The Molecular Epidemiology of Human Immunodeficiency Virus-Type 1 in Six Cities in Britain and Ireland

A. J. Leigh Brown,[1,*] D. Lobidel,[1] C. M. Wade,[1] S. Rebus,[1] A. N. Phillips,[2] R. P. Brettle,[3] A. J. France,[4] C. S. Leen,[3] J. McMenamin,[5] A. McMillan,[6] R. D. Maw,[7] F. Mulcahy,[8] J. R. Robertson,[9] K. N. Sankar,[10] G. Scott,[6] R. Wyld,[3] and J. F. Peutherer,[11]

[1]*Centre for HIV Research, Institute of Cell, Animal and Population Biology, and* [11]*Department of Medical Microbiology, University of Edinburgh, Waddington Building, West Mains Road, Edinburgh, EH9 3JN Scotland:* [2]*Department of Public Health Medicine, Royal Free Hospital, London NW3, England;* [3]*Infectious Diseases Unit, City Hospital, Edinburgh, Scotland;* [4]*King's Cross Hospital, Dundee, Scot/and;* [5]*Ruchill Hospital, Glasgow, Scot/and;* [6]*Department of Gen/to-Urinary Medicine, Royal infirmary of Edinburgh, Edinburgh, Scotland;* [7]*Department of Genito-Urinary Medicine, Royal Victoria Hospital, Be/fast, Ireland;* [8]*Department of Genito-Urinary Medicine, St James Hospital, Dublin, Ireland;* [9]*Muirhouse Medical Group, Edinburgh, Scot/and; and* [10]*Department of Genito-Urinary Medicine, Newcastle General Hospital, Newcastle, England*

Received February 21, 1997; accepted June 4, 1997

We have sequenced the p17 coding regions of the *gag* gene from 211 patients infected either through injecting drug use (IDU) or by sexual intercourse between men from six cities in Scotland, N. England, N. Ireland, and the Republic of Ireland. All sequences were of subtype B. Phylogenetic analysis revealed substantial heterogeneity in the sequences from homosexual men. In contrast, sequence from over 80% of IDUs formed a relatively tight cluster, distinct both from those of published isolates and of the gay men. There was no large-scale clustering of sequences by city in either risk group, although a number of close associations between pairs of individuals were observed. From the known date of the HIV-1 epidemic among IDUs in Edinburgh, the rate of sequence divergence at synonymous sites is estimated to be about 0.8%. On this basis we estimate the date of divergence of the sequences among homosexual men to be about 1975, which may correspond to the origin of the B subtype epidemic. © 1997 Academic press

The high genetic diversity which usually characterizes HIV-1 has been exploited in many investigations of postulated linkage between infections. These include the identification of individuals belonging to an infection cluster of individuals attending a dental practice in Florida (Ou et al, 1992; Hillis and Huelsenbeck, 1994), of haemophiliacs in Scotland and Germany (Baife *et al.,* 1990; Holmes *et al*., 1995; Kleim *et al*., 1991; Chanter *et al*., 1993), and of a rape victim and their assailant in Sweden (Albert *et al*., 1994), It has also allowed the exclusion of nosocomial infection as a source of HIV infection from an HIV-positive surgeon (Rogers *et al*., 1993; Holmes *et al*., 1993), from a second U.S. dentist (Jaffe *et al*., 1994), and from a British health care worker (Arnold *et al*., 1995). Studies of emerging epidemics have revealed high levels of sequence similarity among patients infected in a short period of time (Ou *et al*., 1993; McCutchan *et al*., 1992; Dietrich *et al*., 1993; Grez *et al*., 1994) and in a recent investigation of a lengthy transmission chain, sequence data have been able to reconstruct the transmission routes reported (Leitner *et al*., 1996).

These studies have been carried out with sequences from the V3 region of the *env gene* (Ou *et al*., 1992; Hillis and Huelsenbeck, 1994; Balfe *et al*., 1990; Kleim *et al*. 1991; Chant *et al*., 1993), from the p17 coding region of the *gag* gene (Holmes *et al*., 1993, 1995; Albert *et al*., 1994), and, more recently, with datasets consisting of the entire *env* gene (Arnold *et al*., 1995) or of combined *env* and *gag* gene sequences (Leitner *et al.,* 1996). We have shown that the p17 coding sequence on its own can reconstruct known cases of epidemiological linkage and have used it to describe the molecular epidemiology of HIV-1 within Edinburgh (Holmes *et al.,* 1995), In that investigation, injecting drug.users (IDUs) and haemophiliacs were identified as forming distinct infection clusters, with patients infected following heterosexual transmission grouping

---

Sequence data from this article have been deposited with the *EMBL*/GenBank Data Libraries under Accession Nos. AF014163-AF014362.

* To whom correspondence and reprint requests should be addressed. Fax: +44-131-650-8673. E-mail: A. Leigh-6rown@ed.ac.uk.

This article is a reproduction of that published in: *Virology* (1997) 235, pp.166-177. Pagination may not match that of the original.

with the IDUs. An investigation in Amsterdam based on V3 region sequences also suggested a separation of IDUs from other risk groups, although only two synonymous nucleotide positions in the V3 loop were consistently associated with the distinction (Kuiken and Goudsmit, 1994).

We have analyzed nucleotide sequence data in order to reconstruct and compare the HIV epidemics in Scotland, Ireland, and the north of England. Samples were obtained from individuals infected by sexual contact between men and by injecting drug use from three cities in Scotland (Edinburgh, Glasgow, and Dundee), from Newcastle in the north of England, Belfast (Northern Ireland), and Dublin (Republic of Ireland). Sequences of the p17 coding region were obtained from over 200 individuals in this survey and we have analyzed the phylogenetic relationships .among these sequences using both parsimony and neighbor-joining (N-J) techniques. Our results indicate lasting distinctions between these risk groups which extend between cities and suggest that HIV-1 was sufficiently rapidly transmitted among IDUs in these cities for those patients to be defined as a single group r from their viral genotypes.

MATERIALS AND METHODS

Patients

Samples were obtained with informed consent from 211 HIV-1-seropositive homosexuals and IDUs attending genito-urinary medicine and infectious diseases clinics in Edinburgh, Glasgow, Dundee, Newcastle, Belfast, and Dublin between 1993 and 1995. All patients were resident in the city where they attended, although a few reported travelling overseas and were suspected to have acquired their infections outside the British Isles. Approximately 20 individuals from each risk group within each city were included in the study. When samples from more individuals were available, a subset was chosen at random for sequencing. Most were seroprevalent cases, although seroconversion dates were available for a minority. Where the number of patients available was less than 20, then all samples in that category (risk group and city) were sequenced.

Sample preparation and DNA Extraction

EDTA-treated whole blood samples were collected, shipped to Edinburgh, and processed within 48 hr. The blood samples were separated into plasma and peripheral blood mononuclear cell (PBMC) fractions by ficoll Hypaque (Pharmacia) density gradient centrifugation at 1500 *g* for 30 min and the PBMC fraction was stored immediately in liquid nitrogen. DNA extraction was performed from 10^6 to 10^7 uncultured PBMCs as described by Simmonds *et al.* (1990).

