularly pro-

nounced in Edinburgh (Lothian region), which accounts for ~52% of all HIV infections in Scotland: ~54% of these infections are thought to have been acquired through intravenous drug use. There is a similarly high incidence of HIV infection in IVDUs in Dundee (Tayside region). The numbers of seropositive IVDUs in Edinburgh increased dramatically in the last quarter of 1983 and the first quarter of 1984 [8]. Retrospective testing suggested that HIV first entered the Edinburgh IVDU community in 1983 and was present in Dundee by 1984 and in Glasgow (Strathclyde region) by 1985 [9].

Little is known, however, about the nature of the virus variants involved in the HIV epidemic in Edinburgh. We previously reported nucleotide sequence data from members of a small cohort of hemophilia A patients infected by non-commercial factor VIII in 1984 [10–13]. Analysis of transfusion records implicated one common batch as the likely source of infection. Eighteen of 32 patients who received this batch seroconverted [10]. Phylogenetic analysis of the V4 and V5 regions of the *env* gene supported this contention because sequences from most of the patients studied (referred to as the main hemophilic cohort) grouped closely together on phylogenetic trees [11, 12]. Two patients (p82 and p74), however, had more divergent sequences even though they also received factor VIII from the implicated batch. Of these 2 patients, p82 shared identical sequences in the V3 and V4 regions of gp120 with another HIV-infected hemophilic (p80) who had never received factor VIII from the main implicated batch [13].

To clarify the number and origin of the batches of factor VIII involved in the HIV infection of the Edinburgh hemophilic population, we extended our studies by including sequences obtained from the p17 region of the *gag* gene. This region was chosen for phylogenetic analysis because many of the nucleotide substitutions observed are synonymous: Such

Received 9 February 1994; revised 2 August 1994.

Financial support: Medical Research Council (UK) AIDS Directed Programme and Committee for Epidemiological Studies on AIDS (G9209918, SPG9027555).

Reprints or correspondence: Dr. Andrew J. Leigh Brown, Centre for HIV Research, ICAPB, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, UK.

* Current affiliations: Dept. of Zoology, University of Oxford, UK (E.C.H.). Aaron Diamond AIDS Research Center, New York (L.Q.Z.).

The Journal of Infectious Diseases 1995;171:45–53
© 1995 by The University of Chicago. All rights reserved.
0022-1899/95/7101-0008\$01.00

substitutions are generally regarded as reliable markers of evolutionary history [14, 15]. Although p17 sequences evolve more slowly than regions of the *env* gene more commonly used in phylogenetic studies of HIV-1, there does appear to be sufficient information to resolve many evolutionary relationships [4]. Another part of the *gag* gene, the p7/p9 coding region, has provided important information about the number of distinct HIV variants circulating in Finland [6]. So that a more detailed investigation of the evolution of HIV within a single community could be undertaken, homologous sequence data were also obtained from a number of local IDUs and persons thought to have been infected through heterosexual contact. For comparative purposes, sequences of the V3 region of the *env* gene from the same patients were analyzed similarly to investigate the patterns of evolution in a more rapidly evolving region of the genome.

Materials and Methods

Origin of HIV sequence data. The numbers of unique plasma RNA and lymphocyte-associated DNA *gag* p17 and V3 *env* sequences obtained from each patient are presented in table 1. These patients are believed to have been infected with HIV in Edinburgh with the exception of Sc2, a patient from Dundee who was infected through heterosexual contact. One hemophiliac, p95, became infected after exposure to commercial blood products prepared outside of Scotland (Watson H., personal communication). All hemophiliacs except p95 seroconverted in 1984, as did all the IDUs. Patient p95 seroconverted in 1982. The patients thought to have been infected through heterosexual contact, Sc1, Sc2, and Sc3, seroconverted in 1988, 1991, and 1992, respectively.

Primers. All primers used in this study were synthesized by Oswel DNA Service (Department of Chemistry, University of Edinburgh) and purified by high-pressure liquid chromatography (HPLC). Primer sites were chosen for the greatest conservation between published sequences of HIV-1 taken from the Los Alamos HIV sequence data base [18]. No more than one mismatch with any published sequence was permitted, and no mismatch near the 3' terminus was permitted. The sequences of the *gag* primers, with coordinates relative to HIV_{HTLV-III} (clone HXB2R) given in parentheses, are as follows: *gag*1, 5'-GCGA-GAGCGTCAGTATTAAGCGG (sense, 342); *gag*2, 5'-GGG-AAAAAATTCGGTTAAGGCC (sense, 382); *gag*3, 5'-CTTCT-ACTACTTTTACCCATGC (antisense, 795); *gag*4, 5'-TCTGATAATGCTGAAAACATGGG (antisense, 843).

Nucleic acid purification and amplification by nested polymerase chain reaction after limiting dilution and direct sequencing (primers *gag*2 and *gag*3) were done as described [12] using DNA from 2×10^6 lymphocytes or viral RNA from at least 500 μ L of plasma. Viral RNA samples were treated with RNase-free DNase (Boehringer Mannheim, Lewes, UK) prior to reverse transcription with the antisense primer *gag*4. The sequences of the *env* primers have been presented elsewhere [16].

