# Phylogenetic Analysis of the Earliest *nef* Gene from Hemophiliacs and Local Controls in Korea

Young-Keol Cho,[1] Jung-Eun Kim,[1] and Brian T. Foley[2]

## Abstract

Twenty hemophiliacs (HPs) were found to have human immunodeficiency virus type-1 (HIV-1) 1–2 years after exposure to Factor IX manufactured in Korea in late 1989. Plasma samples collected from donors O and P during their pre-seroconversion acute infection stage were used to manufacture clotting factors, including Factor IX, to treat these patients. To assess whether a genetic relationship exists between the viruses infecting HIV-1-positive HPs and those infecting plasma donors, we evaluated the *nef* sequences in 216 individuals. Frozen-stored serum samples obtained 1–3 years after the diagnosis of HIV-1 in the 20 HPs were used for amplification of the *nef* gene by reverse transcriptase–polymerase chain reaction, and amplicons were subjected to direct sequencing. Phylogenetic analysis revealed that the *nef* sequences from 143 of the samples belonged to the Korean subclade of HIV-1 subtype B (KSB). Sequences of the *nef* gene from donors O and P and the 20 HPs comprised two subclusters within KSB together with several local control (LC) sequences. In addition, signature pattern analysis revealed the presence of conserved nucleotides at eight positions in donors O and P compared with LCs ($p < 0.01$). These nationwide and comprehensive *nef* data support the previous conclusion that HPs were infected with HIV-1 from the clotting factor, although the stringency of *nef* is weaker than for the *pol* and *vif* genes.

**Key words:** hemophilia; HIV-1; *nef*; phylogenetic analysis; signature pattern analysis

## Introduction

**P**HYLOGENETICS AND SIGNATURE PATTERN ANALYSIS have been employed to provide evidence for human immunodeficiency virus type-1 (HIV-1) transmission events, not only for epidemiological purposes but also for resolution of legal cases.[1–9] Previous studies in patients with short time lag intervals from HIV infection to diagnosis have been performed via the detection of the *env* and *gag* genes,[3,4,9] since *pol* is the most highly conserved of the nine HIV genes.[10] Recently, we used a similar method to determine the genetic relatedness of the *pol* and *vif* genes between plasma donors and hemophiliacs (HPs).[11,12] These studies showed that *pol* and *vif* gene analyses are more suitable in cases with significant time lags between primary infection and sampling because these gene regions evolve more slowly than *env* and are subjected to less individual patient-specific selection pressures. On September 29, 2011, the Supreme Court of Korea ruled that Company X was responsible for the infection with HIV-1 of Korean HPs B after reviewing our studies on *pol* and *vif* genes (Case no. 2008 Da16776).[11,12] At present, cash-paid plasma donation is not permitted by law, and heat inactivation is not included in the original procedure, even though RNA from the Hepatitis A virus was amplified in vials of Factors VIII and IX from 1998 to 1999.[13]

In our present study, we amplified the *nef* genes by reverse transcriptase–polymerase chain reaction (PCR) using serum samples obtained 1–3 years after the diagnosis of HIV in 216 patients, including 20 HPs and 4 plasma donors, and compared the *nef* sequence identity with previously detected *pol* and *vif* genes.[11,12] The use of plasma donors complements the findings of previous studies that lacked plasma donor O sequences.[14] It is suspected that viruses from donors O and P contributed to HIV-1 infections in HPs in Korea.[9,11,12] We performed phylogenetic and signature pattern analyses to the samples to investigate the genetic relationships in the *nef* sequences among plasma donors, HPs, and local controls (LCs), and compared the results with those of previous analyses that used the *pol* and *vif* genes. Although analyses of *nef* gene sequences are weaker than those of other genes used in previous reports, this nationwide and comprehensive *nef* gene analysis supports previous findings that two of the

---

[1]Department of Microbiology, University of Ulsan College of Medicine, Seoul, South Korea.
[2]HIV Databases, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico.

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 performed at several plasma centers run by Company X, which manufactures various blood products. Four HIV-1-seropositive plasma donors, O, P, Q, and R, were found to have HIV-1

during seroconversion. Details of these patients are described in Table 1.[9,11,12]

### HIV-1-infected HPs and LCs

Twenty HPs, designated HP1–HP20, tested positive for HIV-1 between 1990 and 1994 (Table 2).[9,11,12] In addition, 196 patients (including 4 plasma donors, 3 [HP21–23] infected outside Korea, and 44 with non-subtype B) were included in the analysis (Table 3). Finally, 37 HIV-1 patients diagnosed before 1991 were presumed to be infected with Korean

