Published in final edited form as:

Haemophilia. 2012 March ; 18(2): 291–299. doi:10.1111/j.1365-2516.2011.02620.x.

Molecular evidence of HIV-1 transmission in 20 Korean individuals with haemophilia: phylogenetic analysis of the *vif* gene

Y.-K. CHO*, Y. JUNG*, J.-S. LEE*, and B. T. FOLEY†

*Department of Microbiology, University of Ulsan College of Medicine, Seoul, South Korea

†HIV Databases, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, USA

Summary

To assess whether a genetic relationship exists between the viruses infecting HIV-positive patients with haemophilia and those infecting plasma donors, we determined the *vif* sequences in 169 individuals, including 20 haemophilia patients, 3 plasma donors, and 146 local controls. Twenty haemophilia patients were diagnosed with HIV-1 at 1–2 years after exposure to factor IX (FIX) manufactured in Korea, beginning in 1989–1990. Plasma samples from donors O and P were used to manufacture clotting factors including FIX used to treat the 20 haemophiliacs. The *vif* gene from frozen stored serum samples obtained 1–3 years after diagnosis was amplified by RT-PCR, and subjected to direct sequencing. Phylogenetic analysis revealed that *vif* sequences from 128 of the samples (including haemophilia patients and donors) belonged to the Korean subclade of HIV-1 subtype B (KSB). Sequences from 41 other participants were identified as subtype B, but outside the Korean subclade. Sequences of the *vif* gene from donors O and P plus the 20 individuals with haemophilia comprised two subclusters within KSB. In addition, signature pattern analysis disclosed the presence of conserved nucleotides at two positions in donors and haemophiliacs only. Together with information on KSB, dates of plasma donations and seroconversion of haemophilia patients, our results suggest that the haemophiliacs examined here became infected by viruses in the domestic clotting factor used for treatment.

Keywords

AIDS; haemophilia; HIV-1; phylogenetic analysis; signature pattern analysis; *vif*

Introduction

HIV-1 is an extremely variable virus [1], partly owing to lack of a proofreading mechanism and the consequential high error rate, a feature shared by all RNA viruses [2], high replication rates, and apparent high tolerance and selection for change [1]. Thus, some degree of genetic variation in HIV-1 is present among individuals related epidemiologically and even within individuals. Signature pattern analysis facilitates the identification of

©2011 Blackwell Publishing Ltd

Correspondence: Young-Keol Cho, Department of Microbiology, University of Ulsan College of Medicine, 88, Olympicro 43-gil, Songpogu, Seoul 138-736, Korea. Tel.: 82-2-3010-4283; fax: 82-2-3010-4259; ykcho2@amc.seoul.kr; and Brian T. Foley, HIV Databases, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, USA. Tel.: 1-505-665-1970; fax: 1-505-665-3493; btf@lanl.gov.

Disclosures

The authors stated that they had no interests which might be perceived as posing a conflict or bias.

particular sites in amino acid or nucleic acid alignments of variable sequences distinctly representative of a query set of sequences relative to a background set [3,4]. Phylogenetic methods have been employed to provide evidence for HIV-1 transmission events, not only for epidemiological purposes but also resolution of legal cases [5–12]. Recently, we applied these methods to investigate the genetic relatedness of the *pol* gene between plasma donors and haemophiliacs [12]. We concluded that *pol* analysis is more suitable in cases with a significant time lag between primary infection and sampling, similar to the present Korean group. In contrast, previous studies involving short time-lag intervals were mainly performed on *env* and *gag* genes [3,4,9], as *pol* is conserved to the highest degree among the nine genes [13]. Thus, a more conserved region of the genome appears appropriate for analysis with a longer lag time between infection and sampling. However, *vif* gene analysis for these purposes has not been performed to date.