Sequence analysis. A total of 186 sequences (average length, 343 bp) were obtained from the *gag* gene of HIV-1 from infected persons in Edinburgh. The region begins at amino acid 40 in

Table 1. Origin and number of HIV-1 nucleotide sequences obtained from patients in Edinburgh.

Group, patient	p17		V3	
	RNA	DNA	RNA	DNA
Hemophiliacs				
p28	—	9*	—	15*†
p56	7‡	12‡	—	—
p58	3‡	—	—	—
p70	3‡	—	—	—
p72	—	—	—	8‡
p74	15*	2‡	40*	6‡
p77	—	15*‡	—	17*‡
p79	—	11*‡	—	15*‡
p80	—	2‡	—	8‡
p82	14*	—	67‡	—
p84	3‡	8*	11*	11*
p87	—	2‡	—	9‡
p89	—	2‡	—	—
p91	—	2‡	—	5‡
p95	—	1‡	—	—
p110	3‡	—	—	—
p180	3‡	—	—	—
Heterosexual contact				
Sc1	21*‡	—	26*	—
Sc2	18*	—	29*	—
Sc3	12*	7*	24*	29*
Intravenous drug users				
IDU2	—	1‡	—	—
IDU5	—	2‡	—	—
IDU6	—	1‡	—	—
IDU9	—	2‡	—	—
IDU10	—	2‡	—	—
IDU11	3‡	—	—	—

Sequences: *from Zhang et al. [16], †from Balfe et al. [11], ‡generated in this study, §from Simmonds et al. [12], ¶from Simmonds et al. [13], *from Holmes et al. [17].

p17 and extends to the 5' 60 bases in p24 (homologous with positions 451–793 in HXB2r). Many of the patients were represented by multiple sequences, but when identical sequences were obtained from the same source, they were represented only once in the phylogenetic analysis for ease of computation. Thus, there were 101 sequences. Some of these *gag* sequences have been published (see table 1 for references and accession numbers); sequences generated in this study have been assigned GenBank accession numbers L35950–L36017.

Fifty-four unique sequences were obtained from the V3 region of the *env* gene from 12 of the hemophiliac patients (table 1). Additional unique sequences (67) taken from patient p82 during 7 years of infection [17] were also included. The average length of the region analyzed was 237 bp (homologous to HXB2r positions 6598–6834). All these sequences have been published elsewhere (table 1). Finally the V3 loop, which corresponds to the principal neutralizing determinant of HIV-1 [19], was extracted from the complete V3 region of the sequences described above and analyzed separately. Seventy-six unique V3

loop sequences (~105 bp long and homologous to HXB2r positions 6656–6763) were available.

A number of sequences from published isolates of HIV-1 were used to provide a background to the Edinburgh population: HIV_{CDC4}, HIV_{SF2}, HIV_{MN}, HIV_{JRCSE}, and HIV_{NY5CG} (USA); HIV_{ELI} (Zaire); HIV_{HAN} (Germany); HIV_{JH3} (Japan); HIV_{LAI} (France); HIV_{OVI} (Gabon); and HIV_{RF} (Haiti). All these sequences were taken from the Los Alamos HIV sequence data base [18] and, with the exception of HIV_{ELI}, have been assigned to HIV-1 subtype B [1]. HIV_{ELI} is a member of subtype D [1, 18] and was used as an outgroup to root the phylogenetic trees. By combining the (unique) Edinburgh and the published sequences, data sets consisting of 112, 132, and 87 sequences were available for phylogenetic analysis from the p17 region, complete V3 region, and V3 loop, respectively.

Nucleotide and translated amino acid sequences were aligned by the CLUSTALV program [20] as implemented in the GDE package (version 2.0; Harvard Genome Laboratory, Boston). Length varied in the 3' region of the p17 coding sequence (between nucleotides 687 and 713 of the HXB2r sequence); alignments were checked visually. After alignment, a 382-bp sequence (including gaps) from the p17 region was available for phylogenetic analysis. Less internal length variation was observed in the V3 region, and 246 bp were used for phylogenetic analysis after alignment; 111 bp from the V3 loop were analyzed after alignment.

Phylogenetic analysis was done using programs from version 3.5c of the PHYLIP package [21]. Because of the large numbers of sequences obtained, phylogenetic trees were reconstructed using the neighbor-joining clustering method [22] (program, NEIGHBOR). Nucleotide sequence distances were estimated under a model that allows different rates of transition and transversion and different frequencies of the four nucleotides [21] (program, DNADIST). Trees were run with different input orders and with different subsets of data. To test the significance of the groupings obtained and to assess the relative likelihood of other hypotheses of epidemiologic importance, different phylogenetic trees, representing possible epidemiologic relationships, were artificially constructed (program, RETREE) and given log likelihoods (program, DNAML). Log likelihoods were then compared using the test of Kishino and Hasegawa [23], which assigns significance between trees ($P < .01$) if the mean difference in log likelihood is >1.96 SD. The method of bootstrap resampling (100 replications) was used on neighbor-joining trees (programs, SEQBOOT and CONSENSE) to place approximate confidence limits on individual branches.

