

---

## Completing the molecular investigation into the HIV outbreak at Glenochil prison

---

D. L. YIRRELL<sup>1\*</sup>, S. J. HUTCHINSON<sup>2</sup>, M. GRIFFIN<sup>1</sup>, S. M. GORE<sup>3</sup>,  
A. J. LEIGH-BROWN<sup>1</sup> AND D. J. GOLDBERG<sup>4</sup>

<sup>1</sup> *Centre for HIV Research, Waddington Building, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, UK*

<sup>2</sup> *MRC BIAS, Waddington Building, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, UK*

<sup>3</sup> *MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK*

<sup>4</sup> *Scottish Centre for Infection and Environmental Health, Clifton House, Glasgow G3 7LN, UK*

(Accepted 7 May 1999)

### SUMMARY

In a molecular investigation into the outbreak of HIV in Glenochil during the first 6 months of 1993, we previously demonstrated that 13 out of the 14 HIV positive inmates were infected with a virtually identical strain, and discounted 2 others as potential sources. Here we investigate a further 8 potential contacts and 4 potential sources which were identified in the companion paper [1]. We were able to examine viral sequence from all but one of these 12 and results have revealed them to be distinct both from each other and the original 14. Thus, despite an intensive follow-up investigation, we have been unable to identify any further HIV infections that might have been part of the 1993 outbreak. It is possible that persons who were infected at that time remain undetected; however this and the companion report strongly suggest that if this were the case the likely numbers would be few.

### INTRODUCTION

In mid 1993 an infection control exercise in Glenochil prison in central Scotland identified 14 HIV positive individuals [2]. From the dates of seroconversion and entry into prison conclusive proof was provided that at least 6/14 had acquired their infection whilst in that institution. Subsequently we carried out a molecular epidemiological investigation using viral sequence analysis to demonstrate that 13 (the Glenochil Cohort) of the 14 were infected with a virtually identical strain of HIV [3]. Such unprecedented similarity between viruses from different individuals led us to the conclusion that this cohort of 13 was infected from a common source within a short space of time. In the light of previous findings it followed that they were all

infected in Glenochil prison. The fourteenth individual harboured a virus quite distinct from the rest. We also examined the virus from two individuals who were known to be HIV positive prior to entry to Glenochil, and who were present in the prison during the outbreak of HIV but found their viruses to be different again from the others. At that time therefore, we were unable to identify the source of the infection although we were equally unable to discount the possibility that 1/13 had acquired his HIV outside prison and entered Glenochil in the very early stages of infection (i.e. prior to seroconversion). Since, however, a known heterosexual transmission from one of the cohort had viral sequences which significantly associated with the 13 we felt confident in the ability of molecular techniques to identify both contacts and sources.

\* Author for correspondence.

Table 1. *Dates and nature of samples from which sequences were obtained and presented in Figures 1a and 1b*

Case	Sample date	Sample	No. of sequences	
			gag	env
C1	03/06/96	PLASMA	2	2
C2	10/11/93	PLASMA	1	1
C2	12/01/95	PBMC	1	1
C3	31/03/93	PLASMA	0	1
C4	22/07/96	PLASMA	2	1
C5	21/10/97	PLASMA	2	2
C6	17/07/96	PLASMA	2	2
C7	08/12/93	PLASMA	0	0
C8	31/05/96	PLASMA	1	2
SA	13/07/95	PBMC	2	2
SB	17/08/93	PBMC	2	2
SC	03/11/94	PBMC	3	3
SD	11/06/93	PBMC	1	0

In this paper we document the extension of the molecular epidemiological investigation surrounding the 1993 outbreak. In the data linkage component of the study, in which a soundex coded list of inmates in Glenochil prison during the first half of 1993 was matched with Scotland's HIV and AIDS diagnosis register [1], a further 12 cases were identified. Of the 12, 4 were known to have been infected prior to 1993 and were considered potential sources, whilst the remaining 8 were diagnosed HIV positive subsequent to January 1993 and thus represented potential contacts of the Glenochil cohort. By applying the same methods as those which proved successful in our original investigation [3], we hoped to identify other transmission events which had occurred at that time.

## SUBJECTS AND METHODS

Details of the subjects and how they were identified are presented in the companion paper [1]. Identification numbers used for potential contacts are prefixed with the letter C and coincide with those used in that publication. To further preserve the anonymity of potential sources their samples (prefixed S) were assigned letters A–D. Thus SA–SD in this publication are the same as S3–6 in reference 1, but remain unidentifiable.