In this study, we employed RT-PCR to amplify the *vif* genes from 169 patients, including 20 with haemophilia (HP), 3 plasma donors, 105 local controls infected with Korean subclade B of HIV-1 (KSB) and 41 infected with other strains of HIV-1. Phylogenetic and signature pattern analyses were applied to investigate the genetic relationships among the Korean *vif* sequences. The genes from 20 HP cases were identified as belonging to KSB, indicating infection in Korea by the domestic-made Factor IX (DCF). These nationwide and comprehensive *vif* data confirm previous findings, and further suggest that two of the four identified HIV-infected plasma donors were the most likely sources of viral infection in the 20 HPs.

Materials and methods

Plasma donors O and P: sources of HIV-1

In 1989, cash-paid plasma donations were obtained at several plasma centres run by Company X, which manufactures various blood products, including albumin and Factors, via non-heat inactivation. Four HIV-1-seropositive plasma donors, O, P, Q and R, who had sex with men, were diagnosed during primary infection. Details of these patients are described elsewhere [11,12] (Table 1).

HIV-1-infected HPs and local controls

Twenty patients with HP, designated HP1–HP20, tested positive for HIV-1 between 1990 and 1994 [11,12]. Patients had been treated primarily with imported clotting factors prior to the start of local DCF production. Three HPs (HP21–HP23) were infected with factors purchased outside Korea before 1994 (Table 1). Thirty-seven HIV-1 patients diagnosed up to 1990 were presumed to be infected with KSB [12], of which 27 were included in our study. Written informed consent was obtained from living HPs.

RNA preparation and *vif* gene amplification

Total RNA was extracted from 300 µL serum samples using a QIAamp Ultra sense Viral RNA kit (Qiagen, Hilden, Germany), as described previously [12]. The *vif* gene was amplified via nested PCR with TaKaRa R-Taq. First and second PCR reactions were performed in 20 µL and 50 µL reaction mixtures respectively. The outer primer pairs were 545 (5'-GCAGTACAAATGGCAGTATTCATC-3') and LA 106K (5'-TGRTAGAGRAACTTG ATGRTYCTT-3'), and the inner primer pairs were 547K (5'-GCTTCTCTGGAAAGGTGAA GG-3') and LA 102K (5'-TACAAGGAGTCTTGGG CTGACTTC-3') respectively. After initial denaturation at 95°C for 10 min, 35 cycles were run at 95°C for 30 s, 54°C for 30 s and 72°C for 2 min, followed by a final extension step at 72°C for 10 min. The second PCR was performed with 10 µL of the first PCR product. Cycling conditions were as follows: 95°C for 15 s, 57°C for 15 s, 72°C for 1 min and final

extension at 72°C for 10 min. Subsequent amplicons were directly sequenced using Applied Biosystems 3730XL (Macrogen Inc., Seoul, Korea).

Sequences

The GenBank accession numbers are AY581320-423, DQ072731-850, DQ295192-196, JF957864-8081 and JN049928-936.

Phylogenetic analysis

Sequences obtained from the earliest serum available from each patient and local controls were aligned to the HIV-1 subtype reference set from the HIV Sequence Database (http://hiv-web.lanl.gov/content/hiv-db/Subtype_REF/align.html), and phylogenetic trees built using the DNAML Maximum Likelihood method [14], PAUP maximum parsimony and PhyML programmes [12]. Trees generated using each of these methods produced the same donor O and donor P subclades, which were, in turn, within KSB of HIV-1 subtype B.

Viral signature pattern analysis

The VESPA (<http://www.hiv.lanl.gov/content/sequence/VESPA/vespa.html>) programme was used to identify sites within each group of sequences distinct for each group [9,12].

Statistical analysis

Data were expressed as means $\pm$ standard deviations. Statistical significance was estimated using the Student's two-tailed *t*-test, chi-square test and Fisher's exact test, using SPSS package version 12.0. Statistical significance was defined as $P < 0.05$.

Results

Origin of the Korean subclade of subtype B

A total of 327 HIV-1 infected patients were diagnosed in Korea up to 1993 [12]. Based on epidemiological data, 172 patients were domestically infected within Korea [12]. The main local circulating virus was classified as KSB, which had already begun to be transmitted in 1988 [12].