Results

Phylogenetic analysis of gag p17 sequence data. To estimate the number and origin of HIV-1 variants in Edinburgh, comparison with sequences taken from a wider geographic background was necessary. Comparison of all Edinburgh *gag* and *env* sequences with consensus sequences representing the 5–7 reported subtypes of HIV-1 [1, 18] clearly identified these sequences as being members of subtype B, the predominant subtype found in North America and Europe. There-

fore, to understand the distribution of HIV variants in Edinburgh in more detail, phylogenetic analyses were done with a background of 10 published subtype B sequences obtained from various parts of the world (see Materials and Methods). Sequences from a subtype D Zairian isolate, HIV_{ELI} [1, 18], were also included to root the phylogenetic trees.

The rooted neighbor-joining tree for the 112 unique *gag* p17 sequences from the 25 Edinburgh patients and the sequences of 11 published isolates is presented in figure 1. In most cases, p17 sequences from the same patient clustered most closely with others from the same patient. The sequences from different patients clearly fall into a number of distinct groups.

This analysis confirmed the close clustering of sequences from most of the hemophiliac patients (p28, p56, p58, p70, p77, p79, p84, p87, p89, p91, p110, and p180: the main hemophiliac cohort), some of which had been identified as sharing sequences in the V4 region [11, 12]. The analysis also demonstrated that hemophiliacs p80 and p82, although also infected from locally prepared factor VIII, grouped away from the main hemophiliac cohort, even though p82 had received factor VIII from a small number of bottles from the common implicated batch. Another hemophiliac, p74, was separated from the main cohort by a single published sequence from the United States, HIV_{NY5CG}. Furthermore, the hemophiliac patient known to have been infected following exposure to commercial factor VIII, p95, occupies a distinct phylogenetic position.

In contrast to the multiple origins of HIV infection found in the Edinburgh hemophiliac population, the sequences from IVDU patients fall into a single group (figure 1). Sequences from 3 patients infected by heterosexual contact also fell into this group, even though one of these patients seroconverted in Dundee, ~110 km from Edinburgh.

These groupings were also found in a bootstrap resampling analysis (table 2), although the branches supporting the common ancestry (monophyly) of the heterosexual-IVDU group and the main hemophiliac cohort (excluding p74 and p82) were found in only 28% and 33% of replicates, respectively. This suggests that more sequence data are required before these groupings can be confidently assigned, although the low bootstrap support for the main hemophiliac cohort is partly due to the ambiguous position of p74 (see below and Discussion). A second neighbor-joining analysis was done on the 101 sequences obtained from the 25 Edinburgh patients (i.e., published isolates removed; trees not shown). The major groupings observed were the same as those in the full data set, but the level of supporting bootstrap replicates was generally greater, especially for the heterosexual-IVDU and p80-p82 groupings (found in 84% and 89% of replicates, respectively; table 2).

Thus, there is evidence that there are several HIV variants within Edinburgh that are associated with different risk groups. These results are particularly interesting because the

