Phylogenetic Analysis of Near Full-Length HIV Type 1 Genomic Sequences from 21 Korean Individuals

Young-Keol Cho,¹ Jung-Eun Kim,¹ and Brian T. Foley²

Abstract

The Korean subclade of subtype B (KSB) is the most prevalent HIV-1 strain found in Korea. To date, only two near full-length HIV-1 sequences from Korean patients have been reported. Here, we analyzed a total of 24 near full-length genomes of HIV-1 strains that were isolated from 17 antiretroviral therapy (ART)-naïve patients and four ART-exposed patients. Proviral DNA from peripheral blood mononuclear cells was PCR amplified and directly sequenced. Phylogenetic analyses were used to classify viruses from 19 patients as KSB, from one patient as subtype B, from one patient as subtype D, and three viruses from one patient as CRF02_AG. All KSB viruses demonstrated TAAAA instead of TATAA at the TATA box in the LTR. Of the 19 KSB patients, their sequence identities at the nucleotide level ranged from 89.8% to 97.1% from the lowest *env* gene to the highest *pol* gene. Other than the CRF02_AG viruses, no recombination events were noted in any of the 19 KSB patients, which is consistent with our previous studies on the *pol*, *vif*, and *nef* genes. Except for one strain, all of the strains were classified as non-syncytium-inducing strains. This is the first report to describe near full-length KSB.

SINCE 1985, 8,535 INDIVIDUALS IN KOREA have been diagnosed with HIV-1 infection, excluding foreign workers and U.S. army soldiers. In 2011, 888 patients were newly diagnosed. Our previous study reported that a total of 327 HIV-1-infected patients were diagnosed prior to 1994.¹ Of the 327 patients diagnosed prior to 1994, 155 patients (47.4%) were infected abroad, as overseas sailors ($n=116$), or as the wives of individuals who were infected overseas based on epidemiological data. Various subtypes and circulating recombinant forms (CRFs) of HIV-1 have been imported by overseas sailors and workers.¹ However, it seems that HIV-1-infected sailors were not involved in the significant spread of HIV-1 in Korea. Since 1992, domestic infections (excluding overseas sailors and their wives) have become a major portion (75%; 58 of 77) of newly acquired infections compared with 7% (1 of 14), 14% (3 of 22), 32% (12 of 37), 39% (21 of 54), and 46% (21 of 46) that were reported in 1985–1987, 1988, 1989, 1990, and 1991, respectively¹ (www.aidsinfo.or.kr/?r=home&c=1/9/55&uid=2975).

Among various subtypes and CRFs in Korea, domestically infected Korean individuals are mainly infected with the Korean subclade of subtype B (KSB) at the level of the *env*, *nef*, *pol*, and *vif* genes.^{1–6} KSB sequences were first detected in homosexual Korean HIV-1 patients in 1988.¹ Our previous

studies show that these KSB sequences represent a subclade of the globally represented subtype B, indicating a founder effect. All of these KSB sequences were found in domestically residing homosexual and heterosexual men and women who did not have sexual contact with foreigners.⁴

To date, only two near full-length KSB sequences have been reported by our team.^{7,8} Here, we analyzed 24 near full-length sequences from 21 HIV-1-infected Korean individuals (Table 1) and performed phylogenetic analyses.^{9–12} Sequence amplification of 11 overlapping fragments [each about 1 kilobase pair (kbp)] from peripheral blood mononuclear cell (PBMC) DNA using nested PCR, followed by direct sequencing of the products, was performed as previously described.^{7,8}

Sequences were aligned with the HIV-1 subtype reference set from the HIV Sequence Database (http://hiv-web.lanl.gov/content/hivdb/Subtype_REF/align.html) and phylogenetic trees were built using the PHYLIP DNAdist (F84 model, Ts:Tv 1.7) plus Neighbor, DAMBE maximum likelihood, and BEAST Bayesian maximum likelihood programs. Based on the results of the phylogenetic tree analyses, the viruses were classified as KSB ($n=19$), subtype B ($n=1$), subtype D ($n=1$), and CRF02_AG ($n=3$) (Fig. 1). Bootstrap support for the presence of KSB within subtype B was 100%.