Molecular epidemiologic data on the *vif* gene

We determined the *vif* sequences in 169 patients, specifically, 105 LCs, 3 cash-paid plasma donors and 20 HPs (HP1–HP20) infected with KSB, and 41 non-KSB cases, including 3 HPs (HP21–HP23) (Table 2). Phylogenetic analysis revealed that 127 sequences belonged to KSB (Fig. 1). The 127 KSB sequences were subdivided into two clusters (marked 'O' and 'P' in Fig. 1), which included the 20 HPs and plasma donors O and P. Donors O and P were associated with 8 (1–4, 6, 8, 10 and 18 with cluster O) and 12 (5, 7, 9, 11–17, 19 and 20 with cluster P) HPs respectively. These findings were similar to those obtained for the *pol* gene [12]. However, the earliest of all 16 sequences in HP-4 did not belong to cluster O (Fig. 1). Only one of the 16 sequences, 6249 (JF958031), belonged to cluster O. In addition, we observed separation between the blood donor (94JWK12) and two recipients (91HJY7/92KGS) in the tree, although seven pairs clustered together (Fig. 1).

Nucleotide/amino acid signature patterns

Assessment of signature pattern nucleotides in clusters O and P revealed that seven positions showed statistically significant relationships with nucleotide sequences in the 118 and 114 LCs ($P < 0.01$), along with a significant difference between the two clusters ($P < 0.01$), except position 5420 (Table 3). Nucleotide changes at four positions were non-synonymous,

whereas those at three positions were synonymous, compared with the consensus nucleotide sequences of the LCs. Interestingly, none of the LCs contained nucleotide A at positions 5070 and 5253 (Table 3), which was specifically identified in clusters O and P respectively (Fig. 2).

With regard to signature pattern amino acids, two positions in clusters O (K141T and G185E) and P (T47P and R127H) showed significantly high frequencies, compared with those in the background set ($P < 0.01$) (Fig 2). K141T was detected among all patients of cluster O, although minor in HP-4 (1/15), but absent in cluster P, and present in 17 among the background set ($P < 0.0001$). G185E was detected in all patients of cluster O, but not in cluster P, and present in 12 among the background set ($P < 0.0001$). T47P was detected in all patients of cluster P (T47S in HP-17 in all three amplicons in 1993 and T47P in 17 amplicons after 2003) and in two among the background set ($P < 0.0001$), but not among the 179 *vif* amplicons in cluster O.

Nucleotide A at position 5070 is present in all patients comprising cluster O, in contrast to other Korean patients

We determined 1475 *vif* sequences, irrespective of subtype. Surprisingly, nucleotide A at position 5070 was conserved in the nine patients constituting cluster O. We obtained 23 *vif* amplicons from donor O between 1991 and 1999. Except for the 179 *vif* amplicons from the nine cluster O patients, all 1296 *vif* amplicons in the remaining 160 patients contained nucleotide G only.

Sequence identities of donors O and P

We sequenced 23 and 4 *vif* genes from seven donor O and two donor P samples respectively. In donor O, the intrapersonal sequence identities between five amplicons obtained in October 1991 were 100–99.3%, and ranged 97.9–98.3% between 1991 and January 2002. In donor P, the intrapersonal identity between two amplicons in February 1993 and between two amplicons in October 1993 was 100% and 99.8%, respectively, ranging from 99.7% to 99.8% between February 1993 and October 1993. The *vif* sequence obtained from donor O in October 1991 displayed $98.94 \pm 0.53\%$ identity to that of cluster O (range 99.48–97.75%), and that from Donor P in February 1993 displayed $98.86 \pm 0.74\%$ identity to that of cluster P (range 99.65–97.24%). Samples from four HPs obtained after 2000 exhibited the lowest sequence identity (97.24–98.61%). With the exception of these four HP samples, sequence identities in clusters O and P were estimated as $99.11 \pm 0.74\%$ and $99.19 \pm 0.37\%$ respectively. Notably, sequence 6693 (JF957908) obtained from donor O showed 100% identity with that from HP-1 (AY581322).