TABLE 1. CHARACTERISTICS OF CLOTTING FACTOR IX AND FOUR CASH-PAID PLASMA DONORS

| Plasma donor/ date sampled | CD4+ T cell count | Treatment | GenBank accession no. | | |
| | | | nef | vif[12] | Full-length pol[11] |
|---|---|---|---|---|---|
| Donor-O | | | | | |
| 16 Oct 1989 | Negative | | | | |
| 30 Nov 1989 | Diagnosis | FIX lot no (9005) | | | |
| Jun 1990 | 1178 | | | | |
| Mar 1991 | 388 | | | | |
| Oct 1991 | 409 | | HM237141-44 | JF957908-12 | |
| Jan 1992 | 603 | ZDV & KRG | HM237145-47 | JF957913-14 | HM210883-34 |
| Apr 1992 | 271 | ZDV & KRG | | | |
| Feb 1993 | 395 | ZDV & KRG | HM237148-55 | JF957915-21 | |
| Feb 1999 | 299 | ddI & 3TC | FJ615003-05 AY899339-40 AY584756-57 | JQ248127-28 | |
| Feb 2001 | 314 | KRG | FJ615006 | JQ248129-31 | |
| Jun 2001 | 423 | KRG | AY584758-59 | AY581320-21 | |
| Jan 2002 | 216 | KRG | AY121450-51 AY260769-70 | JQ248132-33 | HQ026610-11 |
| Donor-P | | | | | |
| Early 1990 | (Primary infection symptoms; fever for 10 days) | | | | |
| 30 Mar 1990 | Indeteminate by Western blot, FIX lot no (9006 & 0001) | | | | |
| Apr 1991 | 1085 | | | | |
| Sep 1991 | 632 | | AY363309-10[9] | | |
| Mar 1992 | 751 | | | | |
| Feb 1993 | 779 | | HM237156-59 | JF957935 JQ248134 | HM210885 |
| Jul 1993 | 391 | | AF063929[14] | | |
| Oct 1993 | 433 | | HM237160-62 | JF957936-37 | HM210886 |
| Donor-Q | | | | | |
| 06 Feb 1992 | Diagnosis | | | | |
| Jan 1994 | 263 | | | | |
| Jun 1994 | 271 | | AF063918 | | |
| Jan 1999 | | TB-medication | | | |
| Donor-R | | | | | |
| Mar 1992 | (Admission for 10 days for high fever) | | | | |
| 03 Jun 1992 | Diagnosis | | | | |
| Jun 1992 | 511 | | HM237163-65 | JN049934-35 | HM210887-88 |
| Dec 1992 | 1105 | | | | |
| Mar 1993 | 900 | | HM237166-69 | JQ248517-19 | |
| Feb 1997 | 1002 | | | | |

First FIX had been manufactured from 1986 (NMC Medical record showed that HP13 got injected with two vials of FIX on 7 March 1987). Second FIX has been manufactured since November 1989. Lot number 9005 was manufactured on 12 December 1989. Lot numbers 9001–9004 were not described in the report by Company X.

FIX, Clotting Factor 9; ZDV, zidovudine; KRG, Korean red ginseng.

TABLE 2. DATE OF DIAGNOSIS AND SAMPLING IN HEMOPHILIA PATIENTS AND PLASMA DONORS

| Hemophiliacs no. | Date of HIV-1 diagnosis | Start of HAART | Date of sampling | Present status | GenBank accession no.[a] |
|---|---|---|---|---|---|
| Twenty hemophiliacs infected with HIV-1 in Korea | | | | | |
| HP1 | Nov 1990 | None | Feb 1991 | Death | JN247463-71 |
| HP2 | Feb 1991 | Oct 1993 | Jan 1992 | HAART | JN247472-75 |
| HP3 | Apr 1991 | Mar 2003 | Aug 1992 | HAART | JN247476-79 |
| HP4 | Aug 1991 | Aug 1996 | Oct-2002[b] | HAART | JN247480-81 |
| HP5 | Jan 1992 | None | Aug 1992 | Death | JN247482-85 |
| HP6 | Jan 1992 | 1999 | Sep 1992 | HAART | JN247486-89 |
| HP7 | Jan 1992 | Jan 1997 | Oct 1992 | HAART | JN247490-91 |
| HP8 | Feb 1992 | Jun 2007 | May 1992 | Death | JN247492-512 |
| HP9 | Feb 1992 | Jul 2007 | Feb 1993 | HAART | JN247513-15 |
| HP10 | Mar 1992 | Jan 2005 | Aug 1992 | HAART | JN247516-21 |
| HP11 | Feb 1992 | Nov 2003 | Jul 1992 | HAART | JN247522-24 |
| HP12 | Feb 1992 | Apr 2002 | May 1992 | HAART | JN247525-27 |
| HP13 | Feb 1992 | Feb 2006 | Jul 1992 | HAART | JN247528-31 |
| HP14 | Feb 1992 | Apr 2009 | Mar 1992 | HAART | JN247532-33 |
| HP15 | Sep 1992 | Early 2011 | Mar 2003[b] | HAART | AY260782-83 |
| HP16 | Dec 1992 | 2006 | Dec 1992 | HAART | JN247534-36 |
| HP17 | Feb 1993 | Oct 2001 | Mar 1993 | HAART | JN247537 |
| HP18 | Mar 1993 | Apr 1993 | May 1993 | Death | JN247538 |
| HP19 | Jul 1993 | 1999 | Dec 1994 | HAART | AF063921 |
| HP20 | Aug 1994 | Nov 1994 | Oct 2002[b] | HAART | JN247539-41 |
| Three hemophiliacs infected with HIV-1 outside Korea | | | | | |
| HP21 | Jun 1987 | Not yet | Nov 1991 | HAART | AF238268 |
| HP22 | Mar 1991 | Dec 1996 | Dec 1994 | HAART | AF462737 |
| HP23 | Oct 1994 | Feb 2002 | Nov 2002 | HAART | AY221714 |

[a]GenBank accession no. from the earliest sample.
[b]Samples obtained after 2000 were used for three people with hemophilia.
HIV-1, human immunodeficiency virus type-1; HP, hemophiliac; HAART, highly active antiretroviral therapy.

subclade of HIV-1 subtype B (KSB),[11] of which 27 were included in our study. Of the 20 HPs in our cohort, 4 died. Two of these deaths occurred before the implementation of highly active antiretroviral therapy (HAART), HP8 owing to poor compliance and intracranial hemorrhage in August 2011, and HP18 from full-blown AIDS in December 2011. Written informed consent was obtained from the 16 living HP patients who remained to participate in this study.

