Journal of
HIV/AIDS & Infectious Diseases

JScholar
Fostering Scholarly Communication

Research Article Open Access

Evolutional History of HIV-1 in Korea: Sequence Analysis of env gene in HIV-1 Korean B Subtype from 2006 to 2011

Sangmi Ryou[1], Yoon-Seok Chung[2], Myeongsu Yoo[3], Kisoon Kim[4,5], Sangsoo Kim[6] and Mee-Kyung Kee[4,7]*

[1]Division of Clinical Research, Center for Emerging Virus Research, Korea National Institute of Health, Korea Disease Control and Prevention Agency, Chungbuk, Republic of Korea
[2]Honam Regional Center for Disease Control and Prevention, Regional Centers for Disease Control and Prevention, Korea Disease Control and Prevention Agency, Gwangju, Republic of Korea
[3]Division of Public Health Emergency Response Research, Director General for Public Health Emergency Preparedness, Korea Disease Control and Prevention Agency, Chungbuk, Republic of Korea
[4]Division of Viral Disease Research, Center for Infectious Diseases Research, Korea National Institute of Health, Korea Centers for Disease Control and Prevention, Cheongju, Republic of Korea
[5]Department of Microbiology, Korea University Collegar of Medicine, Seoul, Republic of Korea
[6]Department of Bioinformatics and Life Science, Soongsil University, Seoul, Republic of Korea
[7]International Tuberculosis Research Center, Seoul, Republic of Korea

***Corresponding authors:** Mee-Kyung Kee, Division of Viral Disease Research, Center for Infectious Diseases Research, Korea National Institute of Health, Korea Centers for Disease Control and Prevention, International Tuberculosis Research Center, Seoul, Republic of Korea, Tel: +82-10-3127-1368, Fax: +82-50-4341-1368, E-mail: keemeekyung@gmail.com

Received Date: August 07, 2021 **Accepted Date:** September 07, 20211 **Published Date:** September 09, 2021

Citation: Sangmi Ryou (2021) Evolutional History of HIV-1 in Korea: Sequence Analysis of env gene in HIV-1 Korean B Subtype from 2006 to 2011. J HIV AIDS Infect Dis 8: 1-15.

Abstract

A previous molecular epidemiological study (1990–2005) reported that the transmission of Korean subtype B (Korean B), the human immunodeficiency virus (HIV)-1 variant, was predominant in Korea. We investigated the HIV-1 subtype diversity and evolutionary patterns of Korean B from 2006–2011.

Sequences of the *env* variable C2V3 region and epidemiological data were obtained from 392 newly diagnosed HIV-1-positive cases from 2006–2011. HIV-1 subtypes were determined, and genetic distances of sequence pairs were calculated by the Kimura two-parameter model. A phylogenetic tree was constructed using the maximum likelihood method. HIV-1 co-receptor usage was inferred by analyzing the V3 nucleotide sequence.The evolutionary rate was estimated by Bayesian inference and maximum clade credibility trees.

In Korea, subtype B (91.8%) was the most prevalent, and classified as Korean B (86.7%) and global subtype B (global B) (13.3%). The mean diversity of Korean B (0.097 ± 0.009) was lower than that of global B (0.127 ± 0.009) and subtype G (0.157

± 0.013) (P < 0.01). GPGS (41.2%) was the predominant motif in the V3 loop in Korean B, global B (24.4%), and other subtypes (10.0%) (P < 0.001). CCR5 was more commonly used by HIV-1 than CXCR4 (95.5% vs. 5.0%) (P < 0.001). The estimated evolutionary rates of Korean B and the global B were 4.29×10^{-3} (95% highest posterior density (HPD): 3.10×10^{-3} – 5.49×10^{-3}) and 7.46×10^{-4} (95% HPD: 1.39×10^{-7} – 2.42×10^{-3}) substitutions/site/year, respectively (P < 0.0001).