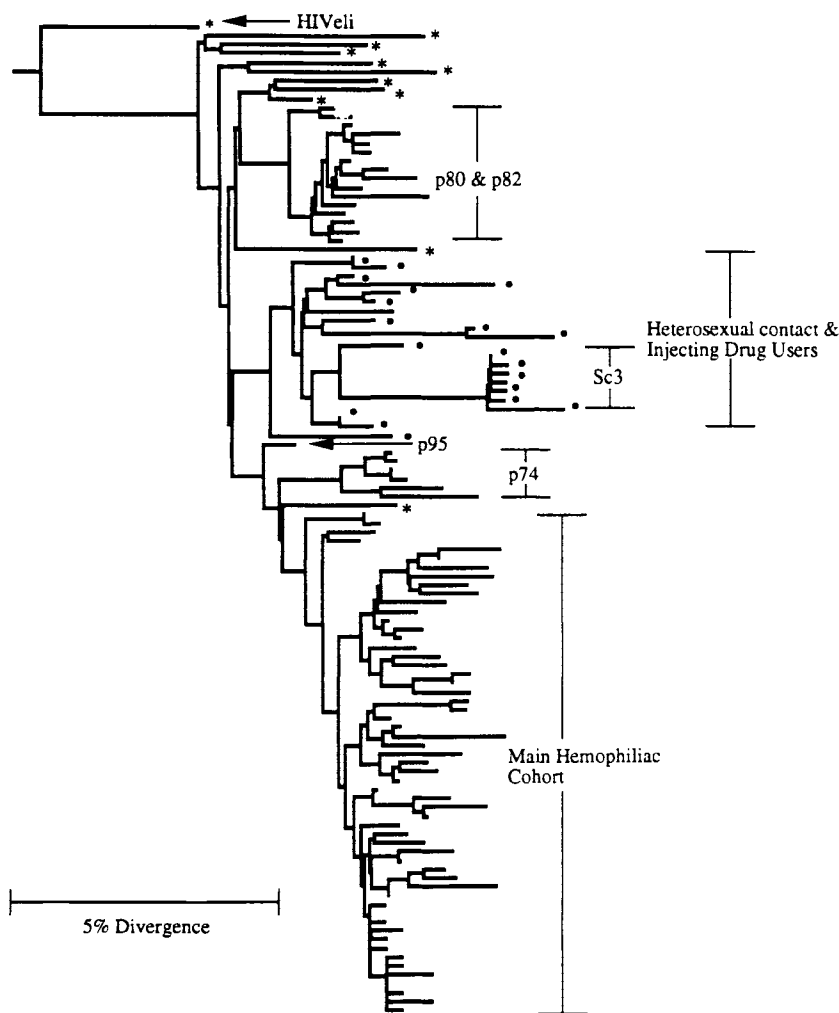


Figure 1. Rooted neighbor-joining tree for 112 unique *gag* p17 sequences taken from 25 Edinburgh patients, 10 worldwide isolates of HIV-1 subtype B, and subtype D isolate HIV_{ELI}, which also serves as an outgroup. Branches without symbols at tip = sequences obtained from hemophiliac patients; ● = sequences from heterosexually infected persons and intravenous drug users; * = sequences of published worldwide isolates of HIV-1 subtype B; Sc = patient with heterosexual contact; p = hemophiliac patients. Horizontal branch lengths drawn to scale.

positions of hemophiliac patients p74 and p82 are away from the main hemophiliac cohort even though they received factor VIII from the implicated batch of factor VIII and because the sequences found in patients infected by heterosexual contact grouped with those found in IVDUs. To test more rigorously these and other hypotheses, trees were compared using a maximum likelihood significance test. This involved assigning log likelihoods to the trees obtained in the neighbor-joining analysis and to trees that depict the following alternative epidemiologic relationships: (1) both p74 and p82 are members of the main hemophiliac cohort and (2) the sequences from patients infected through heterosexual contact and IVDUs do not group together. Alternative trees can then be compared using the significance test of Kishino and Hasegawa [23].

It is clear from the results on alternative 1 that we cannot separate p74 from the main hemophiliac cohort with confidence, since a tree that places p74 with the main cohort actually has a slightly higher likelihood (table 3) than the tree that depicts this patient as being a separate transmission (fig-

ure 1). On the other hand, there does appear to be good evidence that p82 is not a member of the main cohort because the optimal tree depicting this relationship was significantly less likely than one that grouped p82 with p80 (table 3 and data not shown). Patient p80 was never exposed to factor VIII from the batch implicated in the infection of the main cohort but shared at least one other batch with p82.

Alternative 2 was tested by moving the sequences taken from the patient infected by heterosexual contact to different places on the phylogenetic tree and estimating the log likelihoods of each topology as before. In all cases, these trees had lower likelihoods than those placing this group with the IVDUs, although only in some cases was there a significant difference between these trees at the $P < .01$ level using the Kishino and Hasegawa test [23] (table 3).

We also examined whether HIV has entered Edinburgh on one occasion only (single origin) or more than once (multiple origin). The major groups of Edinburgh sequences described above were sometimes separated from each other by sequences obtained from persons infected outside the city

Table 2. Numbers of bootstrap replications supporting different epidemiologic groupings of HIV-1 in Edinburgh.

Grouping	Supporting bootstrap replication (%)	
	All sequences*	Edinburgh sequences†
p74 as an independent transmission	61	74
p82 with p80	76	89
Heterosexuals with IVDUs	28	84
Main hemophilic cohort, monophyletic	33	35

NOTE. Data are inferred from p17 sequences. IVDUs = intravenous drug users.

* Including 10 worldwide subtype B sequences and HIV_{ELI} (total, 112).

† Excluding 10 worldwide subtype B sequences and HIV_{ELI} (total, 101).

(worldwide isolates of subtype B), suggesting a multiple origin. However, in this analysis the multiple origin is not significantly more likely than a single origin (table 3).

Phylogenetic analysis of the V3 region. A rooted neighbor-joining tree for 132 unique V3 sequences from 12 Edinburgh patients and 11 sequences from published worldwide isolates of HIV-1 is presented in figure 2. There is considerable resemblance between the epidemiologic relationships shown in this analysis and those from the phylogenetic trees of the p17 sequences. Of particular note, sequences obtained from members of the main hemophilic cohort tend to cluster together, although there is a generally less coherent picture of relationships: some of the sequences obtained from p79 fall some distance from the remaining main cohort sequences, and there is a wider mix of sequences from individ-

ual patients than seen in p17. The position of p74 also appears to lie outside the main cohort by likelihood analysis although, as in p17, this was not significant in the Kishino-Hasegawa test [23]. Of interest, only a minority of the V3 region sequences obtained from p80 cluster with those from p82; the remainder group closer to the main cohort even though this patient never received factor VIII from the main implicated batch. The separation of p82 from the main cohort was significant in the maximum likelihood analysis (table 3). Only 2 of the 3 heterosexually infected patients grouped together (Sc2 and Sc3); the separation of the third patient was not significant.