### RNA preparation and nef gene amplification

Total RNA was extracted from 300 $\mu$L serum samples using a QIAamp Ultra sense Viral RNA kit (Qiagen, Hilden, Germany), as described previously.[11,12] The *nef* gene was amplified via nested PCR with TaKaRa Taq™ (Takara Biotechnology, Dalian, China). First and second PCRs were performed in 20 and 50 $\mu$L reaction mixtures, respectively.[11,12] Subsequent amplicons were directly sequenced using Applied Biosystems 3730XL (Macrogen, Inc., Seoul, Korea)

### Sequences of the nef gene

The GenBank accession numbers for the genes investigated in this study are as follows: AF063921, AF462696, AF462699, AF462702, AF462704, AF462708, AF462717, AF462723, AF462726, AF462767, AF462770, AF462780, AF462788, AF462791, AY121440, AY121441, AY121446, FJ615009, AY121477, AY121478, AY260782, AY363365, AY584754, AY584800, AY899350, AF063918, AY899359, AY899370, AY899371, AY899379, AY899382, AY899391, AY899401, AY899378, AY899402, DQ121756, DQ121762, DQ121819, DQ121870, DQ121882, DQ121886, DQ121902, DQ121912, DQ121926-DQ121933, DQ121957, DQ121963, DQ121965, DQ121984, DQ121986, DQ121987, DQ121991, DQ121999, DQ122019, DQ122023, DQ122026, DQ122029, DQ122041, DQ122044, DQ122045, DQ122058, DQ122061, DQ122069, DQ122071, DQ122075, DQ122083, DQ122088, DQ122093, DQ122096, DQ122098, DQ122100, DQ122102, DQ122108, DQ122110, DQ122114, DQ122116, DQ122118, DQ122122,

TABLE 3. DISTRIBUTION OF 216 HUMAN IMMUNODEFICIENCY VIRUS TYPE-1 INFECTED PERSONS WITH *NEF* GENE SEQUENCED

| Year of diagnosis | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993–2006 | Total |
|---|---|---|---|---|---|---|---|---|---|
| KSB-local control | | | 1 | 11 | 14 | 12 | 29 | 56 | 123 |
| KSB-HPs | | | | | 1 | 4 | 11 | 4 | 20 |
| Western B[a] | 1 | 5 | 2 | 1 | 4 | 3 | 7 | 6 | 29 |
| Non-B | | 1 | 5 | 6 | 10 | | 6 | 16 | 44 |
| Total | 1 | 6 | 8 | 18 | 29 | 19 | 53 | 82 | 216 |

[a]Western B includes three hemophiliacs (HP21–HP23) infected outside Korea.
KSB, Korean subclade B of subtype B.

DQ400965, DQ400967, FJ615085, HM237141-HM237162, HM369811, HM369938, JQ007433-JQ007505, JN247463-JN247541, and JN377681-JN377703.

### Phylogenetic analysis

Sequences obtained from the earliest serum available from each donor, patient, and LC 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). Phylogenetic trees were built using the DNAML Maximum Likelihood method based on a data-specific model,[15] PAUP maximum parsimony, and PhyML programs.[16] Trees generated using each of these methods produced the same donor O and donor P subclades, which were, in turn, within the KSB of HIV-1 subtype B.

### Viral signature pattern analysis

The VESPA (www.hiv.lanl.gov/content/sequence/VESPA/vespa.html) program was used to identify sites that were distinct within each group of sequences.[7]

### Statistical analysis

Data are expressed as means ± 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 KSB

A total of 327 HIV-1-infected patients were diagnosed in Korea before 1994.[11] On the basis of epidemiological data, 172 of these cases were domestically infected within Korea.[11] The main local circulating virus was classified as KSB, which had already begun to be transmitted in 1988. The earliest *pol* and *vif* genes were isolated from a patient diagnosed in 1988 (HQ026557 and JF957897, respectively).[11,12]

### Molecular epidemiologic data on the nef gene

We determined the *nef* sequences in 216 patients, including 189 LCs, 4 cash-paid plasma donors, 20 HP patients (HP1–HP20) infected with KSB (Table 2), and 3 (HP21–HP23) infected outside Korea. Phylogenetic analysis revealed that 143 sequences were positive for KSB (Fig. 1), and 73 sequences were non-KSB. The 143 KSB sequences were subdivided into two clusters (marked "cluster O" and "cluster P" in Fig. 1) through their similarity to plasma donors O and P, respectively. We did not include bootstrap values in the phylogenetic tree because there is low bootstrap support for each subclade within KSB, which is already a subclade of subtype B.[12]

### Nucleotide/amino acid signature patterns

Assessment of signature pattern nucleotides in clusters O and P revealed that eight positions showed statistically significant relationships with nucleotide sequences in 127 and 131 LCs ($p < 0.01$), respectively, along with a significant difference between the two clusters ($p < 0.01$; Table 4). Nucleotide changes at five positions were nonsynonymous, whereas those at three positions were synonymous, compared with the consensus nucleotide sequences of the LCs (Table 4). Among the five nonsynonymous changes, two and three positions were in clusters O (I113V and V145T) and P (I165V, C174S, and V194E), respectively ($p < 0.01$) (Fig. 2).