Proviral DNA was extracted from peripheral blood mononuclear cells (PBMC) and viral RNA extracted from plasma obtained from patients as indicated in

Table 1. Viral RNA was subjected to reverse transcription in both the p17 region of the gag gene and the v3/v4 region of the env gene. Both DNA and cDNA were then amplified by nested PCR in the same two regions of the HIV genome and sequenced directly on an ABI 373A automated sequencer as detailed elsewhere [3]. All sequences were obtained blind, the technician having no knowledge of the identity of the individual or whether they represented potential sources or contacts.

Sequences were aligned using version 2.2 of the Genetic Data Environment (GDE) package [4] and phylogenetic analyses were performed using programs taken from version 3.52c of the Phylogeny Inference Package (PHYLIP) [5].

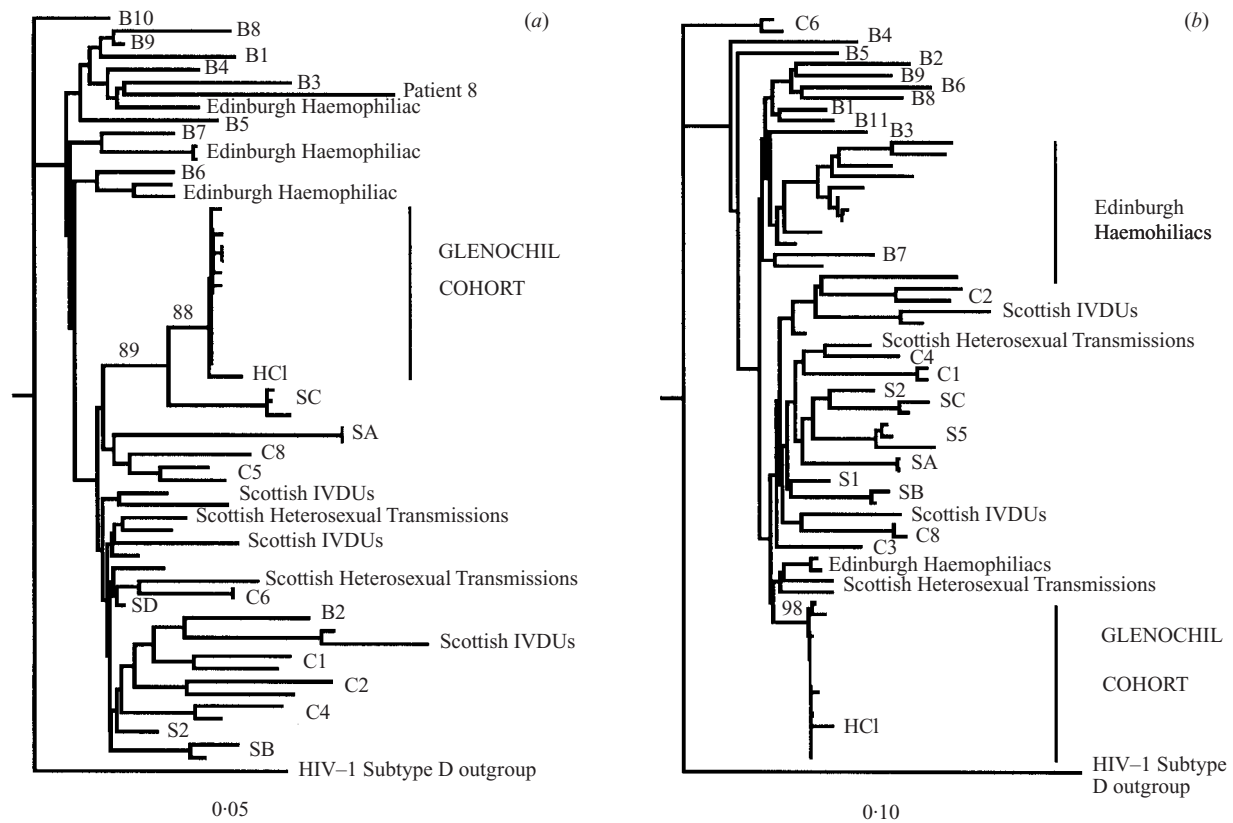
Neighbour joining phylogenetic trees were constructed for both p17 gag and v3/v4 env ideally using two sequences from each of the above patients and rooted using a subtype D isolate, ELI. To assign support these were subjected to bootstrap resampling [6] (500 replicates) using the programmes seqboot and consense. Maximum Likelihood trees were then constructed using the same sequences using the modified PHYLIP program FASTDNAML, and well supported bootstrap values (i.e. > 70%) applied to these figures.