PCR amplification and sequence analysis

An approximately 390-base pair (bp) fragment of the p17 coding region of the gag gene between positions 69 and 453 in the HIV-HXB2 genome (Myers *et al.*, 1995) was amplified for each patient by nested polymerase chain reaction (PCR) essentially as described by Leigh Brown and Simmonds (1995), using primers "*gag* 1-4" of Holmes *et al*., (1993). Single-stranded DNA was purified on streptavidin-coated magnetic beads (Dynal-Dyna-beads M280) and sequenced using an Applied Biosys-tems PRISM Sequenase Terminator Single Stranded DNA Sequencing kit according to ABI operating instructions, as described elsewhere (Leigh Brown and Simmonds, 1995).

The raw nucleotide sequences were assembled with the TED and XBAP sequence editors (Staden, 1993) and .aligned using the CLUSTAL V algorithm (Higgins and Sharp, 1988), as implemented in version 2.2 of the Genetic Data Environment (GDE) package (Smith *et al.*, 1994). The final alignment was improved manually. Phylo-genetic analyses were performed using programs taken from version 3.52c of the Phylogeny Inference Package (PHYLIP; Felsenstein, 1989) using the neighbor-joining method (Saitou and Nei, 1987; program

"NEIGHBOR"), maximum parsimony ("DNAPARS"), and bootstrap resampling (Felsenstein, 1985) ("SEQBOOT" and "CONSENSE"). Nucleotide distances were estimated using the generalized two-parameter (maximum likelihood) model (Kishino and Hasegawa, 1989) ("DNADIST"). Principal coordinates analysis was performed using the program "PCOORD" (Higgins, 1992).

RESULTS

Two sequences of the p17 coding region of *gag* were obtained for all newly sequenced patients. Preliminary neighbor-joining phylogenetic analysis of 411 sequences from 211 patients showed that for all patients for whom more than one sequence was obtained, the sequence which grouped most closely was the other from the same individual. This preliminary analysis also included sequences from 24 HIV-1 subtype B reference isolates, as a check for contamination. Subsequently, a single sequence was used from each patient for all further analyses so as to reduce computation time.

Of the approximately 390 nucleotide sites sequenced no fewer than 268 were variable in the dataset. At the amino acid level, 99 of 130 residues showed some variation. The variable residues were distributed throughout the sequence, with some evidence for greater conservation toward the 3' end. The amino acid sequences obtained from 116 homosexual men and 84 IDUs are shown in Fig. 1 aligned against HIV-1_{MN}; the sequences from additional patients analyzed from Edinburgh will be presented elsewhere (C. M. Wade *et al.,* submitted).

Nucleotide distance comparisons

Nucleotide distances were calculated for all possible pairwise comparisons of the sequences from 211 new patients together with those of 24 reference isolates. The mean distances between individuals were calculated by city and risk group (Table 1); the mean distance among reference sequences was 6.03%. Among homosexual men, the mean distances for each city were very similar with an overall mean of 7.2% (range 6.7% (Belfast)–7.5% (Edinburgh)). However, for four cities, there was even greater uniformity among IDUs (range 4.3-4.4%). The mean distance among Belfast IDUs was higher, at 7.3%, but only four IDUs were available from this city and only one from Newcastle. There was a marked difference between the means for the two risk groups: the means for homosexual men were significantly higher than those for IDUs ($P < 0.001$; based on binomial standard errors given in Table 1). In addition the mean distances among 16 haemophiliacs from Edinburgh was 3.3 ± 0.23%. This group, all of whom seroconverted following exposure to a common batch of factor VIII, had previously been found to show a low level of nucleotide diversity (Holmes *et al.,* 1995).

Comparisons between sequences from different cities reveal an unexpected feature in both risk groups. The mean divergence between patients from the same risk group for any pair of cities was no greater than the within-city diversity for the same risk group. The range of intercity mean distances among IDUs was 4.3-4.6% and for homosexual men was 6.8-7.4%. Thus, overall, there was no effect whatever of geographical origin on nucleotide distance, but a highly significant and consistent effect of risk group.

Estimating the frequencies of synonymous (d_s) and nonsynonymous substitutions (d_n) separately revealed that the difference between risk groups came from both classes of substitution (Table 1). However, the mean synonymous divergence among sequences from homosexuals was nearly twice that among IDUs, while the ratio for nonsynonymous substitutions was 1.4. A corresponding difference in the ratio d_s/d_n between the risk groups was observed. For samples from homosexual men-this was 3.0, while for IDUs it was 2.3.