### Sequence identities of donors O and P

We sequenced 37 and 9 *nef* genes in seven and three samples from donors O and P, respectively. In donor O, the intrapersonal sequence identities between four amplicons obtained in October 1991 were 99.2 to 98.6%, and decreased to 96.0–94.7% between October 1991 and January 2002. In donor P, the intrapersonal identity between two amplicons in September 1991 and between four amplicons in October 1993 was 100% and 100 to 99.5%, respectively, and decreased to 99.0–99.4% between September 1991 and October 1993. The *nef* sequence obtained from donor O in October 1991 displayed 97.62 ± 1.6% identity to that of cluster O (range, 98.71–93.55%), and that (AY363309) from Donor P in September 1991 displayed 97.34 ± 2.1% identity to that of cluster P (range, 99.19–91.35%). Four patients (HP4, HP15, HP17, and HP20) revealed insertion of nucleotides (24, 27, 3, and 3 base pairs, respectively) compared with donors O and P.

TABLE 4. FREQUENCY OF SIGNATURE PATTERN NUCLEOTIDES FROM PLASMA DONORS O AND P

| Position in HIV-1 NL4-3 | 8850 | 8975 | **9087** | 9113 | **9183** | **9243** | **9270** | **9331** |
|---|---|---|---|---|---|---|---|---|
| Signature (in Donor O) | $A_{1.00}$ | $A_{1.00}$ | **$G_{1.00}$** | | **$C_{1.00}$** | | | |
| Cluster-O ($n = 9$) | 1.00 | 0.88 | 0.88 | | 1.00 | | | |
| Background ($n = 131$) | 0.29 | 0.069 | 0.26 | | 0.41 | | | |
| NS change | | | **I113V** | | **V145T** | | | |
| Signature (in Donor P) | | | | $T_{1.00}$ | | **$G_{1.00}$** | **$A_{1.00}$** | **$A_{1.00}$** |
| Cluster-P ($n = 13$) | | | | 0.92 | | 1.00 | 0.77 | 0.77 |
| Background ($n = 127$) | | | | 0.08 | | 0.45 | 0.20 | 0.34 |
| NS change | | | | | | **I165V** | **C174S** | **V194E** |

Signature patterns in clusters O (Oct 1991) and P (Sep 1991 and Feb 1993) show sets of four and six nucleotides, respectively, that were the most common, but atypical in background sequences (a majority signature). $A_{1.00}$ at 8850 and $T_{1.00}$ at 9113 denote A and T at 4 of 4 and 6 of 6 sequences, respectively. Lines from cluster O and P depict the frequencies of these nucleotides among viral sequences of 8 and 12 people with hemophilia, respectively. Each four of eight nucleotide positions revealed significantly higher frequency in clusters O or P, compared with background ($p < 0.01$ for all). Five positions with nonsynonymous changes are presented in bold. Three sequences were excluded from this analysis.
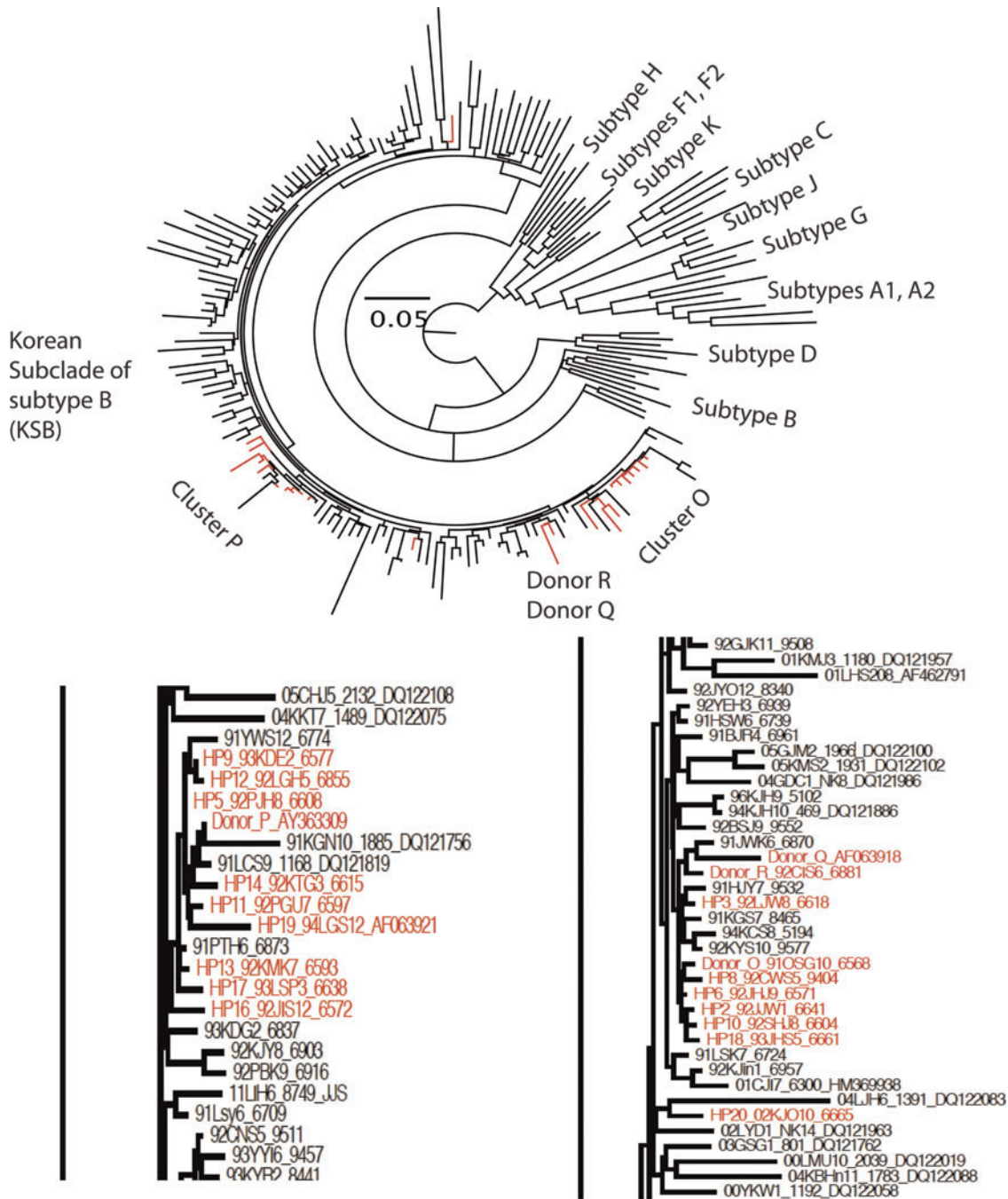
NS change, nonsynonymous change in amino acid.