The most intriguing feature of this tree, however, and one that is in marked contrast to the relationships observed in the phylogenetic analysis of the p17 region, is that the sequences obtained over 7 years from patient p82 appear not to have a single origin (figure 2). We showed in our previous phylogenetic analysis of sequence sets according to the time they arose that all sequences obtained from p82 could be derived from the single V3 sequence observed at seroconversion [17]. The diversity that developed during the infection of p82 is even more apparent when the V3 loop is analyzed in isolation (see below). This demonstrates that the extent of V3 variation that develops within a person over several years can reach the same order as that existing between persons.

Phylogenetic analysis of the V3 loop. The rooted neighbor-joining tree for the 87 unique sequences from the V3 loop of the *env* gene is presented in figure 3. This tree does not depict the correct phylogenetic relationships among the HIV-infected patients in Edinburgh. In particular, V3 loops identical at the amino acid level are found in unrelated pa-

Table 3. Likelihoods of four pairs of alternative or competing hypotheses of epidemiologic relationship of HIV-1 in Edinburgh.

Competing hypotheses	Log likelihood		
	p17	All V3	V3 loop
p74 as an independent transmission vs. p74 with main hemophilic cohort	-3336.249	-3901.909	-1399.779*
p82 with p80 vs. p82 with main hemophilic cohort	-3336.249*	-3901.909*	-1399.779*
Heterosexuals with IVDUs vs. Heterosexuals separate from IVDUs	-3336.249	-3903.513†	-1398.705†
Edinburgh population multiple origin vs. Edinburgh population single origin	-3336.249	-3881.483	-1413.138*
	-3352.041	-3891.792	-1460.912

* Significantly better than that of alternative ($P < .01$).

† Only includes persons infected through heterosexual contact; sequences for intravenous drug users (IVDUs) not available.

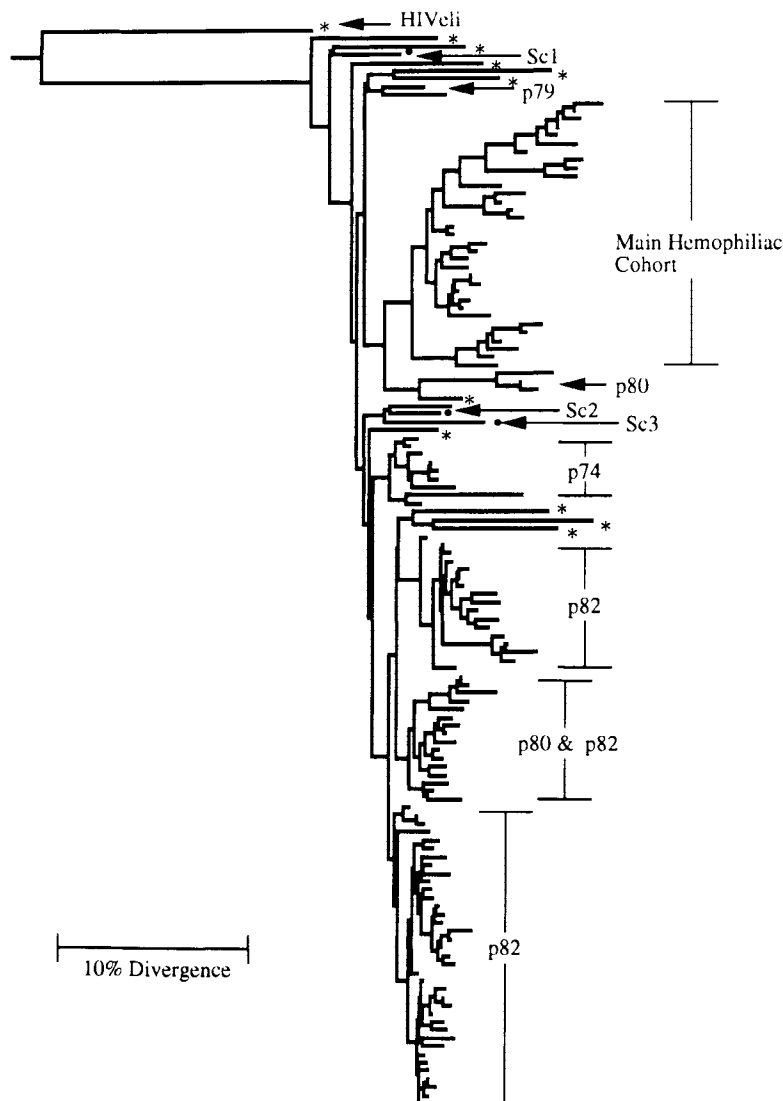


Figure 2. Rooted neighbor-joining tree for all 132 unique V3 sequences. Branches without symbols at tip = sequences obtained from hemophiliac patients; ● = sequences from heterosexually infected persons and intravenous drug users; * = sequences of published worldwide isolates of HIV-1 subtype B; Sc = patients with heterosexual contact; p = hemophiliac patients. Horizontal branch lengths drawn to scale.

tients (namely, p82 and p74, p82 and Sc3). From these data, all patients from whom more than one sequence was obtained appeared to have multiple origins of their virus population. Furthermore, many of the groupings of patients conflict with both epidemiologic evidence and that deduced from the phylogenetic analysis of the *gag* gene. The most notable conflict is that all members of the main hemophiliac cohort do not group together. Similar results were obtained in both the neighbor-joining bootstrap and maximum likelihood analyses (data not shown). Thus, the variation observed in the V3 loop does not always reflect the history of transmission events between patients. Therefore, it appears that there is not sufficient information in this region alone to resolve the epidemiologic relationships among infected persons. Such a finding is in accord with studies suggesting that the evolution of the V3 loop reflects the action of complex adaptive forces, such as the escape from immunologic recognition and the need to infect a variety of cell types [24–28].