¹Department of Microbiology, University of Ulsan College of Medicine, Seoul, Korea.

²HIV Databases, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico.

TABLE 1. CHARACTERISTICS OF FULL-LENGTH HIV-1 SEQUENCES FROM 21 KOREAN PATIENTS

Patient (code ^a)	Year of diagnosis	Date of collection	Subtype	CD4 ⁺ T cells/μl	Viral load (copy/ml)	Date of initiation for	
						HAART ^b	Korean red ginseng
10 patients reported in previous studies ^{9,10}							
CSR(87-05)	1987	Mar 2005	B	117	26,900	None	Jun 1994
KJS(89-17)	1989	Aug 2004	KSB	109	162,000	Feb 2002	Nov 1991
LSK(90-05)	1990	Aug 2004	KSB	225	4,500	None	Nov 1991
LSH(90-18)	1990	Jan 2003	KSB	501	116,000	None	Dec 1991
KJin(90-50)	1990	Aug 2004	KSB	332	104,000	None	Feb 1993
JWK(91-20)	1991	Jul 1997	KSB	701	5,825	None	Dec 1991
		Jul 2004		642	17,800		
CWS(92-13)	1992	May 2004	KSB	212	14,600	None	May 1993
KYR(93-04)	1993	Aug 2004	KSB	343	31,900	None	Mar 1993
KMH(93-60)	1993	May 2004	KSB	454	5,750	None	Nov 1996
YJN(96-51)	1996	Feb 2005	KSB	640	29,200	None	Jan 1997
7 patients reported in previous studies ^{8,10,11,14}							
HYH(89-26)	1989	Feb 1999	KSB	18	ND ^c	None	None
OSG(90-01)	1990	Jan 2002	KSB	216	97,800	Jan 1999	Jan 1992
HJY(91-22)	1991	Aug 2003	KSB	238	17,000	None	Aug 2002 ^d
KGS(91-23)	1991	May 2003	KSB	15	386,197	Jan 2001	Oct 1992
KYY(92-23)	1992	Apr 2007	KSB	921	8,680	None	Nov 1992 ^d
MHI(Pt. B)	1988	Oct 2004	CRF02_AG	574	821	None	Dec 1991
		Nov 2008		1154	49,800		
LHS	2001	Jun 2004	KSB	566	786,133	Feb 2001	None
2 patients reported in previous studies ^{12,15}							
KDE(HP-9)	1992	Nov 2003	KSB	502	11,500	None	Nov 2002 ^d
KMK(HP-13)	1992	May 2004	KSB	424	4,000	None	Aug 2000 ^d
2 patients not previously reported							
YGS(90-47)	1990	Mar 2003	KSB	578	698	None	Apr 1992
KBH(93-01) ^e	1993	Aug 2004	D	473	35,800	None	Apr 1993

^aThe number or expression in parentheses was used in the previous study.

^bHighly active antiretroviral therapy.

^cNot determined.

^dCompliance with Korean red ginseng (KRG) intake was defined as poor, or the duration of KRG intake was short.

^ePatient KBH had taken KRG from April 1993 to August 2004 (10,800 g of KRG supplied). Thereafter, follow-up was stopped. CD4⁺ T cell count increased from 199/ μ l to 473/ μ l for the period.