**FIG. 1.** Molecular phylogenetic analysis using the maximum likelihood method. The evolutionary history was inferred using the maximum likelihood method based on the data-specific model. The tree with the highest log likelihood (−15860.8756) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial trees for the heuristic search were obtained automatically as follows. When the number of common sites was < 100 or < 1/4 of the total number of sites, the maximum parsimony method was used; otherwise, the BIO neighbor-joining method with maximum composite likelihood distance matrix was used. A discrete Gamma distribution was used to model evolutionary rate differences among sites (eight categories; +G, parameter = 0.4391). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 4.6206% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 190 nucleotide sequences from 20 HPs, plasma donors O, P, Q, and R, 119 LC patients infected with KSB (*n* = 143), and 46 reference strains. Codon positions included were 1st + 2nd + 3rd + noncoding. All positions containing gaps and missing data were eliminated. There were a total of 526 positions in the final dataset. Evolutionary analyses were conducted in MEGA5. The resulting tree file was rendered as a radial tree to show all taxa and as a phylogram to zoom in on the O and P subclusters, using FigTree. Two clusters, O and P, were broadly clustered with intervening of some LCs infected with KSB. The two digits after the number (1–20) of HPs represent the year of sampling. HPs, hemophiliacs; KSB, Korean subclade of human immunodeficiency virus type-1 subtype B; LCs, local controls.
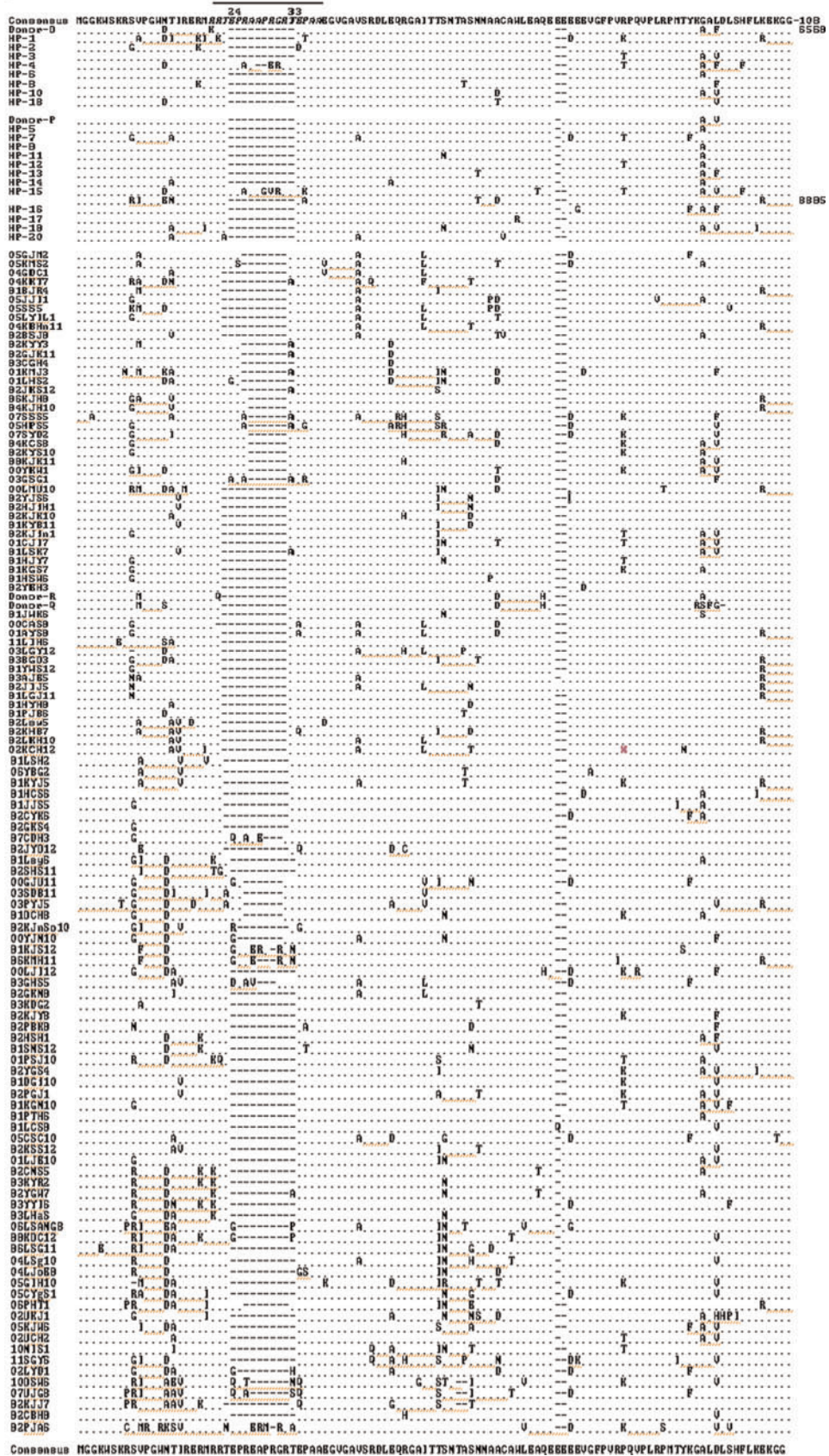
**FIG. 2.** Alignment of Nef protein sequences from 20 HPs, 4 plasma donors, and 116 LCs infected with KSB (LCs). Three sequences from three LC patients with premature stop codons were excluded. Significantly high frequencies of amino acids at four positions in cluster O and at four in cluster P were observed, compared with the earliest sequences in 126 and 130 LCs, respectively (*p* < 0.001 for all). The dot indicates the same sequence from Korean consensus sequences. The hyphen denotes deletions. Italic characters (*RRTEPRAAPRGRTEPAA*) indicate the variable region.
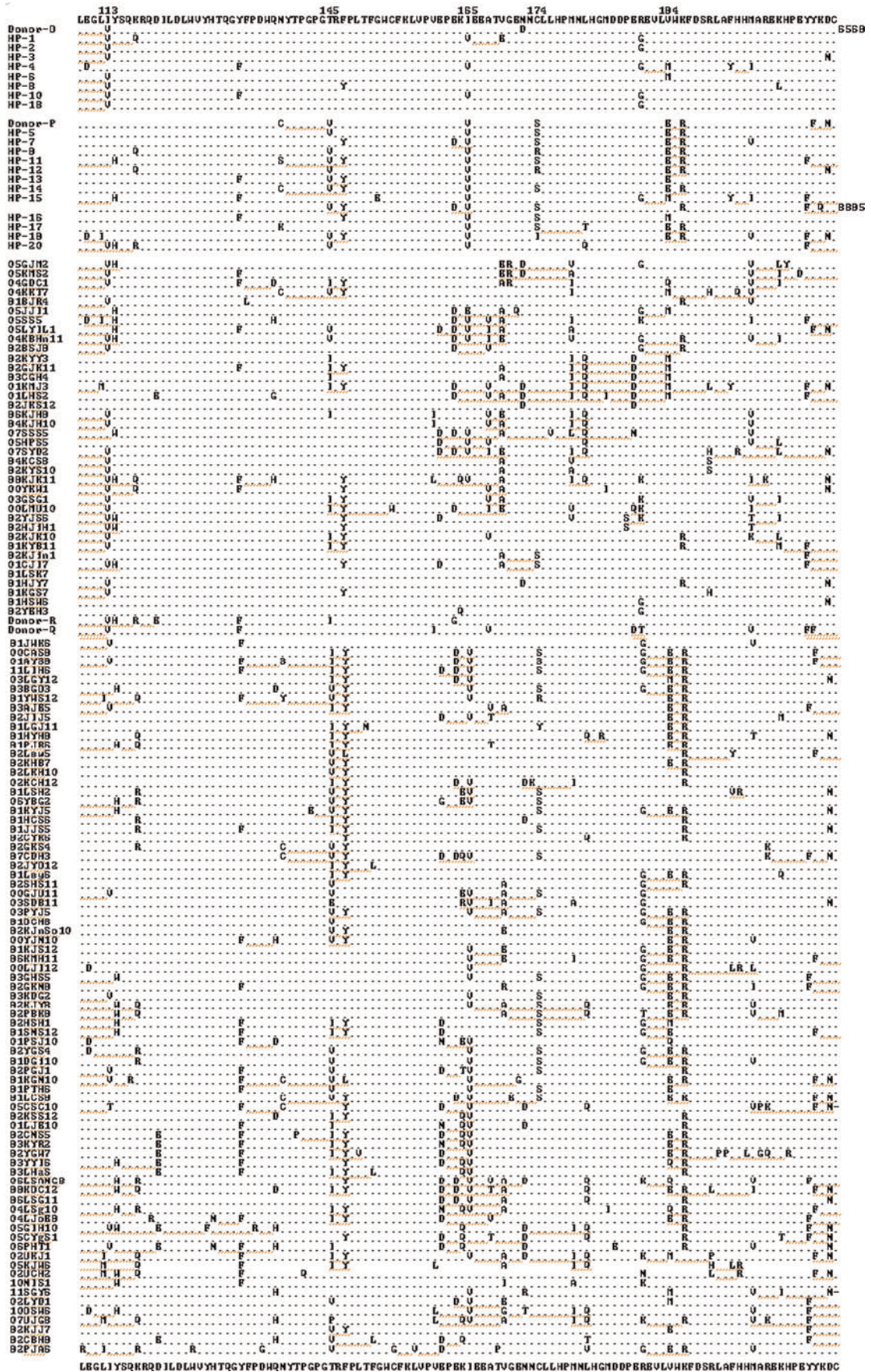
**FIG. 2.** (Continued).