Correlation between sampling intervals and number of nucleotide differences

We determined the correlation between sampling intervals from the accident and number of nucleotide differences, compared with that in the corresponding plasma donor. Four patients for whom first sampling was performed in 2002 or 2003 displayed the highest difference of more than 8 (range 2–16, median 5.5). The correlation coefficient, γ , was 0.84 ($P < 0.0001$) (Fig. 3).

Absence of discordant subtypes between paired *vif* and *pol* genes

In a previous study, we determined the *pol* gene sequences in 109 patients [12], which are included in the present study. No discordant subtypes were evident between paired *vif* and *pol* genes.

Discussion

Korean subclade B of HIV-1 forms a distinct monophyletic clade that is not related to any of the international sequences from Los Alamos Database as of December 2010 [12]. For the *nef* gene, DNA distances within KSB were 1.9–8.8% [15]. In KSB-infected homosexual Korean men, the diversities in *env* genes V1 to V5 were significantly lower than those in other countries, and recorded as 2.1% and 2.5% respectively [16]. In total, 11 residues of KSB *vif* sequences have been identified as the Korean clade signature [17]. In the present study, we amplified the *vif* gene from sera obtained earlier than PBMC samples used in prior studies [11,15–18], facilitating extension of the follow-up period from 1991. Within these KSB with highly limited variations, donors O and P formed two clusters designated O and P respectively. However, we did not include bootstrap values in the phylogenetic tree, as there is low bootstrap support for each subclade within KSB, which is already a subclade of subtype B. There are only a few sites in the data, which link donor O to some individuals with HP and donor P to others, and bootstrap re-sampling obviously tends to erase this signal. The major point is not which of these two donors infected which of the 20 people with HP, but that these 20 people were infected in Korea with Korean-made clotting factor. Both O and P were cashpaid donors who sold plasma used for the production of clotting factors, and individuals with HP seroconverted within a few months of seroconversion of donors O and P after using the Korean-made factor.

Nucleotide A at position 5070 induces a non-synonymous change in integrase (V281M). The 100% presence of V281M in cluster O and absence from the remaining 160 patients may have resulted from the higher degree of genetic conservation of *vif* and *pol*, compared with other genes [10,12], which may also be an advantage of the highly conserved gene analysis. Consistent with our data, another study recently revealed the absence of nucleotide A at position 5070 in 75 Korean patients [18]. In keeping with the Florida dental cases ($n = 35$) [9], our study leads to the more stringent conclusion that the conserved nucleotide (A at 5070 and 5253 in clusters O and P respectively) present in each sequence set is completely absent in background sequences ($n = 118$).

In the present study, sera sampled from different individuals at various time-points were used. Despite the variable time-points of sampling and possibility of super-infection by both donors, a few signature pattern amino acids appeared well conserved. Thus, epidemiologically linked sequences were clustered among 127 KSB sequences (Fig. 1). However, it is also possible that most sequences from HP-4 did not belong to cluster O (Fig. 1), although all *nef* [11], *pol* [12] and 5' LTR/full *gag* (unpublished data) sequences from HP-4 belonged to cluster O. As an extreme example, discordant subtypes in paired *pol* and *env* sequences in Portuguese HIV-1 patients were 25.7% [19]. To date, we have not identified discordant subtypes or subclades in Korean patients. Therefore, this separation in the phylogenetic tree between cluster O and HP-4 is natural due to the long time lag (>11 years) between primary infection and sampling. Similarly, a genetic relationship between *gag* and *env* could not be established in another study in which specimens were collected > 3 years after transmission [20].

In addition to genetic relatedness of HIV-1 between HPs and DCF, Korean NIH reported a hepatitis A virus spread among 57 Korean haemophiliacs [21]. In an earlier study, we showed that all HIV-1-positive HPs were infected with HCV [11].