Clinical characteristics including changes in CD4 T cell count, RNA copy, Korean Red Ginseng (KRG) therapy, and frequent genetic defects in the *nef* and 5'-LTR/*gag* in the first 10 patients in Table 1 were described elsewhere.^{9,10} In brief, all patients were treated with KRG for a significant period. Consequently, they were defined as long-term slow progressors whose annual decrease of CD4 T cells was less than 20/ μ l.⁹ Clinical characteristics and frequent genetic defects in the 5'-LTR/*gag* genes in the second set of seven patients in Table 1 have been previously described.^{10,11} Two hemophilic patients (KDE and KMK) in Table 1 were described as HP-9 and HP-13, respectively.¹² The last two patients were not reported elsewhere. Interestingly, three patients (CSR, MHI, and KYY) have survived for 25, 24, and 20 years in the absence of highly active antiretroviral therapy (HAART), respectively, because they have been prescribed KRG for >10 years.

The average nucleotide sequence identity of the coding regions in the 19 KSB patients was 93.7%, and the nucleotide sequence identity between the earliest sequence (AF224507) of the 19 KSB patients and the 1984 reference strain HXB2 (K03455) was 92.5%. The average pairwise nucleotide sequence identity in the coding regions between KSB sequences

and the non-KSB subtype B sequences is 91.4%. The sequence identity at the nucleotide level between two patients who were husband (04KMH5) and wife (04KJS8) and between OSG and CWS who were donor and recipient were 95.1% and 93.2%, respectively. These epidemiologically linked cases provide the highest sequence identity of the 19 KSB patients who were investigated.

We also checked to determine if the KSB sequences recombine with other various subtypes using data provided by the Recombinant Identification Program of the Los Alamos National Laboratory (www.hiv.lanl.gov/cgi-bin/RIP3/RIP.cgi). No recombination events were associated with any of the KSB sequences. This finding is consistent with our previous studies on *vif*, *pol*, and *nef* genes.^{1,5,6} TATA box sequences in both LTR were classified as TATAA for subtype B and almost all other subtypes, and CRFs. However, TAAAA instead of TATAA was identified in CRF01_AE, some cpx strains, and CPZ.US.85.US_Marilyn.¹³ All of the KSB strains revealed TAAAA at the TATA box. This was found to be a common feature of all KSB strains, as previously reported.^{7,8,14,15}

Six patients were found to have a premature stop codon in at least one gene (Table 2). All of the premature stop codons

FIG. 1. Phylogenetic tree analysis of the near full-length genomes of 19 KSB, 1 subtype D (04KBH8), 1 subtype B (05CSR3), 3 CRF02_AG (MHI), and reference HIV-1 sequences. Patients 04WK7, 03KGS5, and 03HJY8, 04CWS5 and 02OSG1, and 04KMH5 and 04KJS8 were epidemiologically linked patients, respectively.

TABLE 2. DISTRIBUTION OF CHARACTERISTIC SEQUENCES IN 13 PATIENTS

Patient	Premature stop codon	Deletion	Insertion or duplication ^a	TATA box
KJS			9 bp in <i>gag</i>	TAAAA
LSK			9 bp in <i>gag</i> 1 bp in <i>pol</i>	TAAAA
LSH	<i>tat</i>	6 bp in <i>nef</i>	6 bp in 5' LTR 6 bp in <i>gag</i>	TAAAA
KJin		6 bp in <i>vpr</i>		TAAAA
WK	<i>pol</i> and <i>env</i>			TAAAA
CWS			9 bp in <i>env</i>	TAAAA
OSG	<i>rev</i>			TAAAA
KYY	<i>rev</i>			TAAAA
LHS		6 bp in <i>vif</i>		TAAAA
KMK			9 bp in <i>gag</i>	TAAAA
YGS	<i>pol</i>			TAAAA
KBH	<i>gag</i>	15 bp in <i>nef</i> 24 bp in both LTRs	3 bp in <i>gag</i>	TATAA

^aDescribed as an insertion or duplication, except for the findings shown in Fig. 2.