## BLAST search

We searched in BLAST for the earliest sequences from donors O (AY363309) and P (AY363309), respectively. Sequences from six HPs (in order: HP8, HP6, HP2, HP18, HP3, and HP10) and four HPs (HP12, HP5, HP11, and HP9) most highly matched donors O and P, respectively.

## Correlation between sampling intervals and number of nucleotide differences

We determined the correlation between sampling intervals from the exposure (presumed January 1990) and number of nucleotide differences, compared with that in the corresponding plasma donor (HM237144 in October 1991 and AY363309 in September 1991). Three patients for whom the first sampling was performed in 2002 or 2003 displayed the highest difference of $>25$ nucleotides (range, 5–54; mean, 30), with a correlation coefficient, $\gamma$, of 0.74 ($p < 0.0001$; Fig. 3).

Intra-patient variation between the earliest and latest detection of *nef* genes in 22 patients (20 HPs plus 2 plasma donors) was $0.039 \pm 0.023$ over $90 \pm 57$ months prior to HAART. The correlation coefficient, $\gamma$, for intra-patient sequence divergence between the earliest and latest detection of *nef* genes in these patients was 0.588 ($p < 0.01$).

## Presence of premature stop codons

Three sequences from three LC patients infected with KSB (two in DQ122083-04LJH6, two in DQ122116-05Lsh5, and one in DQ122108-05CHJ5) revealed premature stop codons. These sequences are excluded from Figure 2.

## Comparison of sequence variation of the nef gene with vif and pol genes

We compared the sequence identity among three genes how much each gene varied from each donors. Sequence

**Table 5. Comparison of Sequence Identity with Donors O and P Within Each Cluster**

| Gene | Cluster O (n=8)[a] | Cluster P (n=12)[a] |
|------|------|------|
| *pol* | $99.07 \pm 0.47$[11] | $99.18 \pm 0.35$[11],* |
| *vif* | $98.94 \pm 0.53$[12] | $98.86 \pm 0.74$[12] |
| *nef* | $97.62 \pm 1.60$ | $97.34 \pm 2.10$* |

[a]Cluster based on *pol* and *vif* genes.
*$p < 0.01$ in cluster P only.
$p < 0.05$ between *pol* and *nef* genes in cluster O and between *vif* and *nef* genes in both clusters.

identity among patients was highest ($>99.07\%$) for the *pol* gene, followed by *vif* ($>98.80\%$). Despite the use of an earlier sample for *nef* compared with *pol* and *vif* genes (Table 1), the sequence identity of the *nef* gene between two donors and within each cluster was the lowest ($<98.0\%$; Table 5). There was no significant difference in sequence identity for the *pol* and *vif* genes among patients in clusters O and P. In contrast, a significant difference between *nef* and *pol* or *vif* genes was noted between clusters (Table 5).

## Evidence of the highly variable nef gene compared with pol and vif genes

In regard to insertions, HP4 and HP15 revealed similar insertions of 8 and 10 amino acids, respectively, compared with consensus sequences at the same variable region of the Nef protein (Fig. 2). Many sequences from HP4 revealed the same insertion of eight amino acids (24-EPAAERRT) although two sequences (AY221684 and FJ201810) revealed the wild-type sequence without insertion of 24 base pairs. All three sequences from HP15 revealed the same insertion of 10 amino acids (24-EPAAAGVRRT-33). In addition, donors O and P did not reveal variable sequences in the variable region. This may be because insertions may originate from the wild-type *nef* gene. In contrast, there were no insertions in the *pol* gene, and there was an amino acid insertion in the *vif* gene in one patient.

## Absence of discordant subtypes among paired pol, vif, and nef genes

In previous studies, we determined the *pol*[11] and *vif*[12] gene sequences in 109 and 169 patients, respectively, all of whom are included in the present study. No discordant subtypes were detected among the three genes.

## Discussion

In our present study, we amplified the *nef* gene in sera obtained earlier than the peripheral blood mononuclear cell samples used in previous studies,[9,13] thereby facilitating extension of the follow-up period from 1991. Within these KSB cases with highly limited variations, donors O and P formed two clusters designated O and P, respectively. Both O and P were cash-paid donors who sold plasma for the production of clotting factors, and HPs seroconverted within a few years after using the Korean-made factor. KSB forms a distinct monophyletic clade that is not related to any of the international sequences from Los Alamos Database described in December 2010.[11,12]
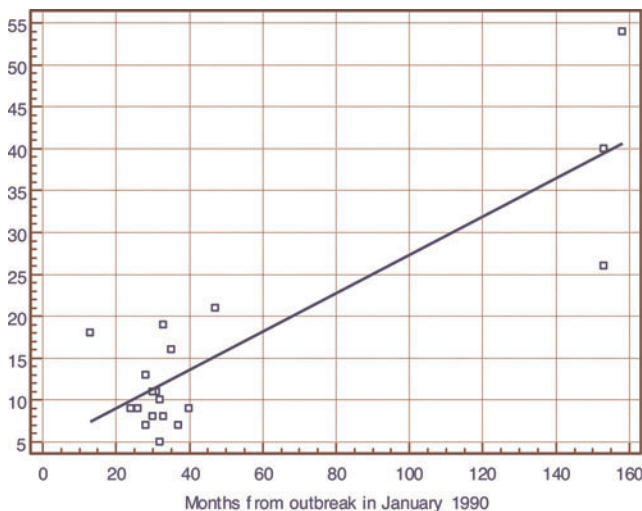