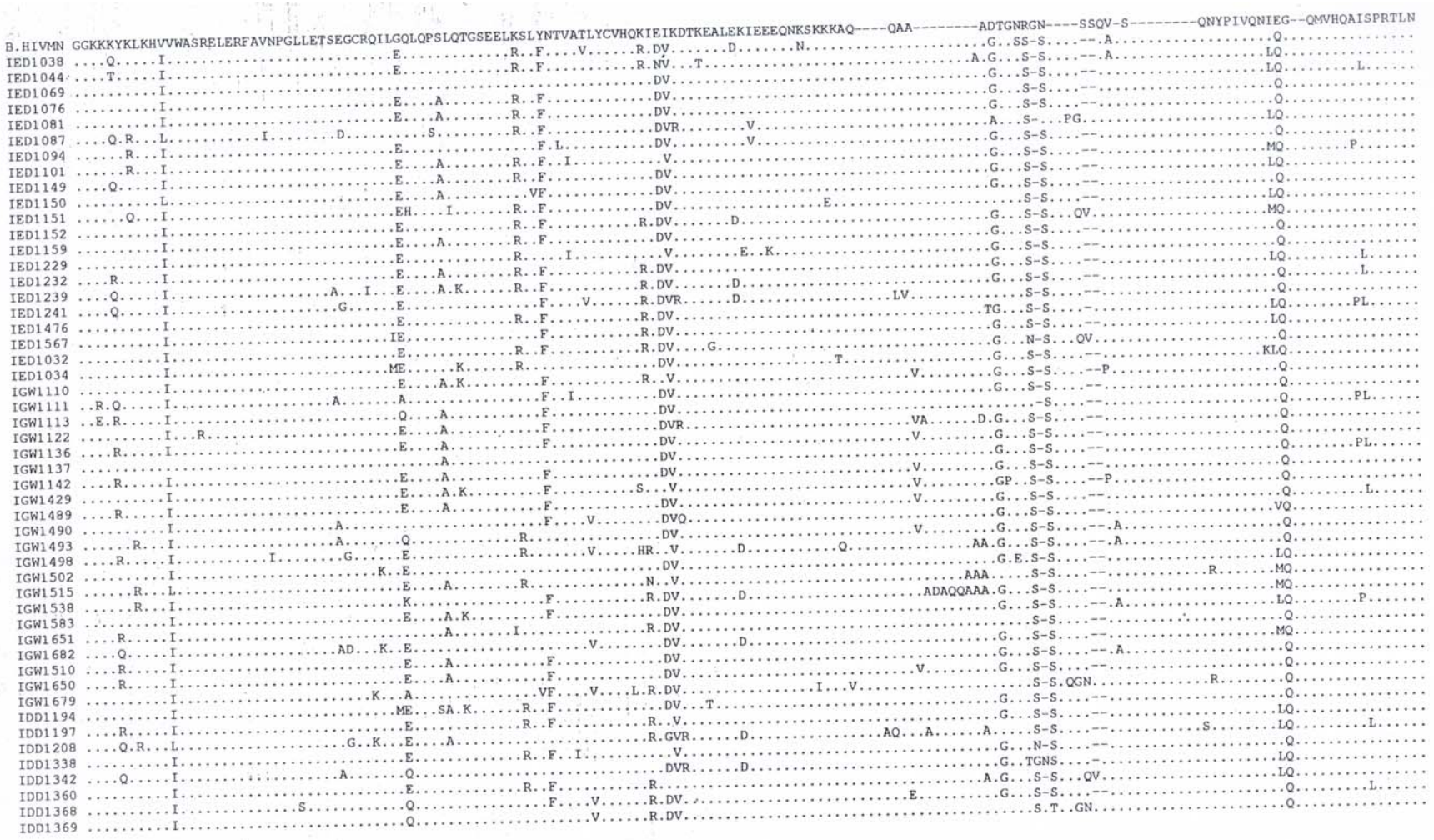

FIG. 1. Amino acid sequences of the MA protein coding region (p17) of the *gag* gene from 200 patients, aligned against the sequence of HIV-1_{MN}. Patient risk group codes: I.., injecting drug user; G.., homosexual man. City codes: .ED, Edinburgh; .GW, Glasgow; .DD, Dundee; .NC, Newcastle; .DB, Dublin; .BF, Belfast.