Korean B accounts for most nationwide cases of HIV-1 in Korea. The present results provide novel insights to further the current understanding of the characteristics and dynamics of the HIV-1 epidemic in Korea. Continued research is needed to monitor the spread of the virus in Korea.

Keywords: env gene in HIV-1; HIV

List of abbreviations: AIDS: Acquired Immune Deficiency Syndrome; HIV: Human Immunodeficiency Virus; HTS: Heterosexual; MSM: Men who have Sex with Men; CCR5: C-C chemokine receptor type 5; CXCR4: C-X-C chemokine receptor type 4

Introduction

The cumulative number of human immunodeficiency virus (HIV) cases in Korea since the first reported case of HIV infection in 1985 was reportedly 18,724 in 2019. The sex ratio was 9.9:1, and the proportion of foreigners was approximately 12.1%. Of all the cases, HIV infection through sexual contact accounted for > 99%. Although cases of HIV infection in Korea are relatively low compared to those in other countries, the number of newly diagnosed HIV cases is increasing annually. Furthermore, the number of HIV-infected young men has been increasing yearly since 2011 [1].

Previous studies on *env* gene sequences of HIV-1 revealed a unique strain of HIV-1 subtype B known as Korean clade B subtype (Korean subtype B; Korean B). Korean B accounted for over 80% of subtype B infection cases in Korea. A most recent common ancestor (MRCA) analysis revealed that HIV-1 subtype B originally emerged in 1961 in the United States, and Korean B evolved in 1967, and its prevalence gradually increased until the mid-1990s. The evolutionary rate of Korean B was estimated to be approximately 3–5-fold less than that of non-Korean B (global subtype B; global B) isolates [2-5].

Envelope glycoprotein gp120 contains constants (C1-C5) and variable (V1–V5) regions. Several amino acid residues in the V3 loop (amino acid residues 31-39) are highly conserved among HIV-1 variants owing to their functional importance. GPGQ is the most common tetrameric tip motif in the V3 loop among all HIV-1 subtypes, and GPGR is the predominant motif in subtype B in America and Africa. The substitution of proline (P) for tryptophan (W), creating the GWGR motif, has been reported in Brazil. In Korea, however, an unusual tetrameric tip motif, GPGS, is present in a slight majority of sequences [4,6]. The V3 loop of HIV-1 gp120 plays a key role in viral entry into target cells and influences tropism and biological phenotype. Furthermore, molecular recognition of chemokine receptors is predominantly mediated through the V3 loop fragment. Upon V3 loop-co-receptor interaction (CCR5 or CXCR4 or both), a series of rearrangements occur in the envelope glycoproteins, leading to the fusion of the host cell membrane and viral envelope [7-10].

Analysis of the *env* C2V3 sequence is useful for subtype determination and phylogenetic analysis, and is important for evaluating the regional circulation of the virus and developing effective strategies to prevent epidemics. Moreover, temporal monitoring of changes in the frequency of certain variants within different risk groups would facilitate the characterization of transmission clusters and networks to examine the diversity of HIV-1 [11,12].

In Korea, the previous study estimated the origin and evolution charecteristics for Korean HIV-1 subtype B using bayesian phylogenetic analysis. Those results suggested that the growth rate of prevalent HIV-1 strains in Korea was lower than in other countries and the evolution of HIV Korean clade B was relatevely slow [2]. Although genotype testing was routinely performed for the diagnosis of HIV-1 infection in Korea, few studies have focused on Korean B. In this study, to investigate the evolutionary history of HIV-1 subtype B in Korea, we successfully performed epidemiological and phylogenetic analyses based on env C2V3 sequences from HIV-1 infected Koreans over the past 6 years (2006-2011) and analyzed the HIV-1 subtype diversity and evolutionary patterns of Korean B.