**FIG. 3.** Correlation between months from the exposure (around January 1990) to the sample being used and the number of nucleotide differences in the 20 HPs, compared with corresponding plasma donors. Three patients (HP4, HP15, and HP20) who provided serum samples after 2002 displayed the highest nucleotide differences with $\geq 25$ nucleotide changes. The correlation coefficient, $\gamma$, was 0.74 ($p < 0.0001$).

Owing to the higher degree of genetic conservation of *pol* and *vif* compared with other genes,[10] our previous studies on *pol* and *vif* revealed three and two nucleotide positions that were 100% conserved in clusters O and P, respectively, and 0% conserved in LCs. However, there was no such conserved nucleotide position in the *nef* gene. In other words, the significance of statistics in the frequency was less in the *nef* gene than in the *pol* and *vif* genes, although eight nucleotide positions revealed statistical significance in the frequency between clusters O or P and LCs. Despite the absence of a conserved nucleotide position, showing a contrast of 100% versus 0%, epidemiologically linked sequences were weakly clustered in the two groups (Fig. 1). In the four patients, sampling was performed after 2002, suggesting a long lag between primary infection and sampling. Therefore, significant divergence may have occurred.

Thus, in the phylogenetic analysis, the *pol* gene showed two strongly demarcated clusters (O and P) that included only 8 and 12 HPs, respectively. The *vif* gene also revealed two strong clusters with only 7 and 12 HPs, respectively. However, the *nef* gene analysis revealed some dispersion in the LCs included in the analysis. Thus, several sequences (HP4, HP6, HP7, HP15, and HP20) did not belong to clusters O or P (Fig. 1), differing from the distribution seen for the *pol* and *vif* genes.

To date, the discordant subtypes or subclades of HIV-1 in Korean patients have not been identified. Therefore, the separations in the phylogenetic tree, between cluster O and HP4 and between cluster P and HP7, HP15, and HP20, are natural due to the long time lag (>11 years except HP7) between primary infection and sampling (after 2002). In the same context, a genetic relationship between *gag* and *env* could not be established in another study in which specimens were collected >3 years after transmission.[16] Our data suggest that using a viral gene that undergoes rapid mutation may not be the best choice to use in epidemiological analyses, especially with samples that were collected from patients in whom viral infection occurred many years ago. Therefore, appropriate selection of the target viral gene is important in epidemiological studies for HIV-1 transmission evidence.

## Acknowledgments

## Author Disclosure Statement

The authors have no competing interests to disclose.

## References

1. Korber B, Myers G. Signature pattern analysis: a method for assessing viral sequence relatedness. AIDS Res Hum Retroviruses. 1992;8:1549–1560.
2. Leitner T, Escanilla D, Franzén C, et al. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci USA. 1996;93:10864–10869.
3. Metzker ML, Mindell DP, Liu XM, et al. Molecular evidence of HIV-1 transmission in a criminal case. Proc Natl Acad Sci USA. 2002;99:14292–14297.
4. Albert J, Wahlberg J, Leitner T, et al. Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 *pol* and *gag* genes. J Virol. 1994;68:5918–5924.
5. Blanchard A, Ferris S, Chamaret S, et al. Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients. J Virol. 1998;72:4537–4540.
6. Holmes EC, Zhang LQ, Simmonds P, et al. Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon. J Infect Dis. 1993;167:1411–1414.
7. Ou CY, Ciesielski CA, Myers G, et al. Molecular epidemiology of HIV transmission in a dental practice. Science. 1992;256:1165–1171.
8. 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;89:7365–7369.
9. Cho YK, Foley BT, Sung H, et al. Molecular epidemiologic study on HIV-1 outbreak in hemophiliacs B through clotting factor 9 after 1990. Vox Sanguinis. 2007;92:113–120.
10. Abecasis AB, Vandamme AM, Lemey P. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. J Virol. 2009;83:12917–12924.
11. 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;27:613–621.
12. Cho YK, Jung Y, Lee JS, et al. Molecular evidence of HIV-1 transmission in 20 Korean individuals with hemophilia; phylogenetic analysis of *vif* gene. Haemophilia. 2012;18:291–299.
13. Jee YM, Go U, Cheon F, 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.
14. 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 Hum Retrovirol. 1998;17:58–68.
15. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52:696–704.
16. 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;182:941–944.

Address correspondence to:
*Young-Keol Cho, MD, PhD*
*Department of Microbiology*
*University of Ulsan College of Medicine*
*88, Olympic-ro 43-gil, Songpa-gu*
*Seoul 138-736*
*Korea*

*E-mail:* ykcho2@amc.seoul.kr

*Brian T. Foley, PhD*
*HIV Databases*
*Theoretical Biology and Biophysics Group*
*Los Alamos National Laboratory*
*Los Alamos, NM 87545*

*E-mail:* btf@lanl.gov