## RESULTS

### Gag sequence data

Figure 1a shows a maximum likelihood phylogenetic tree constructed from gag sequences from cases C1, C2, C4, C5, C6, C8, SA, SB, SC and SD, as indicated in Table 1. Included in the tree are subtype B isolates obtained from international databases and local controls. These consist of 7 Scottish intravenous drug users (IVDUs), 3 Scottish heterosexual transmissions, 5 haemophiliacs from Edinburgh and 1 representative sequence from the previously published Glenochil outbreak (i.e., the Glenochil Cohort, a heterosexual transmission (HC1) between one of its members (patient 5), patient 8, and putative source, S2) [3]. All significant associations (i.e. those supported by 70% or greater bootstrap replicates) other than those between sequences from the same individual are noted.

It is clear that the sequences from all suspected contacts of the Glenochil Cohort from whom gag data was obtained cluster loosely together with Scottish intravenous drug users (IVDUs) and heterosexual



**Fig. 1.** (a) Maximum Likelihood Phylogenetic Tree of the p17 region of the gag gene from duplicate sequences from cases C1, C2, C4, C5, C6, SA, SB, triplicate sequences from SC and single sequences from cases C8 and SD. Included for comparison are single sequences from previously described patients 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14 (the Glenochil Cohort), a heterosexual contact of patient 5 (HC1), patient 8, putative source S2 and 5 haemophiliacs from Edinburgh, 7 Scottish IVDUs, and 3 heterosexual contacts of Scottish IVDUs. B1–10 refer to HIV-1 subtype B strains obtained from international databases (1 = JRCSF, 2 = HAN, 3 = RF, 4 = CDC4, 5 = OYI, 6 = JH3, 7 = NY5, 8 = MN, 9 = LAI, 10 = SF2, 11 = SC). Numbers refer to all bootstrap replicate values greater than 70% (except for those referring to duplicate samples from the same individual). Scales denote branch length measured as % difference in base composition (i.e. genetic divergence). (b) Maximum likelihood phylogenetic tree of the v3/v4 region of the env gene constructed from duplicate sequences from cases C1, C2, C5, C6, C8, SA, SB, single sequences from cases C3, C4, and triplicate sequences from SC. Included for comparison are single sequences from previously described cases 1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14 (the Glenochil Cohort), putative sources S1, S2, a heterosexual contact of patient 5 (HC1), and 16 haemophiliacs from Edinburgh, 3 Scottish IVDUs and 3 heterosexual contacts of Scottish IVDUs. See legend for Figure 1a for other details.

contacts of IVDUs, separate from both haemophiliacs and reference B subtypes. This is also true for putative sources SA, SB and SD. The exception is putative source SC whose association with the Glenochil Cohort is well supported by 89% bootstrap re-samplings.

#### Env sequence data

Figure 1b shows a maximum likelihood phylogenetic tree constructed from env sequences from suspected contacts and sources as indicated in Table 1. Subtype B isolates from international databases and local controls (Scottish IVDUs (3), Scottish heterosexual transmissions (3) and haemophiliacs from Edinburgh

(16)) are included for comparison. In addition, single representative sequences from the previously described Glenochil Cohort, a heterosexual transmission (HC1) from one member (patient 5), and putative sources S1 and S2 are also shown.

Bootstrap values greater than 70% were only obtained for previously described clusters of haemophiliac patients, duplicate samples from the same individual and members of the Glenochil Cohort (including HC1).

The majority of putative sources and contacts cluster loosely together with sequences from heterosexual transmissions from IVDUs. The exception to this is patient C6 whose sequences fail to link with any of the others on the tree.