In conclusion, our findings indicate a clear link between DCF and the 20 HIV-1-infected HPs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a grant from the Korean Society of Ginseng funded by the Korea Ginseng Corporation (2008–09). Cho designed the research study. Cho, Lee and Jung performed the research. Cho and Foley analysed the data and wrote the paper.

References

1. Korber, BT.; Foley, B.; Gaschen, B.; Kuiken, C. Epidemiological and immunological implications of the global variability of HIV-1. In: Pantaleo, G.; Walker, BD., editors. *Retroviral Immunology: Immune Response and Restoration*. Totawa, NJ: Human Press; 2001. p. 1-32.
2. Darke JW. Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci USA*. 1993; 90:4171–75. [PubMed: 8387212]
3. Korber B, Myers G. Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Res Hum Retroviruses*. 1992; 8:1549–60. [PubMed: 1457200]
4. Leitner T, Escanilla D, Franzén C, Uhleén M, Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc Natl Acad Sci USA*. 1996; 93:10864–69. [PubMed: 8855273]
5. Metzker ML, Mindell DP, Liu XM, Ptak RG, Gibbs RA, Hillis DM. Molecular evidence of HIV-1 transmission in a criminal case. *Proc Natl Acad Sci USA*. 2002; 99:14292–97. [PubMed: 12388776]
6. Albert J, Wahlberg J, Leitner T, Escanilla D, Uhlén M. Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 *pol* and *gag* genes. *J Virol*. 1994; 68:5918–24. [PubMed: 7520096]
7. Blanchard A, Ferris S, Chamaret S, Guétard D, Montagnier L. Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients. *J Virol*. 1998; 72:4537–40. [PubMed: 9557756]
8. Holmes EC, Zhang LQ, Simmonds P, Rogers AS, Brown AJ. Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon. *J Infect Dis*. 1993; 167:1411–14. [PubMed: 8501332]
9. Ou CY, Ciesielski CA, Myers G, et al. Molecular epidemiology of HIV transmission in a dental practice. *Science*. 1992; 256:1165–71. [PubMed: 1589796]
10. Cichutek K, Merget H, Norley S, et al. Development of a quasispecies of human immunodeficiency virus type 1 in vivo. *Proc Natl Acad Sci USA*. 1992; 89:7365–69. [PubMed: 1502146]
11. Cho YK, Foley BT, Sung H, Kim YB, Kim JH. Molecular epidemiologic study on HIV- 1 outbreak in hemophiliacs B through clotting factor 9 after 1990. *Vox Sang*. 2007; 92:113–20. [PubMed: 17298572]
12. Cho YK, Jung YS, Foley BT. Phylogenetic analysis of full-length *pol* gene from Korean hemophiliacs and plasma donors infected with Korean subclade B of HIV-1. *AIDS Res Hum Retroviruses*. 2011; 27:613–21. [PubMed: 21062223]
13. Abecasis AB, Vandamme AM, Lemey P. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. *J Virol*. 2009; 83:12917–24. [PubMed: 19793809]
14. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003; 52:696–704. [PubMed: 14530136]
15. Kang MR, Cho YK, Chun J, et al. Phylogenetic analysis of the *nef* gene reveals a distinctive monophyletic clade in Korean HIV-1 cases. *J AIDS and Hum Retrovirology*. 1998; 17:58–68.
16. Kim GJ, Nam JG, Shin BG, et al. National survey of prevalent HIV strains: limited genetic variation of Korean HIV-1 clade B within the population of Korean men who have sex with men. *J Acquir Immune Defic Syndr*. 2008; 48:127–32. [PubMed: 18317230]