Materials and methods

Study population

If HIV infections were confirmed in blood samples upon screening, blood samples of those individuals from the screening sites were sent to the Korea Center for Disease Control and Prevention (KCDC) [15]. KCDC collected blood samples and associated epidemiological data (date of diagnosis, date of birth, transmission route, the reason for HIV testing, marital status, HIV screening institute, CD4+ cell counts, and viral load) from newly diagnosed HIV-1-positive individuals. For the annual representation of newly diagnosed cases of HIV infection, the stratified sampling method was used to select approximately 10% (476 subjects) of 4,715 new Korean individuals on the basis of their demographic characteristics (sex and age) from 2006 to 2011. The sampling method has been changed for new cases of HIV seroconversion since 2012. A total of 392 cases with sequences from 476 blood samples were obtained. From the present epidemiological data, the transmission route was classified into three groups in accordance with self-reported risk factors: the heterosexual (HTS), men who have sex with men (MSM), and unknown. Bisexual individuals were included in the MSM group.

env amplification and sequencing

DNA was extracted, and the 1.2-kb *env* V1–V5 region (HXB2: nt 6556–7801) was amplified primer set (ed3/ed14, ed5/ed12) and sequenced with primer ed5 (6556–6581), ed12 (7822–7792), ed31 (6816–6844), ed33 (7359–7380), r25 (6873–6857), and v3f (7314–7334) as previously reported [5,16]. Direct sequencing was performed using the ABI Prism Big-Dye Terminator Cycle Sequencing 3.1 Ready Reaction Kit (Applied Biosystems, Foster City, CA, USA) with an automated sequencer (ABI Prism 3730 DNA sequencer; Applied Biosystems). V1V5 region sequence data did not cover the entire *env* V1V5 regions (it was a short read sequence of about 7 amino acids); hence, we used *env* C2V3 regions (HXB2: nt 6975-7520). In total, 392 *env* C2V3 sequences were used for further analysis.

Subtyping and data analysis

The HIV-1 subtype (n = 392) was determined by the Los Alamos database HIV BLAST search (http://www.hiv.lanl.gov) and by using REGA (HIV Bioinformatic Bioafrica). Genetic distances of pairs of *env* C2V3 sequences were calculated using the Kimura two-parameter model with MEGA v.6.0. A phylogenetic tree was constructed by the Maximum Likelihood method with 1000 bootstrap replicates. In this analysis, we used 384 sequences (8 sequences were excluded: 7 male, 1 female; 3 HTS, 4 MSM, 1 unknown; 1 Korean B, 6 global B, 1 subtype G) and 56 similar references which accounted for the HIV-1 subtype strains present in Korea. The most similar reference sequences were obtained from HIV BLAST in the Los Alamos database (Supplementary Material).

Characterization of the V3 loop

HIV-1 co-receptor usage was inferred from the V3 nucleotide sequence using Geno2Pheno (http://www.geno2pheno.org). Sequences with a false-positive rate < 10% were considered dual/mixed (X4/DM) tropic. In total, 379 samples were included in this analysis except for 13 sequences (12 male, 1 female; 7 HTS, 4 MSM, 2 unknown; 4 Korean B, 7 global B, 2 other subtypes), for which tropism could not be established.

Molecular clock signal analysis

We cleaned clock-likeness of dataset (392 isolates *env* C2V3 sequences and 140 downloaded from the Los Alamos HIV sequence database for references) by performing linear regression analysis between the parameters 'root-to-tip divergence' and 'sampling data' to maximize the correlation coefficient of R^2 using TempEst v.1.5.3 [17]. One hundred fourteen sequences (fifty-three *env* C2V3 sequences and sixty-one references sequences) were considered outliers in the TempEst analysis and were excluded from further analysis.