Table 2. *Inferred amino acid sequence of a 100 bp fragment of the p17 region of the gag gene derived from single sequences from each member of the Glenochil Cohort and HC1(4), aligned with the homologous region from six replicate sequences from case SC*

Patient no.					
11	E--E--Q--N	--K--S--K-	-K--K--A--	Q--Q--V--A	--A--D--A-
13	.....	.....	.....	.....A.....	.....
14	.....	.....	.....	.....A.....	.....
10	.....	.....	.....	.....A.....	.....
7	.....	.....	.....	.....A.....	.....
12	.....	.....	.....	.....A.....	.....
9	.....	.....	.....	.....A.....	.....
4	.....	.....	.....	.....A.....	.....
6	.....	.....	.....	.....A.....	.....
3	.....	.....	.....	.....A.....	.....
2	.....	.....	.....	.....A.....	.....
5	.....	.....	.....	.....A.....	.....
1	.....	.....	.....	.....A.....	.....
HC1	K.....	.....	.....	.....A.....	.....
SC.1	.....	.....	.....Q..E..	.....A.....	.....-.-.-.
SC.2	.....	.....	.....Q..E..	.....A.....	.....-.-.-.
SC.3	.....	.....	.....Q..E..	.....A.....	.....-.-.-.
SC.4	.....	.....	.....Q..E..	.....A.....	.....-.-.-.
SC.5	.....	.....	.....Q..E..	.....A.....	.....-.-.-.
SC.6	.....	.....	.....Q..E..	.....A.....	.....-.-.-.
Patient no.					
11	-Q--Q--A--	A--A--A--G	--T--G--N-	-S--S-----	Q--V--S--Q
13	.....	.....	.....	.....	.....
14	.....	.....	.....	.....	.....
10	.....	.....	.....	.....	.....
7	.....	.....	.....	.....	.....
12	.....	.....	.....	.....	.....
9	.....	.....	.....	.....	.....
4	.....	.....	.....	.....	.....
6	.....	.....	.....	.....	.....
3	.....	.....	.....	.....	.....
2	.....	.....	.....	.....	.....
5	.....	.....	.....	.....	.....
1	.....	.....	.....	.....	.....
HC1	.....	.....	.....	.....	.....
SC.1	.....	.....	.....	.....S.....	.....
SC.2	.....	.....	.....	.....S.....	.....
SC.3	.....	.....	.....	.....S.....	.....
SC.4	.....	.....	.....	.....S.....	.....
SC.5	.....	.....	.....	.....S.....	.....
SC.6	.....	.....	.....	.....S.....	.....

... indicates identity with the first sequence.

.-. indicated the absence of an amino acid.

### Amino acid sequence data

Table 2 shows a comparison of a section of the amino acid sequence of the p17 region of the gag gene from all members of the Glenochil Cohort, including HC1, with six replicate sequences from patient SC. Although it is clear that there is remarkable similarity in the sequence from SC and the Glenochil Cohort over the majority of the region analysed, there is an eight amino acid insertion in the sequence from the entire Glenochil cohort which is absent from SC and

indeed all other sequences we have so far examined in this context.

### DISCUSSION

Following an infection control exercise conducted in Glenochil Prison in June/July 1993, 14 cases of HIV were identified and serological evidence conclusively showed that 6 had been infected whilst incarcerated there. Using molecular techniques we were able to show that 13 had been infected with a virtually

identical strain of HIV and concluded they had contracted this disease from a common source, most probably whilst in Glenochil. The fact that viral sequences from a known heterosexual transmission from 1/13 inmates also significantly clustered with sequences from the group gave us confidence that we could identify other transmissions and also the possible source of the infection. Independently from the original survey, we were provided with blood samples from two long-term infected individuals known to have been in Glenochil during the 'at risk' period of the first half of 1993. Analysis of these failed to indicate either of them as being the source of the infection [2, 3].

However, during the survey only 162 out of the possible 636 inmates who were in the prison between 1 January and 30 June 1993 were tested for antibodies to HIV. A subsequent survey indicated that the number of HIV positive prisoners in Glenochil during the 'at risk' period could have been in excess of 20 [7]. We therefore suspected that there were other HIV positive individuals who had not been identified who may represent either sources or contacts of the Glenochil Cohort.

A companion data linkage study identified a further 12 HIV positive patients thought to have been in Glenochil prison during the 'at risk' period. Of these, four were candidate sources whilst the rest candidate contacts. We were able to generate good sequence from at least one region from 11/12 (the exception being patient C7).