17. Park CS, Kim MS, Lee SD, Kim SS, Lee KM, Lee CH. Molecular phylogenetic analysis of HIV-1 *vif* gene from Korean isolates. *J Microbiology*. 2006; 44:655–59.
18. Kim JY, Kim EJ, Choi JY. Genetic variation of the HIV-1 integrase region in newly diagnosed antiretroviral drug-naïve patients with HIV/AIDS in Korea. *Clin Microbiol Infect*. 2010 Oct 14;10.1111/j.1469-0691.2010.03392
19. Abecasis AB, Martins A, Costa I, et al. Molecular epidemiological analysis of paired pol/env sequences from portuguese HIV-1 patients. *AIDS Res Hum Retroviruses*. 2011; 27:803–5. [PubMed: 21198411]
20. Birch CJ, McCaw RF, Bulach DM, et al. Molecular analysis of human immunodeficiency virus strains associated with a case of criminal transmission of the virus. *J Infect Dis*. 2000; 182:941–44. [PubMed: 10950794]
21. Jee YM, Go U, Cheon D, et al. Detection of hepatitis A virus from clotting factors implicated as a source of HAV infection among haemophilia patients in Korea. *Epidemiol Infect*. 2006; 134:87–93. [PubMed: 16409654]

Fig. 1.

Phylogenetic tree of *vif* sequences from 20 individuals with haemophilia (HPs), plasma donors O, P and R, 104 local control patients infected with Korean subclade of HIV-1 subtype B (KSB) and 34 non-KSB patients, including HP-21, 22 and 23. Sequences were aligned with the HIV-1 subtype reference set from the HIV Sequence Database, and a phylogenetic tree built using the PHYLIP Dnadist (F84 model, Ts:Tv 1.7) and Neighbour programmes.²³ Sequences from 127 patients belonged to KSB and 34 sequences were classified as outside KSB. Two clusters, O (Donor O, HPs-1–4, 6, 8, 10 and 18) and P (Donor P, HPs-5, 7, 9, 11–17, 19 and 20), were clearly demarcated from KSB-infected local control patients. Seven linked pairs clustered together (04CMU2/94JOE12, HP-21/01HSJ4, 92KJS3/97KMH4, 00CAS9/04AYS12, 91KJin7/01CJI3, 92JYS6/91SNS12, and 91HJY7/92KGS). The two digits after the number (1–20) of HPs represent the year of sampling.

Fig. 2.

Alignment of Vif protein sequences of 20 haemophiliacs, 2 plasma donors and 106 local controls infected with KSB (LCs). Significantly high frequencies of amino acids at two positions each in clusters O (K141T and G185E) and P (R127H and T47P) were observed, compared to the earliest sequences in 106 LCs respectively ($P < 0.001$ for all). Remarkably, donors O and P were the earliest patients containing K141T, G185E and T47P among the 106 LCs. Only one patient displayed T47P among the 106 LCs, while 17 displayed K141T. Among these, donor O was the earliest individual. The dot indicates the same sequence with consensus sequences. One patient (06LSang8) displayed insertion of a single amino acid (D) between positions 60 and 61. 04KKT7 displayed an extension of three amino acids at the terminal codon (WSS). Two amino acids (K141T and G185E) were 100% conserved in all patients of cluster O, whereas there was no such case in the 106 local control patients.

Fig. 3.

Correlation between interval (months) from the accident (around January 1990) to the sample used and number of nucleotide differences in the 20 HPs, compared with corresponding plasma donors. Four patients (HPs-4, 15, 19 and 20) providing serum samples after 2002 displayed the highest differences (8). The correlation coefficient, γ , was 0.84 ($P < 0.0001$).

Table 1

Date of diagnosis and sampling in people with haemophilia and plasma donors.