Discussion

Development of a molecular marker. We have shown that nucleotide sequences from the p17 region of the *gag* gene are a reliable indicator for the divergence of HIV-1 sequences within an infected community. First, we observed that there were no major discrepancies between the transmission groups depicted in the phylogenetic analysis of the sequence data and those previously established on clinical and epidemiologic grounds. Second, we showed that no inconsistencies arose within the molecular data due to the use of viral nucleic acid from different sources. Thus, patients who were thought to be infected from a single batch of factor VIII generally grouped together in the phylogenetic analysis (although not with a significant percentage of supporting bootstrap replications) regardless of whether sequences were obtained from plasma viral RNA, lymphocyte-associated proviral DNA, seroconversion samples, or from samples taken ≤ 7 years after seroconversion. Furthermore, most se-

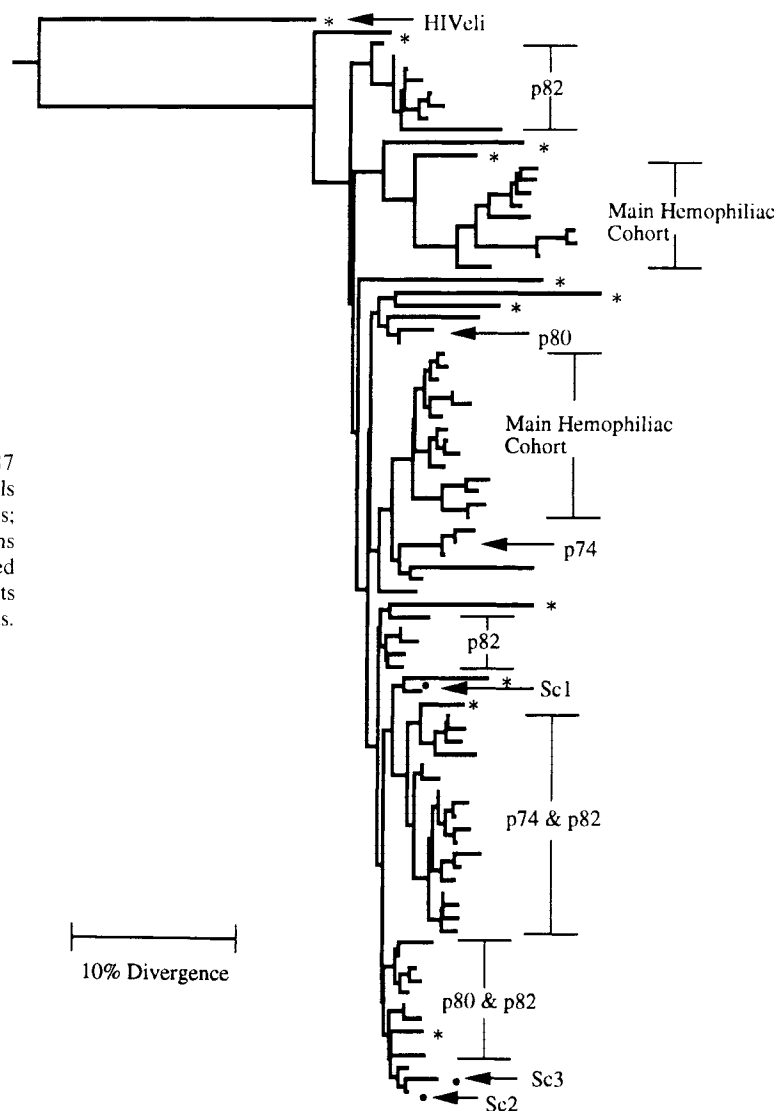


Figure 3. Rooted neighbor-joining tree for all 87 unique V3 loop sequences. Branches without symbols at tip = sequences obtained from hemophiliac patients; ● = sequences from heterosexually infected persons and intravenous drug users; * = sequences of published worldwide isolates of HIV-1 subtype B; Sc = patients with heterosexual contact; p = hemophiliac patients. Horizontal branch lengths drawn to scale.

quences from individuals grouped together on the phylogenetic trees. The exceptions were obtained from persons whose sequences probably diverged within a very short time (e.g., some members of the heterosexual-IVDU group and the main hemophiliac cohort). It will probably be difficult to resolve epidemiologic relationships associated with such rapid transmission events.

One patient, p95, is known to have been infected by commercial blood products prepared outside Edinburgh. The fact that this patient occupies a distinct position on the phylogenetic tree confirms that there is information of epidemiologic value in the p17 region of *gag* and means that this region may be a useful indicator of phylogenetic relationships in future studies of the evolution of HIV within infected communities.