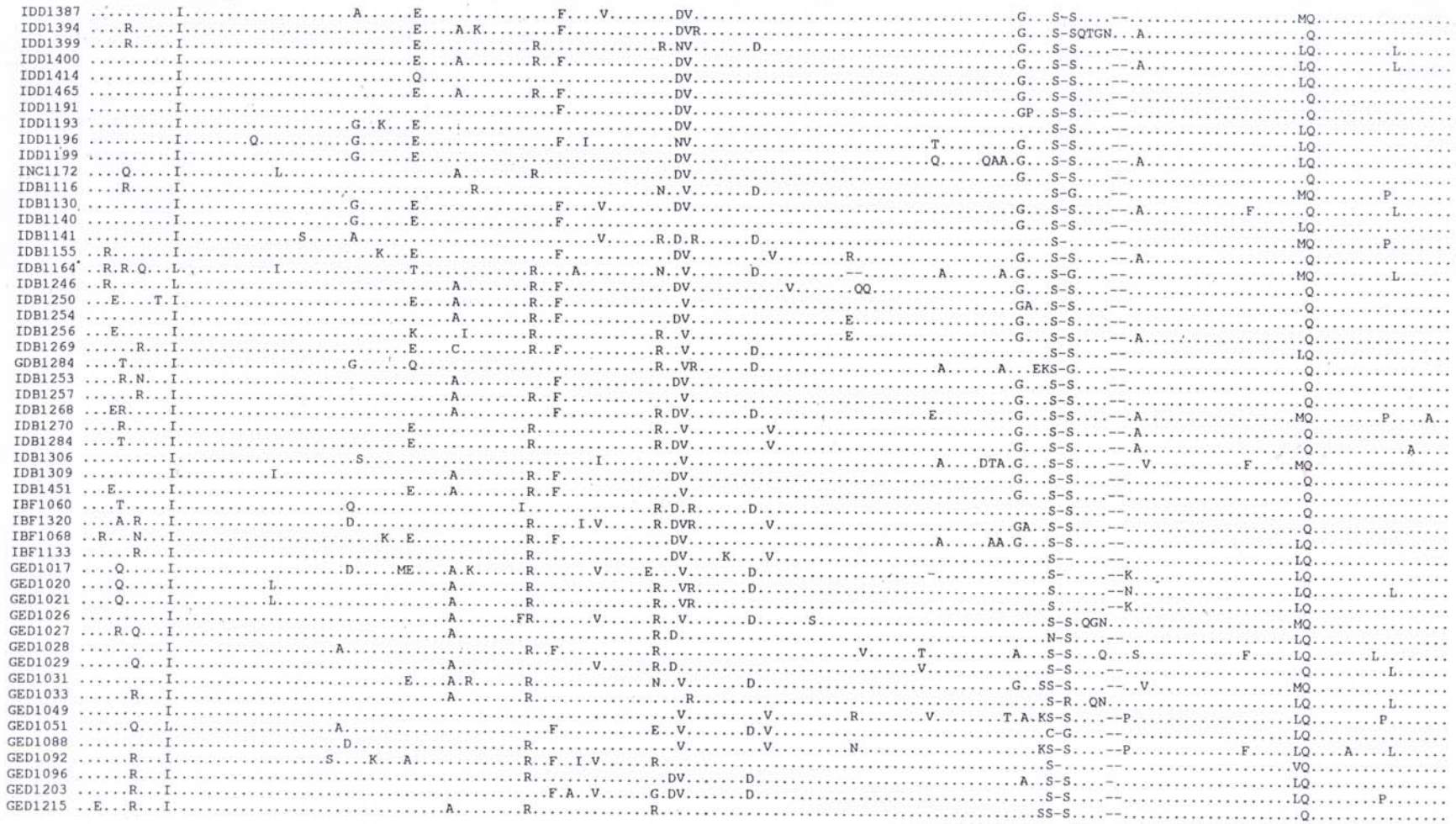

FIG. 1 — *Continued*

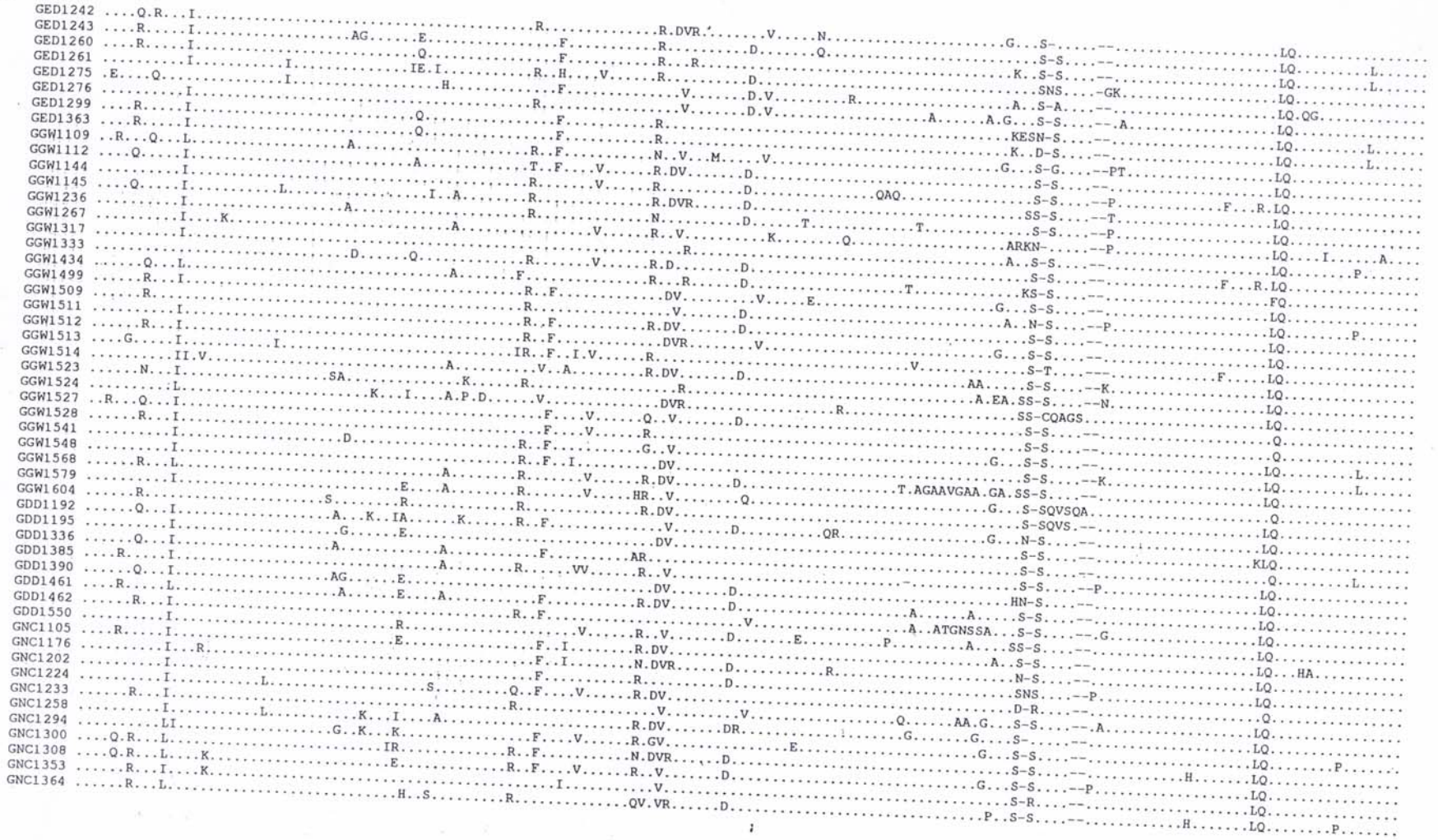

FIG. 1 — *Continued*

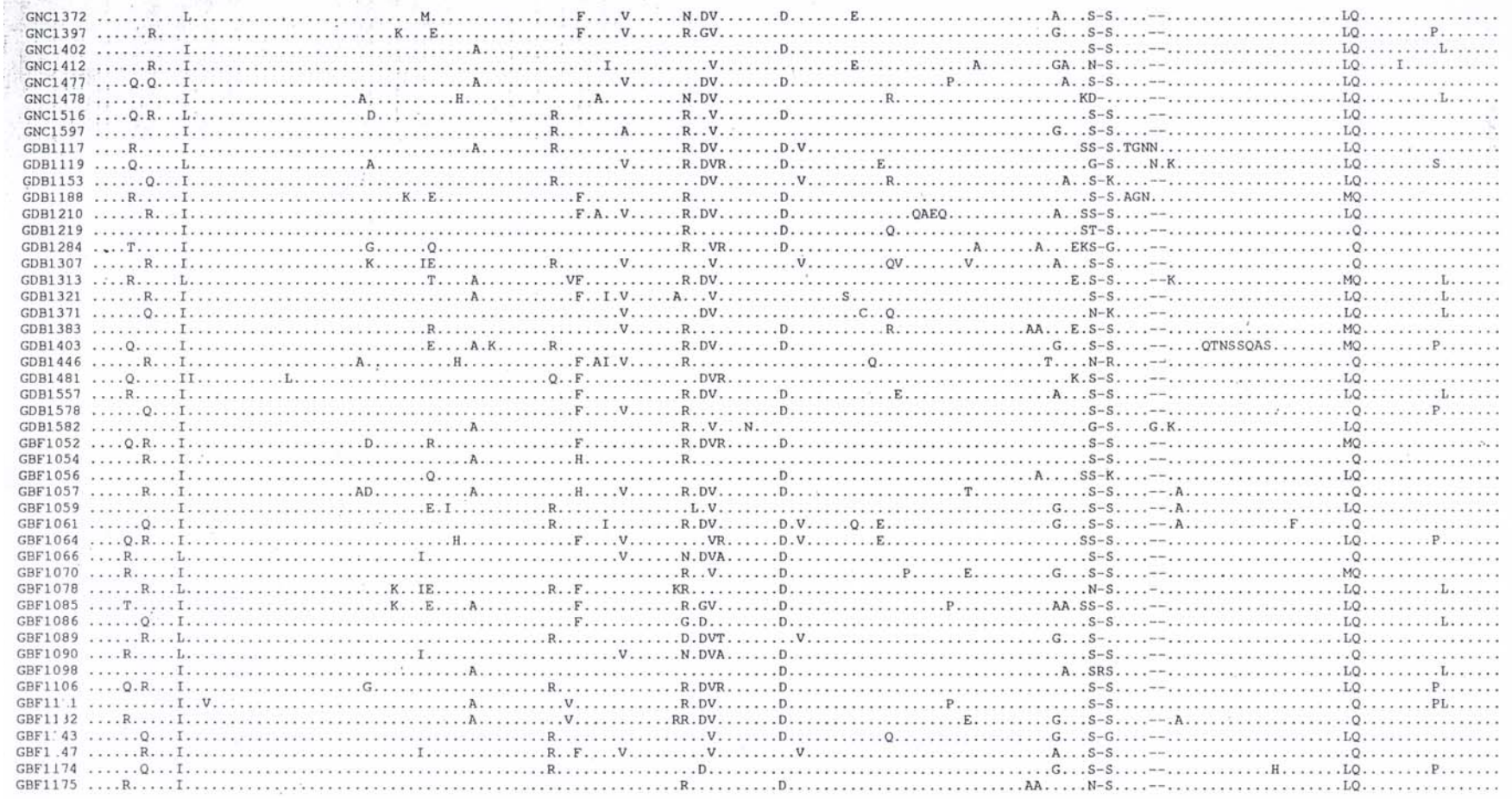

FIG. 1 — *Continued*

TABLE 1

Nucleotide Distances among Homosexual Men and Injecting Drug Users According to City

	IDU		Homosexual	
City	*n*	Nucleotide distance % (±SE)	*n*	Nucleotide distance % (±SE)
Edinburgh[a]	30	4.3 ±0.19	24	7.5 ± 0.28
Glasgow*	22	4.4 ± 0.23	25	7.2 ± 0.27
Dundee[a]	18	4.4 ± 0.25	8	7.1 ± 0.49
Newcastle[a]	1	—	19	6.9 ± 0.30
Dublin[a]	19	4.4 ± 0.24	19	6.9 ± 0.3
Belfast[a]	4	7.3 ± 0.76	22	6.7 ± 0.28
Total[a]	94	4.6 ± 0.11	117	7.2 ± 0.12
Synonymous[b]	94	8.3 ± 0.82	117	15.5 ± 1.1
Nonsynonymous[b]	94	3.6 ± 0.33	117	5.0 ± 0.35
D_s/d_n[b]	94	2.3	117	3.0

[a] Overall nucleotide distances.

[b] Nucleotide distances for synonymous (d_s) and nonsynonymous (d_n) nucleotide substitutions considered separately for the two risk groups.