The gag region of the genome is commonly used for epidemiological studies since its rate of variation is intermediate to the other main genes env and pol. In this study it demonstrated that one long-term infected patient (SC) harboured virus very similar to the Glenochil Cohort, whilst all others tested showed no association at all. However when the env data was examined sequences from this individual were quite distinct from the cohort suggesting no association. We performed further sequencing to confirm our observations and also compared the amino acid sequence of SC with that of the Glenochil Cohort (Table 2). This clearly demonstrated that whilst the sequences from putative source C were virtually identical where comparisons could be made (i.e. along most of the region) they were however, 24 bases (8 amino acids) shorter. The presence of an extra eight amino acids is consistent amongst and characteristic of the Glenochil Cohort, and is absent from any other individuals we have so far examined in this context. In the light of

this we conclude that the data from the gag region does indeed support the observations from the env region that SC does not represent the source of the infection for the 13 members of the Glenochil Cohort. It also serves to illustrate the usefulness, not only of examining more than one region of the viral genome but also the application of visual assessment of sequence data in association with computer generated analyses.

Surprisingly, not only was there no significant association between any of the newly identified HIV positives and the original 13, but also there were no significant associations between any individuals other than the Glenochil Cohort. This suggests that with the exception of the Glenochil Cohort, all in this study appear to be independent infections.

Samples from the potential sources and contacts described here were obtained at various times up to 4 years after the Glenochil outbreak (Table 1). Sequence data described here were mainly obtained from plasma virus in contrast with our original investigation. Since viral sequences evolve over time and dominant plasma virus may vary from dominant proviral DNA, it is plausible that viral variation has accumulated to the extent that linkage is no longer possible. This possibility we do however find unlikely. Our own observations with HC1 show the sequence that was obtained at the end of October 1994 was still associated significantly with the presumed source (patient 5) [3] and that in case C2 the plasma virus and proviral DNA sequences were virtually identical: others have observed significant linkages after periods in excess of 10 years and after several transmission events [8]. Furthermore the use of two separate regions of the viral genome (env and gag) for comparisons further reduces the possibility of divergence impeding association.

We have examined sequence data from 27 of the 28 HIV positive individuals known or suspected to have been in the same prison around the same time. Since it is estimated that in Scotland in excess of 90% of HIV infected IVDUs inside and outside prisons have been identified [9] if any more individuals involved in the Glenochil outbreak remain undetected they are likely to be few in number.

## ACKNOWLEDGEMENTS

We wish to thank Dr S. Burns, Regional Virus Laboratory, City Hospital, Edinburgh, Mr P. McIntyre, Department of Medical Microbiology,

Ninewells Hospital and Medical School, Dundee and Mr J. Whitelaw, HIV Immunology, Royal Infirmary of Edinburgh for assistance with provision of samples. This work was funded by MRC programme grant No. G9209918.

## REFERENCES

1. Hutchinson SJ, Gore SM, Goldberg DJ, et al. Method used to identify previously undiagnosed infections in the HIV outbreak at Glenochil Prison. *Epidemiol Infect* 1999; **123**: 271–5.
2. Taylor A, Goldberg D, Emslie J, et al. Outbreak of HIV infection in a Scottish Prison. *BMJ* 1995; **310**: 289–92.
3. Yirrell DL, Robertson P, Goldberg DJ, McMenamin J, Cameron S, Leigh Brown AJ. Molecular investigation into outbreak of HIV in a Scottish prison. *BMJ* 1997; **314**: 1446–50.
4. Smith SW, Overbeek R, Woese CR, Gilbert W, Gillet PM. The genetic data environment and expandable GUI for multiple sequence analysis. *Comput Appl Biosci* 1994; **10**: 671–5.
5. Felsenstein J. PHYLIP manual version 3.52c. Berkeley University Herbarium, University of California, Berkeley, 1993.
6. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**: 783–91.
7. Gore SM, Bird AG, Burns SM, Goldberg DJ, Ross AJ, McGregor J. Drug injection and HIV prevalence in inmates of Glenochil prison. *BMJ* 1995; **10**: 293–6.
8. Leitner T, Escanilla D, Franzen C, Uhlen M, Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc Natl Acad Sci USA* 1996; **93**: 10864–9.
9. Goldberg D, Davis B, Allardice G, McMenamin J, Codere G. Monitoring the spread of HIV and AIDS in Scotland 1983–1994. *Scot MJ* 1996; **41**: 131–8.