Evolutionary analysis

To better understand the evolutionary dynamics of Korean B, the C2V3 sequences dataset (418 *env* C2V3 sequences, including 339 Korean isolates and 79 reference sequences) used to estimate the nucleotide substitution rate using Bayesian Markov chain Monet Carlo (MCMC) sampling implemented in the BEAST 1.10.4 [18] (Supplementary Material). As the substitution models, we selected independently best-fit models for this analysis using ModelFinder of IQ-TREE v.1.6.12 . [19] The best model was chosen as that with the lowest with the lowest Akaike's information criterion (AIC) through MCMC (AICM). Using the lowest AIC model, Bayesian Evolutionary Analysis Utility (BEAUTi)-generated XML file was then imported into BEAST. We conducted independent runs to 300 million generation (1×10^7, 2×10^7 and 3×10^8 chain lengths) combinations under 4 clock and 6 tree models. To select the best combination of the molecular clock and tree prior, both path sampling and stepping-stone sampling were employed [20,21]. Effective sample sizes (ESS)

from log files by multiple BEAST runs were assessed using Tracer version 1.7.1[22] where the effective sample size (ESS) was no less than 200. The Bayesian maximum clade credibility (MCC) tree was generated using TreeAnnotator v1.8.4, and then visualized using FigTree v1.4.32. To infer past population dynamics, we applied the best-fit model, a Uncorrelated relaxed - Bayesian skyline population (BSP) model [18] about 296 Korean B and 40 variants of global B (CH, Switzerland; CN, China; ES, Spain; FR, France; JP, Japan; NL, Netherlands; UK, United Kingdom; US, United States) in Korea, which was inferred by BEAST 1.10.4 and Tracer 1.7.1. The BSP method followed the MCMC procedure to estimate the distribution of generalized skyline plots and generate a posterior distribution of effective population size through time (or the effective number of infections in the case of viral epidemics) from the collected plots. The results of the Bayesian MCMC were used to calculate a marginal posterior distribution of the demographic inference, and estimated parameters included the effective population size at the most recent sampling, *Ne* (the effective number of prevalent infections).

Statistical analysis

Graphs were generated using Prism 5 software (GraphPad, La Jolla, CA, USA). Significant differences were discerned through t-test, one-way analysis of variance (ANOVA) and Tukey's multiple comparisons post-hoc test. Statistical analysis was performed using SAS 9.4. Chi-square test and Fischer exact (Freeman-Holton) test were used to examine differences for each variable. Statistical significance was set at 0.05 for all cases.

Results

Population characteristics and Subtype classification

We collected DNA sequences from the *env* C2V3 region of 392 individuals, newly diagnosed as HIV-1-positive between 2006 and 2011. Table 1 shows the epidemiological and subtype characteristics of the 392 populations during these six years. The samples were classified in accordance with sex, age, transmission route, CD4+ cell counts, and HIV-1 subtypes. Most of the indi-

Table 1: Epidemiological and subtype characteristics of 392 new cases of HIV infection in Korea (2006-2011)