Molecular epidemiology of HIV-1 in Edinburgh. The phylogenetic analysis of the *gag* p17 sequences presented here is a significant new approach to the reconstruction of the

spread of HIV within a community. The most striking finding was that there are several distinct HIV variants circulating in Edinburgh. The hemophiliac patients appear to be divided into a number of distinct groups whose members were infected by at least two batches of contaminated factor VIII. This is particularly surprising since most of these patients seroconverted at about the same time (spring 1984) and were originally thought to have been infected after exposure to a single common batch [10].

There is no evidence from this study that p82 was a member of the main cohort. However, in the analysis of p17, there was a strong and statistically significant association of p82 with p80, a patient who did not receive any of the main implicated batch. In addition, V3 sequences found at seroconversion in p82 were identical with some of those from p80. Therefore, the similar viruses found in p80 and p82 must have been derived from a factor VIII batch that these 2 patients are known to have shared.

Multiple infected donations to the main implicated batch or polymorphism within the contaminating virus pool could, in principle, account for the divergent position of p74. It is significant, however, that in the present study and others [11, 12], p74 not only possesses variants that are different from those found in the main cohort in samples taken a number of years after transmission but was already distinct at seroconversion in an analysis of the V3 and V4 regions of the *env* gene [16]. Such long-standing distinction could indicate an independent infection from an epidemiologically linked donor. Thus, there appear to have been 2 or 3 HIV-infected donors contributing to the local plasma pool at the time these batches were prepared (1983). This shows there was substantial viral diversity during the early stages of the HIV epidemic in Scotland even though only 53 retrospective HIV-1 antibody tests were reported positive in 1983, and the first reported AIDS patient in Scotland was a homosexual male in 1984 [9].

Whereas a very close similarity was suspected in the sequences from hemophiliacs apparently infected by exposure to a common batch of factor VIII, it was surprising to detect a single evolutionary origin for viruses infecting heterosexuals and IVDUs, especially given the potentially large number of transmission events and the time range over which these patients seroconverted. Although the number of infected IVDUs sampled is small relative to the total number in Edinburgh, the phylogenetic analysis presented here shows it is possible that the epidemic in this group may have been founded by a single HIV variant. This conclusion is compatible with the rapid spread of the virus in the Edinburgh IVDU population in late 1983 and early 1984 [8]. In this context it is also significant that viruses from recently seroconverted heterosexual patients are also clearly related to the other members of this group.

Thus, it appears that the variants that are currently spreading through heterosexual contact derive from the virus variant that first infected the IVDU population in 1983–1984. We are currently examining this hypothesis in more detail. The fact that the heterosexual and IVDU groups are some distance from the hemophiliac groups on all the phylogenetic trees shows that the HIV infections in these populations were independent and refutes suggestions that the hemophiliacs could have been infected from the IVDU community.

Evolution of the V3 loop within a population. In contrast to the simple divergence observed in the p17 region, the evolution of the V3 region is characterized by greater diversity among sequences from the main hemophiliac cohort and by the observation that sequences obtained from patients who have been infected for several years (e.g., p82) do not always group together. In addition, identical amino acid sequences in the V3 loop are found in epidemiologically unlinked patients. This is especially true during primary infection [16], as typified by the identical seroconversion sequences observed in p82, a hemophiliac who seroconverted in 1984,

and Sc3, an unlinked heterosexually infected patient who seroconverted in 1992. The similarity between these V3 loops is even more striking because of the overall distinctiveness in the regions flanking the loop [16] and their very different p17 sequences (figure 1). Thus, nucleotide changes that do not always reflect the evolutionary history of the virus have become fixed in the V3 loop.

These observations may be interpreted to be the outcome of the action of natural selection. While the evolution of the V3 region within a patient appears to be characterized by the generation of substantial amounts of diversity and changes of selective importance [25–29], such diversity is not always reflected in the evolution of V3 at the population level. A similar observation was made in an intensive study of HIV-infected persons in Amsterdam [5]. Although a small increase in diversity in the complete V3 region was detected from 1980 to 1991, very similar V3 loop amino acid sequences were found throughout this period, and there was no evidence of a change in the consensus sequence. This restriction in diversity may be viewed as a consequence of the strength of the selective forces proposed to act during primary infection (or transmission), forces that produce an extremely homogeneous V3 loop population within individuals and generally similar sequences between persons [16]. The principle conclusions from these observations are that phylogenetic reconstructions based on V3 loop sequences alone should not be used to infer epidemiologic relationships among HIV-infected patients and that those based on the V3 region will be most reliable in the first few years of infection.

Acknowledgments

We thank C. A. Ludlam and the staff of the Haemophilia Centre (Royal Infirmary of Edinburgh), J. R. Robertson (Muirhouse Medical Group, Edinburgh), R. P. Brettell (City Hospital, Edinburgh), A. J. France (Kings Cross Hospital, Dundee), and G. E. D. Urquhart (Ninewells Hospital, Dundee) for providing samples; Pamela Barr, Selma Rebus, and laboratory staff for sample preparation and storage (Department of Medical Microbiology, University of Edinburgh); and Catriona Baker for assisting with sequencing and Katherine Fielding for providing epidemiologic data (Centre for HIV Research).