LTR, long terminal repeat.

were identified at tryptophan codons. Specifically, patient LSH was found to have a stop codon in the *tat* gene. Patients KYY and OSG were found to have a stop codon in the *rev* gene. Patients 97Jwk (AF224507; 97JWK position 4171 TGG trp → TAG stop) and YGS (DQ295195; 03YGS3 position 2766 TGG trp → TAG stop) were found to have stop codons in the *pol* gene in 1997 and 2003, respectively.⁷ Using genotyping for the *env* gene C2/V3 region, patient KMK was predicted to carry a syncytium-inducing (SI) strain. His CD4⁺ T cell count was 424 cells/ μ l in May 2005. Subsequently, his CD4⁺ T cell count rapidly decreased to <100 cells/ μ l within 2 years. Viruses from all other patients, except this case, were predicted to be non-SI strains.

Four patients received antiretroviral therapy (e.g., ART, HAART) prior to this study. Patients KJS and OSG took HAART for a month in 2002 and 1999, respectively. Three patients were found to carry drug-resistant mutations. Specifically, patient MHI infected with a CRF02_AG virus was found to carry the nonnucleoside reverse transcriptase inhibitor (NNRTI) resistance mutation N348I in most sequences, although he was not receiving ART. Patient LHS had taken indinavir, ddI, and 3TC since April 2001 following the diagnosis of AIDS, and has since been found to carry many resistance mutations.⁸ Patient YGS had never received ART; he was

found to carry NNRTI-V108I and P236L by G-to-A hypermutations (AY347691), but not in the complete genome sequence (JQ316135).

Several patients were found to have novel insertions, duplications, or deletions (Table 3). However, it is not necessary for us to describe all of the variations. For example, patient LHS (AY839827) was found to have a novel 14-bp insertion in the 5'-LTR region (between base pairs 581 and 582 in HXB2) in June 2004 (Fig. 2A). The same insertion had already been detected in February 2001 (EF370323 and EF370325). Regarding the *pol* gene, patient KJS (JQ316130) was found to have a 24-bp insertion (between base pairs 2161 and 2162) in two amplicons in August 2004 (Fig. 2B). Thereafter, the same insertions were detected in consecutive samples. However, an earlier sample (HQ026504; diagnosed in 1992), and a sample from his spouse KMH (DQ295193), did not reveal the same insertions.

Patient LSK (DQ295192) was found to have a novel 9-bp insertion (between base pairs 5829 and 5830; Fig. 2C), in addition to the insertion of nucleotide A into the *pol* gene. Patient LSH was also found to have a 24-bp insertion (between base pairs 6284 and 6285) in the initial part of the *env* gene (Fig. 2D) and a 6-bp deletion from the *nef* gene. Regarding the *gag* gene, patients KMK (JQ316126), LSK (DQ295192), and KJS (JQ316130) were found to have 9-bp insertions (between base pairs 1167 and 1168) in many samples. Patient CWS was also found to have a 9-bp insertion in the initial part of the V3 loop (6992-CAATAAAAAA-7000), and a 3-bp deletion was found in the terminus of the V3 loop. Patients CSR and KBH were found to have 9-bp deletions (amino acid at position 9–11) after the myristoylation site and 15-bp deletions from the *nef* gene, respectively. In particular, we obtained seven amplicons of the *nef* gene from the earliest sample from patient CSR, which was obtained in April 1991. Three, two, and two amplicons were classified as wild type or 6- or 9-bp deletions, respectively. Most sequences contained 9-bp deletions, indicating that the deletions might have been associated with slowly progressive strains, especially considering the lack of ART for 25 years since diagnosis in 1987.

This is the first study to describe several complete KSB genomes, which is the dominant epidemiological strain of HIV-1 that has spread in Korea since 1992.

Sequence Data

The GenBank accession numbers are AF224507, AY839827, DQ054367, DQ295192-96, DQ837381, JQ316126-38, JQ341411, and JQ429433.

Acknowledgments

This work was supported by a grant from the Korean Society of Ginseng, which was funded by the Korea Ginseng Corporation (2011–2012).

The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Disclosure Statement

No competing financial interests exist.