Category	Total		2006		2007		2008		2009		2010		2011	
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Total	392		70		54		64		74		80		50	
Sex														
Male	366	(93.4)	65	(92.9)	52	(96.3)	60	(93.8)	71	(95.9)	72	(90.0)	46	(92.0)
Female	26	(6.6)	5	(7.1)	2	(3.7)	4	(6.3)	3	(4.1)	8	(10.0)	4	(8.0)
Age (years)														
<30	93	(23.7)	14	(20.0)	10	(18.5)	16	(25.0)	16	(21.6)	24	(30.9)	13	(26.0)
30-49	215	(54.8)	47	(67.1)	30	(55.6)	36	(56.2)	45	(60.8)	35	(43.2)	22	(44.0)
≥50	84	(21.4)	9	(12.9)	14	(25.9)	12	(18.8)	13	(17.6)	21	(25.9)	15	(30.0)
Transmission route*														
HTS	179	(45.7)	26	(37.1)	24	(44.4)	25	(39.1)	40	(54.1)	44	(55.0)	20	(40.0)
MSM	159	(40.6)	34	(48.6)	20	(37.0)	22	(34.4)	27	(36.5)	32	(40.0)	24	(48.0)
Unknown	54	(13.8)	10	(14.3)	10	(18.5)	17	(26.6)	7	(9.5)	4	(5.0)	6	(12.0)
CD4+ cell counts (cell/mm³)*														
<200	74	(23.6)	16	(23.2)	9	(17.3)	6	(9.7)	-		21	(26.3)	22	(44.0)
200-349	100	(31.9)	19	(27.5)	21	(40.4)	27	(43.5)	-		24	(30.0)	9	(18.0)
350-499	79	(25.2)	22	(31.9)	13	(25.0)	17	(27.4)	-		18	(22.5)	9	(18.0)
≥500	60	(19.2)	12	(17.4)	9	(17.3)	12	(19.4)	-		17	(21.3)	10	(20.0)
HIV-1 subtypes														
B	360	(91.8)	66	(94.3)	49	(90.7)	60	(93.8)	67	(90.5)	71	(88.8)	47	(94.0)
CRF01_AE	16	(4.1)	2	(2.9)	3	(5.6)	3	(4.7)	3	(4.1)	4	(5.0)	1	(2.0)
G	7	(1.8)	-		1	(1.9)	-		1	(1.4)	3	(3.8)	2	(4.0)
C	2	(0.5)	-		-		1	(1.6)	-		1	(1.3)	-	
CRF02_AG	2	(0.5)	1	(1.4)	-		-		-		1	(1.3)	-	
A	2	(0.5)	1	(1.4)	-		-		1	(1.4)	-		-	
D	1	(0.3)	-		1	(1.9)	-		-		-		-	
CRF07_BC	2	(0.5)	-		-		-		2	(2.7)	-		-	

*HTS, heterosexual; MSM, men who have sex with men; Unknown, refusal to answer self-report question.

*CD4+ cell counts; 79 populations' data was missed

viduals in the population were males (n = 366 [93.4%]). Regarding the transmission route, the proportion of MSM cases was 40.6%. Furthermore, as a major immunological indicator of the stage of HIV-1 disease progression, CD4+ cell counts were < 200 cell/mm^3 (n = 74 [23.6%]), 200–349 cell/mm^3 (n = 100 [31.9%]), 350–499 cell/mm^3 (n = 79 [25.2%]), and ≥ 500 cell/mm^3 (n = 60 [19.2%]), respectively. Subtype B was detected in 91.8% of the total studied population; the remaining 8.2% accounted for subtype CRF01_AE (4.1%), G (1.8%), and other subtypes (2.3%).

Table 2 shows the characteristics of HIV-1 subtype in 392 HIV-infected individuals classified in accordance with sex and transmission route from 2006 through 2011. Subtype B was divided into Korean B, and global B. Korean B (79.6%) was predominant in our study population and accounted for 86.7% (312/360) of subtype B. Males with Korean B were significantly greater than females with Korean B (81.1% vs. 57.7%, P < 0.001), and the Korean B were significantly more prevalent in the MSM group (86.2%) than in the HTS group (76.8%) (P < 0.001).