Phylogenetic analysis

The initial comparison of the 211 patients' sequences with HIV-1 reference sequences revealed that all belonged to subtype B. All further analyses were undertaken against a background only of subtype B reference sequences. An N-J tree was constructed for the. sequences from all 211 IDUs and homosexual men together with 16 Edinburgh haemophiliacs and 24 subtype B reference sequences. The tree is characterized by long branches to the tips with few, and short, internal branches (Fig. 2). Nevertheless, one subdivision is apparent, which appeared consistently with different numbers of sequences. This contains approximately 80% of the sequences from IDUs and a small number (5) from homosexual men, but none of the reference sequences, the closest being HIV-1_{PH136}. A total of 13 IDUs were located among sequences from homosexuals in the N-J tree (Fig. 2). The location of the Edinburgh haemophiliacs among these sequences confirmed their separation from the IDUs proposed by Holmes *et al*. (1995) and specifically associated the source of this infection with the homosexual risk group (see below).

In order to test the division of the dataset into two risk group-associated clusters, a parsimony analysis was carried out and a majority-rule consensus of 100 of the most parsimonious trees is shown (Fig. 3). This dado-gram also split into two groups, again separating most of the IDUs from the remaining sequences, with none of the B subtype reference sequences among them. To investigate the stability of the IDU cluster, each of the 100 trees was examined, and the location of each IDU sequence recorded. Only the 9 sequences identified by asterisks in Fig. 3, which moved in or out of the IDU group in different trees, showed any inconsistency in their clustering. We conclude that the parsimony analysis strongly supports the existence of two phylogenetically distinct clusters, associated with the two risk groups.

A third method of analyzing nucleotide distance data, based on the principal coordinates technique (Higgins, 1992), was employed to assess the division of the sequences by risk group. Following a transformation of the distances, principal coordinates analysis determines which combination of variables is most suited to reflect the variation in the dataset (the "principal

coordinates"). Although the principal coordinates will be a poor reflection of the total variation in the dataset, they will reflect the major distinctions and consequently the method is particularly useful in resolving major trends within a data-set which may often be invisible or unreliable in a tree. Principal coordinates analysis of the 211 patients from the two risk groups (Fig. 4) clearly distinguished the sequences into two groups containing sequences from IDUs and homosexual men, respectively, and with very little overlap between the two.

Detection of contact networks

As indicated above, although the main IDU cluster was repeatedly identified among the most parsimonious trees, bootstrap resampling of the data did not support it. However, a total of 13 smaller clusters were supported at the 65% level or above, including 6 groups known previously to be linked. These included groups of reference sequences derived from the HIV-1_{LA1} strain, sequences from individuals who were part of an outbreak of HIV-1 infection in a Scottish prison in 1993 (Taylor *et al.*, 1995; Yirrell *et al.*, 1997) and sequences from the

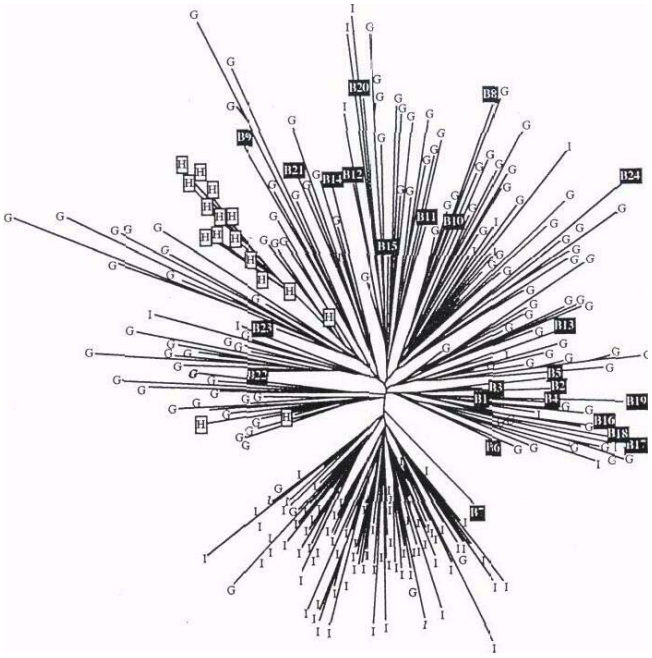

FIG. 2. Neighbor-Joining phylogenetic tree of sequences of the MA protein coding region (p17) of the *gag* gene from 227 patients and 24 subtype B reference sequences. Risk group codes: H, haemophiliac; G, homosexual man; I, injecting drug user. B, Subtype B isolate sequences: 1, HIV-l_{LA1} sequence BRU; 2, HIV-1 BH102; 3, HIV-1_{PH22}; 4, HIV-l_{LA1}, clone HXB2; 5, HIV-I_{MN}; 6, HIV-1_{JH3}; 7, HIV-l_{PH136}; 8, HIV-l_{BZ190}; 9, HIV-l_{RF2}; 10, HIV-l_{NL4-3}; 11, HIV-1_{NY6}; 12, HIV-1_{CDC4}; 13, HIV-1_{D31}; 14, HIV-1_{HOY1}; 15, HIV-l_{YU2}; 16, HIV-1_{PH153}; 17, HIV-1_{JPCSF}; 18, HIV-1_{JRFL}, 19, HIV-1_{TB132}; 20, HIV-1_{BZ167}; 21, HIV-1_{HAN}; 22, HIV-1_{SF2}; 23, HIV-1_{CAM1}; 24, H1V-1_{BZ200}.

Edinburgh haemophiliac cohort (Table 2). The haemophiliacs grouped specifically with a sequence from a homosexual man attending the Edinburgh GUM clinic who was infected in 1984.

A total of seven new groups were identified by bootstrapping of the entire tree with values exceeding 70%, five of which were supported in over 98% of bootstrap replicates (Table 2). Most of these groups contained a single pair of individuals, but one (99%) included three homosexual men from Edinburgh. In all clusters, the individuals identified came from the same city, consistent with the interpretation that they represent true contact networks.

DISCUSSION

In this study we have analyzed nucleotide sequences of the p17 coding region of the *gag* gene obtained from over 200 patients representing the two major risk groups in six cities. Comparisons based on nucleotide distances revealed no evidence that subdividing the data by city led to significant heterogeneity. However, there was a consistent and highly significant effect of risk group, with the sequences, of injecting drug users all being more similar to each other, regardless of their city of origin. Both the neighbor-joining and parsimony methods of tree construction separated the majority of IDUs into a single cluster; none of the 24 HIV-1 subtype B reference isolates grouped with IDUs, instead all were widely distributed in the rest of the tree.