TABLE 3. AVERAGE PAIRWISE NUCLEOTIDE SEQUENCE IDENTITIES OF EACH GENE INCLUDED IN 19 KOREAN SUBCLADE OF SUBTYPE B SEQUENCES

Gene									
	<i>gag</i>	<i>pol</i>	<i>vif</i>	<i>vpr</i>	<i>tat</i>	<i>rev</i>	<i>vpu</i>	<i>env</i>	<i>nef</i>
Identity (%)	95.2	97.1	96.2	95.2	94.5	92.5	91.5	89.8	95.6

References

1. Cho YK, Jung Y, and 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.
2. Kang MR, Cho YK, Chun JS, *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.
3. Kim YB and Cho YK: Monophyletic clade of HIV-1 subtype B in Korea: Evolutionary pressure or single introduction? *AIDS Res Hum Retroviruses* 2003;19:619–623.
4. Cho YK, Foley BT, Sung H, Kim YB, and Kim JH: Molecular epidemiologic study on HIV-1 outbreak in hemophiliacs B through clotting factor 9 after 1990. *Vox Sanguinis* 2007;92: 113–120.
5. Cho YK, Jung Y, Lee JS, and Foley BT: Molecular evidence of HIV-1 transmission in 20 Korean individuals with hemophilia; phylogenetic analysis of *vif* gene. *Haemophilia* 2012; 18:291–299.
6. Cho YK, Kim JE, and Foley BT: Phylogenetic analysis of the earliest *nef* gene from hemophiliacs and local controls in Korea. *BioResearch Open Access* 2012;1:41–49.
7. Cho YK, Lee HJ, and Desrosiers RC: Complete sequences of HIV-1 in a Korean long-term nonprogressor with HIV-1 infection. *J Korean Soc Virol* 1999;29:107–118.
8. Cho YK, Sung H, Bae IG, *et al.*: Full sequence of HIV type 1 Korean subtype B in an AIDS case with atypical seroconversion: TAAAAA at TATA box. *AIDS Res Hum Retroviruses* 2005;21:961–964.
9. Cho YK, Lim JY, Jung YS, Oh SK, Lee HJ, and Sung H: High frequency of grossly deleted *nef* genes in HIV-1 infected long-term slow progressors treated with Korean red ginseng. *Curr HIV Res* 2006;4: 447–457.
10. Cho YK and Jung YS: High frequency of gross deletion in 5' LTR/gag regions in HIV-1 infected long-term survivors treated with Korean red ginseng. *AIDS Res Hum Retroviruses* 2008;24:181–193.
11. Cho YK, Jung YS, Sung H, Sim MK, and Kim YK: High frequency of gross deletions in 5' LTR/gag and *nef* genes in patients infected with CRF02_AG of HIV-1 who survived for over 20 years: An association with Korean red ginseng. *AIDS Res Hum Retroviruses* 2009;25:535–541.
12. Cho YK, Jung YS, and Sung H: Frequent gross deletion in the HIV-1 *nef* gene in hemophiliacs treated with Korean red ginseng; inhibition of detection by highly active antiretroviral therapy. *AIDS Res Hum Retroviruses* 2009;25: 419–424.
13. Oelrichs RB, Workman C, Laukkanen T, McCutchan FE, and Deacon NJ: A novel subtype A/G/J recombinant full-length HIV type 1 genome from Burkina Faso. *AIDS Res Hum Retroviruses* 1998;14:1495–1500.
14. Cho YK and Jung YS: Dosage and duration effects of Korean red ginseng intake on frequency of gross deletions in the *nef* gene. *J Ginseng Res* 2010;34:219–225.
15. Cho YK, Jung Y, Sung H, and Joo CH: Frequent genetic defects in the HIV-1 5' LTR/gag gene in hemophiliacs treated with Korean red ginseng: Decreased detection of genetic defects by highly active antiretroviral therapy. *J Ginseng Res* 2011;35:413–420.

Address correspondence to:

Young-Keol Cho

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