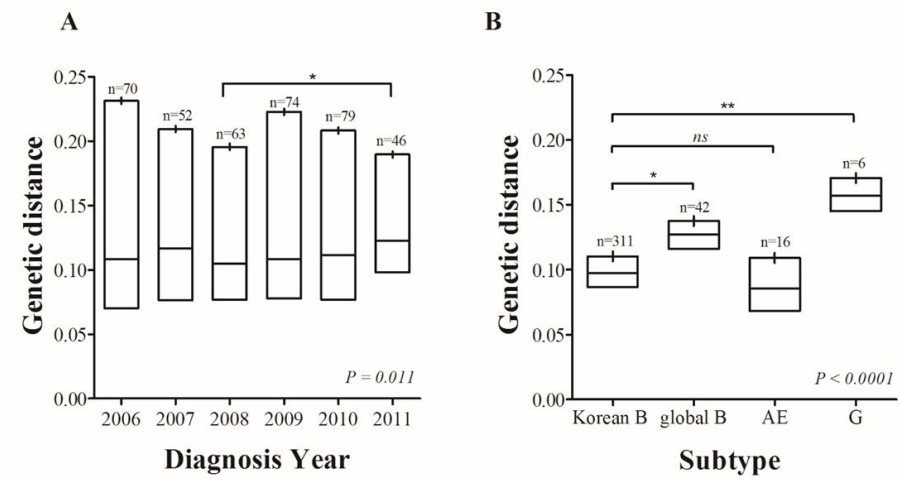
Sequence analysis

Table 2: Subtype characteristics according to Sex and HIV transmission route among 392 HIV-infected person from 2006 to 2011 in Korea

Category	Total	Subtype B				Other Subtypes		P-value
		Korean B		Global B				
		N	(%)	N	(%)	N	(%)	
Total	392	312	(79.6)	48	(12.2)	32	(8.2)	
Sex								<0.001
Male	366	297	(81.1)	45	(12.3)	24	(6.6)	
Female	26	15	(57.7)	3	(11.5)	8	(30.8)	
Male								<0.001
HTS[2]	155	119	(76.8)	16	(10.3)	20	(12.9)	
MSM	159	137	(86.2)	21	(13.2)	1	(0.6)	
Unknown	52	41	(78.8)	8	(15.4)	3	(5.8)	
Female								*
HTS	24	14	(58.3)	3	(12.5)	7	(29.2)	
Unknown	2	1	(50.0)	-		1	(50.0)	

Korean B, Korean subtype B; Global B, global subtype B. HTS, heterosexual; MSM, men who have sex with men; Unknown, refusal to answer self-report question.