There has been considerable discussion about the appropriateness of various methodologies for the reconstruction of HIV-1 phylogenies (Hillis *et al.,* 1994; Leitner *et at.,* 1996). An equally important issue is the choice of gene region for sequencing (Holmes *et al.,* 1995; Arnold

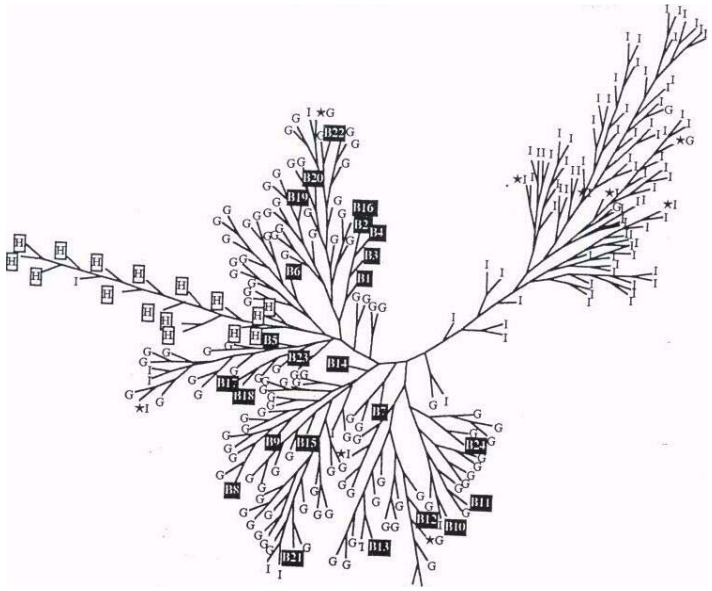

FIG. 3. Parsimony tree of sequences of the p17 protein coding region (MA) of the *gag* gene from 227 patients and 24 subtype B reference sequences. The majority-rule consensus of 100 of the most parsimonious trees is shown. For risk group and subtype B isolate sequence codes, see Legend to Fig. 2. Sequences marked * showed variable location with respect to the main IDU cluster among the 100 trees. All others showed a constant location.

et al., 1995; Leitner *et al.,* 1996). The *gag* gene evolves more slowly than the V3 region of *env* (Leigh Brown and Monaghan, 1988) and it has been suggested that it does not contain sufficient information for the accurate reconstruction of phylogenies (Arnold *et al.,* 1995; Leitner *et al.,* 1996). However, when nucleotide distances are as great as has been observed in this study (Table 1), that is not likely to pose a significant problem. Indeed, we detected no less than five previously unknown linkages between patients, supported in >98% bootstrap replications and. confirmed all seven previously known clusters included in the dataset. We conclude that for community-based studies, the p17 coding sequencers sufficiently informative.

No previous study of the molecular epidemiology of HIV-1 has attempted a systematic comparison of HIV sequences from multiple cities and risk groups. Extensive studies in Thailand have concentrated on two centres of infection, in the north and south of the country, and have compared the viral sequences from IDUs and patients infected by heterosexual contact. These studies revealed a geographic division between two genetically very distinct strains (Ou *et al*., 1993; McCutchan *et al*., 1992), but also that subtype E sequences were associated with a

predominantly heterosexual mode' *of* transmission, while subtype B sequences were found in IDUs (Kunanusont *et al.*, 1995). These epidemics initiated from variants found in Thailand in 1989 and 1988, respectively (Weniger *et al*., 1994), and the-subtype E sequences are now increasing in frequency in the IDU risk group as well (Kalish *et al*., 1995). Overall, the effect of risk group appears to have been more significant than that of geographic location, as has clearly been the case in our study.

The second major study to compare HIV sequence variants in different risk groups has focussed on two risk groups in Amsterdam infected with HIV-1 subtype B. These studies analyzed V3 region sequences and detected a minor, but consistent, difference between sequences from IDUs and homosexual men in the nucleotide sequence of theV3 loop (Kuiken *et el*., 1993; Kuiken and Goudsmit, 1994). Subsequently, differences were reported in other coding regions as well (Kuiken *et al.*, 1996). The statistically significant difference in nucleotide distances we have described, and the similar division in the phylogenetic analyses, suggests that these studies appear to identify the same group of IDUs. We have extended our studies to include IDUs known to have been infected in Amsterdam and have found them to fall into the main IDU cluster we originally identified in Edinburgh (Holmes *et al*., 1995). Intriguingly, this does not

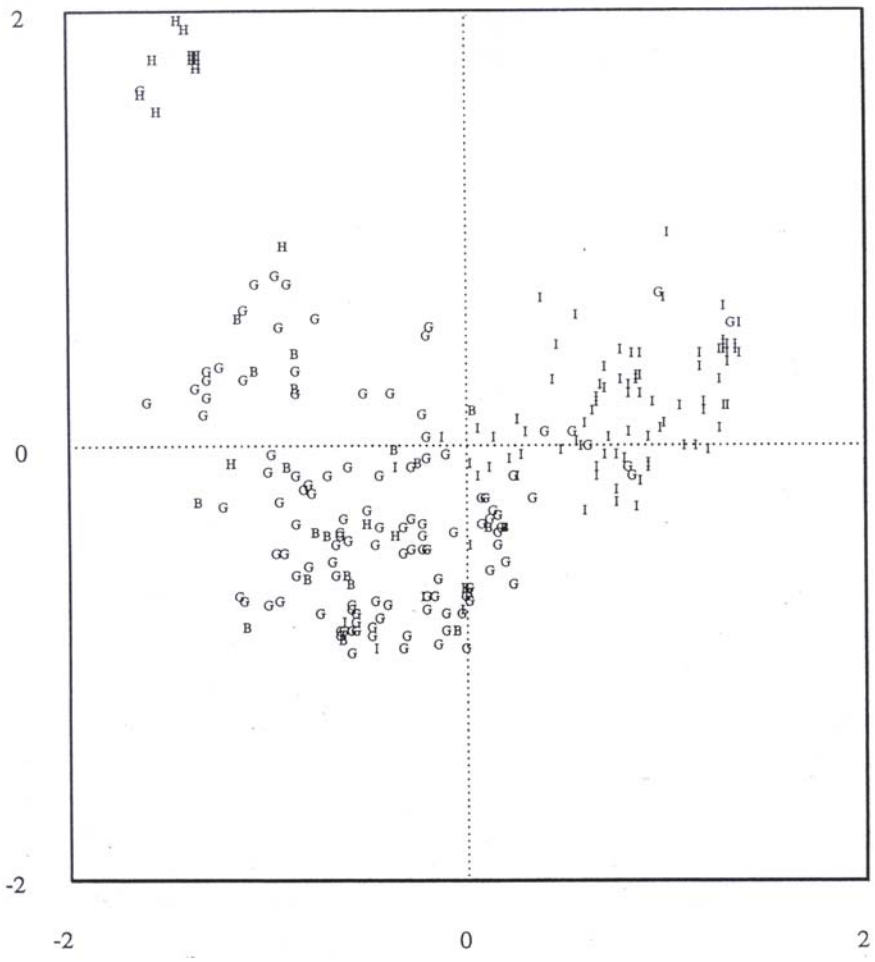

FIG. 4. Principle coordinates analysis separates the main IDU cluster from other risk groups. The plot represents the two largest principle coordinates as ordinate and abscissa, with arbitrary units. The identifiers for risk groups are as described in the legend to Fig. 2.

apply to IDUs from southern Europe, whose sequences scatter widely among those of homosexual men (D. Lobidel and V. Soriano, unpublished data).

From our data, and there is a clear indication that a genetic bottleneck occurred when HIV-1 entered the IDU group we have studied, presumably reflecting infection of a single individual followed by very rapid spread. The major epidemic in Edinburgh occurred in 1983/1984 (Robertson *et al*., 1986), which appears to be earlier than epidemics among IDUs in other cities, including Amsterdam (van Haastrecht *et al.,* 1991). The extreme rapidity of transmission (Robertson *et al.,* 1986), coupled with the lack of variability observed in the first few weeks of the infection (Zhang *et al.,* 1993; Zhu *et al.,* 1993) can account for the rapid spread of almost identical variants among highly susceptible populations, as has also been described in Thailand (Weniger *et al.,* 1992), Bombay (Dietrich *et al.,* 1993; Grez *et al.,* 1994; Pfutzner *et al.,* 1992), and recently among inmates of a Scottish prison (Yirrell *et al*., 1997).