Patient	Date of HIV-1 diagnosis	Number of amplicon [*]	Date of sampling	GenBank accession no. [*]
Twenty people with haemophilia infected with HIV-1 in Korea				
HP-1	Nov 1990	3	Aug 1991	JF958013-15
HP-2	Feb 1991	3	Jan 1992	JF958023-25
HP-3	Apr 1991	3	Aug 1992	JF958026-28
HP-4	Aug 1991	2	Oct 2002 [†]	JF958029-30
HP-5	Jan 1992	7	Aug 1992	JF958032-38
HP-6	Jan 1992	3	Sep 1992	JF958039
HP-7	Jan 1992	4	Oct 1992	JF958040-43
HP-8	Feb 1992	2	May 1992	JF958044-45
HP-9	Feb 1992	6	Feb 1993	JF958046-51
HP-10	Mar 1992	1	Aug 1992	JF958052
HP-11	Feb 1992	3	Jul 1992	JF958053-55
HP-12	Feb 1992	3	May 1992	JF958056-57
HP-13	Feb 1992	3	Jul 1992	JF958058-60
HP-14	Feb 1992	1	Mar 1992	JF958061
HP-15	Sep 1992	3	Mar 2003 [†]	AY581358
HP-16	Dec 1992	3	Dec 1992	JF958062-64
HP-17	Feb 1993	3	Mar 1993	JF958065-67
HP-18	Mar 1993	3	May 1993	JF958068-70
HP-19	Jul 1993	2	Nov 2002 [†]	JF958074-75
HP-20	Aug 1994	6	Oct 2002 [†]	JF958076-78
Three people with haemophilia infected with HIV-1 outside Korea				
HP-21	Jun 1987	3	Nov 1991	JF957893
HP-22	Mar 1991	4	Jan 2003	DQ072778-81
HP-23	Oct 1994	3	Nov 2002	DQ072825-27
Four cash-paid plasma donors with preseroconversion at plasma centre				
Donor-O	Nov 1989	5	Oct 1991	JF957908-12
Donor-P	Mar 1990	3	Feb 1993	JF957935-37

Patient	Date of HIV-1 diagnosis	Number of amplicon*	Date of sampling	GenBank accession no.*
Donor-Q	Feb 1992	0	Jun 1994	No data
Donor-R	Jun 1992	2	Jun 1992	JN049934-35

* Number of PCR amplicon and GenBank no. of the earliest sample.

† Samples obtained after 2000 were used in four people with haemophilia, suggesting a long time lag.

Table 2

Distribution of *vif* gene sequenced by year.

Year of diagnosis	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995-04	Total
KSB			1	11	14	12	28	3	1	38	108
People with haemophilia-KSB					1	4	11	3	1		20
Non-KSB	1	5	3	3	10	2	7	3	2	5	41
Total	1	5	4	14	25	18	46	9	4	43	169

KSB, Korean subclade B of subtype B.

Table 3

Frequency of signature pattern nucleotides from plasma donors O and P.

Position in HIV-1 NL4-3	5070	5179	5253	5420	5462	5553	5594
Signature (in donor O)	A _{0.60}				C _{1.00}		A _{1.00}
Cluster-O (<i>n</i> = 9)	1.0				0.89		1.00
Background (<i>n</i> = 119)	0.00				0.15		0.10
Non-synonymous changes					K141T		G185E
Signature (in donor P)		C _{1.00}	A _{1.00}	A _{1.00}		A _{1.00}	
Cluster-P (<i>n</i> = 13)		0.92	1.00	0.54		1.00	
Background (<i>n</i> = 115)		0.01	0.00	0.11		0.03	
Non-synonymous changes		T47P		R127H			

Signature patterns in clusters O (Oct 1991) and P (Feb 1993) show sets of five and three nucleotides that were the most common, but atypical in background sequences (a majority signature) respectively. A₆₀ at 5070 and C₁₀₀ at 5179 denote A and C at 3 of 5 and 3 of 3 sequences respectively. Cluster-O and -P lines depict the frequencies of these nucleotides among viral sequences of 8 and 12 people with haemophilia respectively. Six of seven nucleotide positions revealed significantly higher frequency in clusters O or P, compared with background (specifically, three positions in cluster O and four in cluster P, relative to background; *P* < 0.001 for all). Six of the seven sites met the criteria for a stringent signature, i.e., perfectly conserved in all donors O and P viral sequences. Notably, nucleotide A at positions 5070 and 5253 was detected only in clusters O and P respectively. Four positions with non-synonymous changes are presented in bold.