References

1. Louwagie J, McCutchan FE, Peeters M, et al. Phylogenetic analysis of *gag* genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* 1993;7:769–78.
 2. Albert J, Wahlberg J, Uhlen M. Forensic evidence by DNA sequencing. *Nature* 1993;361:595–6.
- Ou CY, Ciesielski CA, Myers G, et al. Molecular epidemiology of HIV transmission in dental practice. *Science* 1992;256:1165–71.
- Holmes EC, Zhang LQ, Simmonds P, Rogers AS, Leigh Brown AJ. Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon. *J Infect Dis* 1993;167:1411–4.

5. Kuiken CL, Zwart G, Baan E, Coutinho RA, van den Hoek JAR, Goudsmit J. Increasing antigenic and genetic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of the AIDS epidemic. *Proc Natl Acad Sci USA* 1993;90:9061-5.
6. Salminen M, Nykanen A, Brummer-Korvenkontio H, Kantanen ML, Liitsola K, Leinikki P. Molecular epidemiology of HIV-1 based on phylogenetic analysis of *in vivo gag p7/p9* direct sequences. *Virology* 1993;195:185-94.
7. ANSWER (AIDS News Supplement, CDS Weekly Report). Communicable Disease (Scotland) Unit. Human immunodeficiency virus 1 (HIV-1) quarterly report to 31st March 1993. Glasgow: CD(S)U, Ruchill Hospital, 1993:18.
8. Robertson JR, Bucknall ABV, Welsby PD, et al. Epidemic of AIDS related virus (HTLV-III/LAV) infection among intravenous drug abusers. *Br Med J [Clin Res]* 1986;292:527-9.
9. Communicable Disease (Scotland) Unit. Acquired immune deficiency syndrome and HIV-related disease in Scotland. Report of a working group convened by the Chief Medical Officer, the Scottish office. Glasgow: CD(S)U, Ruchill Hospital, 1993.
10. Ludlam CA, Tucker J, Steel CM, et al. Human T-lymphotropic virus type III (HTLV-III) infection in seronegative haemophiliacs after transfusion of factor VIII. *Lancet* 1985;2:233-6.
11. Balfe P, Simmonds P, Ludlam CA, Bishop JO, Leigh Brown AJ. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations. *J Virol* 1990;64:6221-33.
12. Simmonds P, Balfe P, Ludlam CA, Bishop JO, Leigh Brown AJ. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol* 1990;64:5840-50.
13. Simmonds P, Zhang LQ, McOmish F, Balfe P, Ludlam CA, Leigh Brown AJ. Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 *env* sequences in plasma viral and lymphocyte-associated proviral populations *in vivo*: implication for models of HIV pathogenesis. *J Virol* 1991;65:6266-76.
14. Leigh Brown A, Monaghan P. Evolution of the structural proteins of human immunodeficiency virus: selective constraints on nucleotide substitutions. *AIDS Res Hum Retroviruses* 1988;4:399-407.
15. Li WH, Tanimura M, Sharp PM. Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol Biol Evol* 1988;5:313-30.
16. Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Leigh Brown AJ, Simmonds P. Selection for specific sequences in the external envelope protein of HIV-1 upon primary infection. *J Virol* 1993;67:3345-56.
→ Holmes EC, Zhang LQ, Simmonds P, Ludlam CA, Leigh Brown AJ. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc Natl Acad Sci USA* 1992;89:4835-9.
18. Myers G, Korber B, Berzofsky JA, Smith RF, Pavlakis GN. *Human Retroviruses and AIDS*. Los Alamos, NM: Los Alamos National Laboratory, 1992.
→ LaRosa GJ, Davide JP, Weinhold K, et al. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. *Science* 1990;249:932-5.
20. Higgins DG, Bleasby AJ, Fuchs R. CLUSTALV: improved software for multiple sequence alignment. *CABIOS* 1992;8:189-91.
21. Felsenstein J. PHYLIP (Phylogeny Inference Package) Version 3.5c. Distributed by author. Seattle: Department of Genetics, University of Washington, 1993.
22. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing evolutionary trees. *Mol Biol Evol* 1987;4:406-25.
23. Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J Mol Evol* 1989;29:170-9.
24. Chesebro B, Wehrly K, Nishio J, Perryman S. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell tropic isolates: definition of critical amino acids involved in cell tropism. *J Virol* 1992;66:6547-54.
25. Leigh Brown AJ. Sequence variability in human immunodeficiency viruses: pattern and process in viral evolution. *AIDS* 1991;5:S35-42.
→ Nowak MA, Anderson RM, Mclean AR, Wolfs TFW, Goudsmit J, May RM. Antigenic diversity thresholds and the development of AIDS. *Science* 1991;254:963-9.
27. Schuitemaker H, Koot M, Kootstra NA, et al. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytopathic to T-cell-tropic virus populations. *J Virol* 1992;66:1354-60.
28. Milich L, Margolin B, Swanstrom R. V3 loop of the human immunodeficiency virus type 1 *env* protein: interpreting sequence variability. *J Virol* 1993;67:5623-34.