Although the overall nucleotide diversity observed in IDUs is restricted, the diversity observed among this sample of homosexual men is at least as great as that observed between subtype B sequences from across the United States. In particular, the branch lengths between some of these sequences are substantially greater than

TABLE 2

Significant Associations between Patients Identified from Phylogenetic Analysis of *gag* p17 Sequences

Cluster	Patients	Bootstrap[a]	Risk group	
1	1397	100	Homosexual	Newcastle
	1294		Homosexual	Newcastle
2	1333	70	Homosexual	Glasgow
	1144		Homosexual	Glasgow
3	1066	100	Homosexual	Belfast
	1090		Homosexual	Belfast
4	Go8	100	Prisoner[b]	Glasgow
	1523		Homosexual	Glasgow
5	1122	71	IDU	Glasgow
	1142		IDU	Glasgow
6	1363	99	Homosexual	Edinburgh
	1299		Homosexual	Edinburgh
	1260		Homosexual	Edinburgh
7	1021	100	Homosexual	Edinburgh
	1020		Homosexual.	Edinburgh
8	1028	66	Homosexual	Edinburgh
	EDI		Haemophilia cohort[c]	Edinburgh

[a] Percentage of bootstrap replications in which the cluster was observed. Those shown were all observed in >65% of bootstrap replicates.

[b] Seropositive prisoner identified in Glenochil prison (Taylor *et al.,* 1995; Yirrell *et al.,* 1997).

[c] Edinburgh Haemophilia Cohort (Holmes *et al*., 1995).

those between the reference isolates, and even exceed -the divergence between the prototype B subtype strain HIV-1MN and the prototype subtype D strain HIV-l_{EL1} (Myers *et al.,* 1995). There are three possible factors responsible: first that the virus has entered the homosexual risk group

in northern Europe on many occasions, each of which has established descendent lineages, i.e., there has been no recent bottleneck. Second, the reference sequences have all been obtained from tissue culture-adapted strains; it is possible that the known selection that is imposed on the virus in this process (Meyerhans *et al*., 1989; Kusumi *et al.,* 1992) results in a restriction of diversity in p17. Restriction of amino acid divergence of p17 *in vivo* is indicated by the greater d_s/d_n ratio we have observed among homosexual men than IDUs. This indicates a slowing down of the divergence of amino acid, sequences relative to synonymous sites (Table 1). A final factor is that the isolates referred to were established in the early to mid 1980s, while patient sequences have continued to evolve during the intervening decade.

The availability of an independent date for the IDU epidemic in Edinburgh, approximately 10 years prior to sampling, and the point source origin of the virus in this population, allows the estimation of a mean divergence rate at synonymous sites of 0.83% (±0.08%) per year. The equivalent estimate among sequences from homosexual men was 1.55% (±0.11%) (Table 1). Assuming these have evolved at the same rate, the time between divergence and sampling in this group is 18.7 years (95% confidence interval, 13.4–26.4 years), corresponding to a date of about 1975 (95% CI, 1968-1980). If our sample of homosexual men represent an unbiased sample of the B subtype, this is an estimate of the origin of the B subtype epidemic.

ACKNOWLEDGMENTS

We are very grateful to the patients and the nursing staff of the clinical centres for provision of samples. We are also grateful to Anne P. Leigh Brown for development of the database and Dr. David Goldberg for advice and comments on the manuscript. This work was supported by the Medical Research Council AIDS Directed Programme.

REFERENCES

Albert, J., Wahlberg, J., Leitner, T., Escanilla, D., and Uhlen, M, (1994). Analysis of a rape case by direct sequencing of the human immuno-deficiency virus type 1 pol and gag genes. *J. Virol.* 68, 5918-5924.

Arnold, C., Baife, P., and Clewley, J. P. (1995). Sequence distances between env genes of HIV-1 from individuals infected from the same source: Implications for the investigation of possible transmission events. *Virology* 211, 198-203.

Balfe, P., Simmonds, P., Ludlam, C. A, Bishop, J. 0., and Leigh Brown, A J. (1990). 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.* 64, 6221-6233.

Chant, K., Lowe, D., Rubin, G., Manning, W., O'Donoughue, R., Lyle, D., Levy, M., Morey, S., Kaldor, J., Garsia, R., Penny, R., Marriott, D., Cunningham, A, and Tracy, G. D. (1993). Patient-to-patient transmission of HIV in private surgical consulting rooms. *Lancet* 342, 1548-1549.

Dietrich, U., Grez, M., von Briesen, H., Panhans, B., Geissendorfer, M., Kuhnel, H., Maniar, J., Mahambre, G., Becker, W. B., Meeker, M. L B., and Rubsamen Waigmann, H. (1993). HIV-1 strains from India are highly divergent from prototypic African and US/European strains, but are linked to a South African isolate. *AIDS* 7, 23-27.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39., 783-791.

Felsenstein, J. (1989). PHYLIP-phylogeny inference package (version 3.2). *Cladistics* 5, 164-166.

Grez,.M., Dietrich, U., Balfe, P., von Briesen, H., Maniar, J. K., Mahambre, G., Delwart, E..L, Mullins, J. I., and 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.

Higgins, D. G. (1992). Sequence ordinations: A multivariate analysis approach to analysing large sequence datasets. *CABIOS* 8, 15-22.

Higgins, D. G., and Sharp, P. M. (1988), CLUSTAL; a package for performing multiple sequence alignment on a microcomputer. *Gene* 73, 237-244.

Hillis, D. M., Huelsenbeck, J. P., and Cunningham, C. W. (1994). Application and accuracy of molecular phylogenies. *Science* 264, 671-677.

Hillis, D. M., and Huelsenbeck, J. P. (1994). Support for dental HIV transmission. *Nature* 369, 24-25.

Holmes, E. C., Zhang, L. Q., Simmonds, P., Rogers, A. S., and Leigh Brown, A. J. (1993). Molecular investigation of Human Immunodeficiency Virus (HIV) infection in a patient of an HIV-infected surgeon. *J. Infect Dis.* 167, 1411-1414.

Holmes, E. C., Zhang,.L Q., Robertson, P., Cleland, A., Harvey, E., Simmonds, P., and Leigh Brown, A. J. (1995). The molecular epidemiology of HIV-1 in Edinburgh, Scotland. *J Infect. Dis.* 171, 45-53.

Jaffe, H. W., McCurdy, J. M., Kalish, M. L, Liberti, T., Metellus, G., Bowman, B. H., Neasman, A. R., and Wine, J. J. (1994). Lack of transmission of human immunodeficiency virus in the practice of a dentist with AIDS. *Ann. Intern. Med.* 121, 855-859.

Kalish, M. L, Luo, C. C., Raktham, S., Wasi, C., Baldwin, A., Schochetman, G., Mastro, T. D., Young, N., Vanichseni, S., Rubsamen Waigmann, H., von Briesen. H., Mullins, J. I., Delwart, E., Herring, B., Esparza, J., Heyward, W. L, and Osmanov, S. (1995). The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. *AIDS* 9, 851-857.

Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order in Hominoidea. *J. Mol. Evol*. 4, 406-425.

Kleim, J.-P., Ackerman, A., Brackman, H. H., Gahr, M., and Schneweis, K.E. (1991). Epidemiologically closely related viruses from hemophilia B patients display high homology in two hypervariable regions of the HIV-1 *env* gene. *AIDS Res. Hum. Retroviruses 7,* 417-421.

Kuiken, C. L, Zwart, G., Baan, E., Coutinho, R. A., van den Hoek, J. A. R., and .Goudsmit, J. (1993). 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* 90, 9061-9065.

Kuiken, C. L, Cornelissen, M. T., Zorgdrager, F., Hartman, S., Gibbs, A J., and Goudsmit, J. (1996). Consistent risk group-associated differences in human immunodeficiency virus type 1 vpr, vpu and V3 sequences despite independent evolution. *J Gen. Virol.* 77, 783-792.

Kuiken, C. L, and Goudsmit, J. (1994). Silent mutation pattern in V3 sequences distinguishes virus according to risk group in Europe. *AIDS Res. Hum. Retroviruses* 10, 319-320.

Kunanusont, C., Foy, H. M., Kreiss, J. K., Rerks-Ngarm, S., Phanuphak, P., Raktham, S., Pau, C. P., and Young, N. L (1995). HIV-1 subtypes and male-to-female transmission in Thailand. *Lancet* 345, 1078-1083.

Kusumi, K., Conway, B., Cunningham, S., Berson, A, Evans, C., Iversen, .A K., Colvin, D., Gallo, M. V., Coutre, S., Shaper, E.G., Faulkner, D.V., DeRonde, A, Volkman, S,, Williams, C., Hirsch, M. S., and Mullins, J. I. (1992). Human immunodeficiency virus type 1 envelope gene structure and diversity in vivo and after cocultivation in vitro. *J. Virol.* 66, 875-885.

Leigh Brown, A., and Monaghan, P. (1988). Evolution of the structural proteins of human immunodeficiency virus; selective constraints on nucleotide substitution. *AIDS Res. Hum. Retroviruses* 4, 399-407.

Leigh Brown, A. J., and Simmonds, P. (1995). *HIV: a practical approach; Volume 1, Virology and Immunology* (Karn, J., Ed.) Oxford University Press, Oxford. 161-188.

Leitner, T., Escanilla, D., Franzen, C., Uhlen, M., and Albert, J. (1996). Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc. Natl. Acad. Sci. USA* 93, 10864-10869.

McCutchan, F. E:, Hegerich, P. A., Brennan, T. P., Phanuphak, P., Singharaj, P., Jugsudee, A., Berman, P.W., Gray, A. M., Fowler, A. K., and Burke, D. S. (1992). Genetic variants of HIV-1 in Thailand. *AIDS Res. Hum. Retroviruses 8*, 1887-1895.

Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt Manson, L, Asjo, B., and Wain-Hobson, S. (1989). Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. *Cell* 58, 901-910.

Myers, G., Korber, B., Hahn, B. H., Jeang, K. T., Mellors, J. W., McCutchan, P. E., Henderson, L. E., and Pavlakis, G. N. (1995). *Human retroviruses and AIDS* 1995, Los Alamos National Laboratory.

Ou, C. Y., Ciesielski, C. A, Myers, G., Bandea, C. I., Luo, C. C., Korber, B. T,, Mullins, J. I., Schochetman, G., Berkelman, R. L, Economou, A. N., Witte, J. J., Furman, L, J., Satten, G. A., Maclnnes, K. A., Curran, J. W., and Jaffe, H.'W. (1992). Molecular epidemiology of HIV transmission in a dental practice. *Science* 256, 1165-1171.

Ou, C. Y., Takebe, Y., Weniger, B. G., Luo, C. C., Kalish, M. L, Auwanit, W., Yamazaki, S., Gayle, H. D., Young, N. I., and Schochetman, G. (1993). Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. Lsncer341, 1171-1174.

Pfutzner, A., Dietrich, U., von Eichel, U., voh Briesen, H., Brede, H. D., Maniar, J. K., and Rubsamen Waigmann, H; (1992). HIV-1 and HIV-2 infections in a high risk population in Bombay. India; evidence for the spread of HIV-2 and presence of a divergent HIV-1 subtype. *1. Acquir. Immune. Defic. Syndr.* 5, 972-977.

Robertson, J. R., Bucknall, A. B. V., Welsby, P. D., Roberts, J. J. K., Inglis, J. M., Peutherer, J. F., and Brettle, R. P. (1986). Epidemic of AIDS related virus (HTLV-III/LAV) among intravenous drug abusers. *Br. Med. J.* 292, 527-529.

Rogers, A. S., Froggatt, J. W. Ill, Townsend, T., Gordon, T., Leigh Brown, A. J., Holmes, E. C., Zhang, L Q., and Moses, H. III (1993). An investigation of potential HIV transmission to the patients of an HIV infected surgeon. *J. Am. Med. Assoc*. 269, 1795-1801.

Saitou, N., and Nei, M. (1987). The neighbor-joining method; A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425.

Simmonds, P., Balfe, P., Peutherer, J. F., Ludlam, C. A., Bishop, J. 0., and Leigh Brown, A. J. (1990). Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J Virol.* 64, 864-872.

Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W., and Gillevet, P. M. (1994). The genetic data environment and expandable GUI for multiple sequence analysis. *Comput. Appl. Biosci.* 10, 671-675.

Staden, R. (1993). Staden package update. *Genome News* 13, 12-13.

Taylor, A., Goldberg, D., Emslie. J., Wrench, J., Gruer, L, Camerson, S., Black, J., Davis, B., McGregor, J., Follett, E., Harvey, J., Basson, J., and McGavigan, J. (1995). Outbreak of HIV infection in a Scottish prison. *Br. Med. J.* 310, 289-292.

van Haastrecht, H. J. A., van den Hoek, J. A. R., Mientjes, G. H., and Coutinho, R. A. (1991). Did the introduction of HIV among homosexual men precede the introduction among injecting drug users in the Netherlands. *AIDS* 6, 131-132.

Weniger, B. G., Limpakarnjanarat, K., Ungchusak, K., Thanprasertsuk, S., Cnoopanya, K., Vanichseni, S., Uneklabh, T., Thohgcharoen, P., and Wasi, C. (1992). AIDS 1991—A year in review. *In* "AIDS 5," pp. S71-S85. Current Science, London.

Weniger, B. G., Takebe, Y., Ou, C.Y., and Yamazaki, S. (1994). The molecular epidemiology of HIV in Asia. *AIDS 8* (suppi 2), S13-S28.

Yirrell, D., Robertson, P., Cameron, S., Goldberg, D., and Leigh Brown, A. J. (1997). Molecular epidemiology of an HIV-1 outbreak in a Scottish prison. *Br. Med. J.* 314, 1446-1450.

Zhang, L. Q., MacKenzie, P., Cleland, A., Holmes, E. C., Leigh Brown, A.J., and 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, D. S., Cao, Y., Koup, R. A., and Ho, D. D. (1993). Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. *Science* 261, 1179-1181.