*Expected count less than 5. Chi-squared approximation may be incorrect

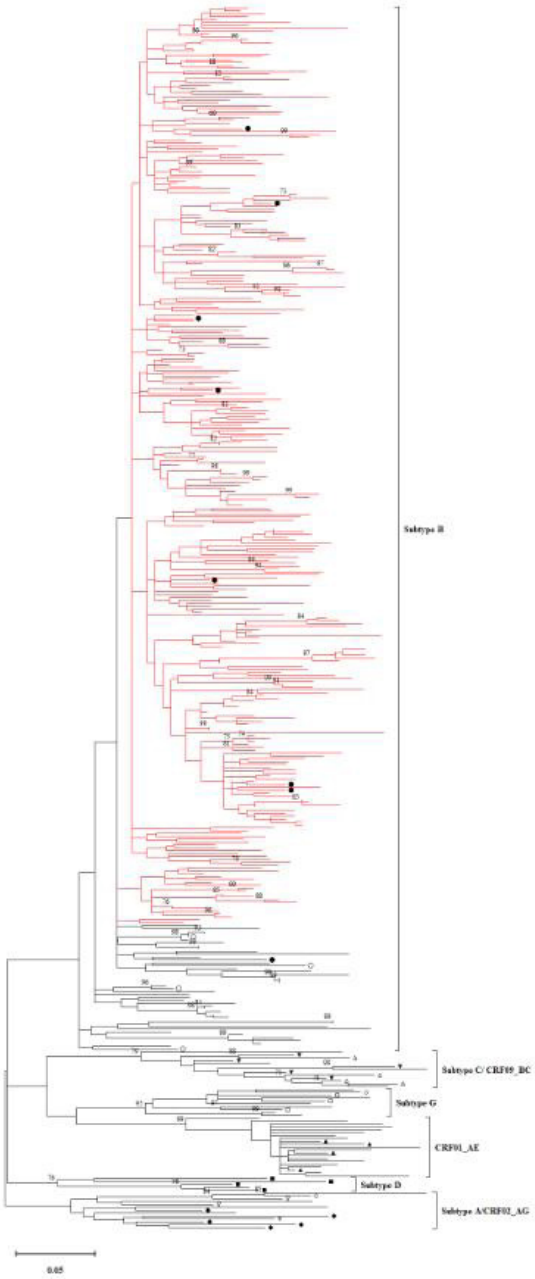


(A) Nucleotide diversity by the year of HIV diagnosis (n = 384). Statistics are presented as the mean (range; min to max). P-values were determined using ANOVA and Tukey's multiple comparisons host-hoc test. The asterisk indicates significantly different values. *P < 0.05. (B) Nucleotide diversity of Korean B and three HIV-1 subtypes. P-values were determined using ANOVA and Tukey's multiple comparisons host-hoc test. The asterisk indicates significant differences between two measured values in each column *P < 0.01; **P < 0.001; ns, non-significant. (KB, Korean B; B, global B; AE, CRF01_AE; G, subtype G)

Figure 1: Nucleotide diversity of HIV-1 variants in Korea, 2006–2011

Pairwise genetic distances were determined on the basis of *env* C2V3 sequences (n = 384) using the Kimura two-parameter model. In the overall population, genetic distances displayed an increasing trend in diagnosis year (P = 0.011) (Figure 1A). Based on the subtype, the genetic distance of Korean B was lower significantly than those of the other subtypes (P < 0.001 vs. global B and P < 0.01 vs. subtype G; Figure 1B). There was no significant difference in the genetic distance by sex (P > 0.05) and by transmission route (P = 0.131).

For 384 sequences and the 56 reference sequences, the result of the maximum likelihood analysis shows that three major taxa (subtype B cluster, subtype AE clusters, and other subtype clusters) were observed among the Korean isolates. This revealed that Korean B formed distinct clusters (70% bootstrap value) (Figure 2). However, no specific epidemiologic characteristic, including age, transmission routes, or diagnosis year, was associated with clustering in this analysis (data not shown).



The tree was constructed with the maximum likelihood method. Bootstrap values of 1000 replicates > 70% are marked at the cluster nodes. The final data set contained 440 sequences (n = 384 found in Korea and n = 56 GenBank reference sequences; see S1 Table). The Korean B clade shown as a red horizontal line. Reference sequences of each subtype are labeled with several symbols. Korean subtype B (●); global B (○); subtype D (▪); subtype G (□); CRF01_AE (▲); CRF09_BC (Δ); subtype C (▼); subtype A (∇); CRF02_AG (♦); CPX (Subtype AD, A2D, DG, GH, CU) (◊).

Figure 2: Phylogenetic tree of the *env* C2V3 gene in HIV-1 subtypes

Characteristics of the V3 loop

Viral subtype diversity and co-receptor tropism were analyzed based on *env* V3 sequences (n = 379) using Gene2Pheno. Overall, we identified 34 tetrameric motifs at the tip of the V3 loop; of these, GPGS (41.2%) was predominant among Korean B isolates (P < 0.001). Upon co-receptor usage prediction analysis, CCR5 and CXCR4 viruses of Korean B were 98.1% and 1.9%, respectively. CCR5 accounted for a significantly greater proportion in the Korean B than in the global B isolates (P < 0.001) (Table 3). Furthermore, CCR5 variants of Korean B contained the GPGS motif (41.4%; 125/302) (data not shown).

Table 3: Characteristics of *env*V3 loop peptide in 379 HIV-infected person from 2006 to 2011 in Korea

Category	Total (%)		V3 tip motif							Viral tropism				
			GPGS		GPGR		Other			CCR5		CXCR4		
	N		N	(%)	N	(%)	N	(%)	P-value	N	(%)	N	(%)	P-value
Total	379		139	(36.7)	126	(33.2)	114	(30.1)		360	(95.0)	19	(5.0)	
Subtype									<0.001					<0.001
Korean B	308	(81.3)	127	(41.2)	97	(31.5)	84	(27.3)		302	(98.1)	6	(1.9)	
Gobal B	41	(10.8)	10	(24.4)	22	(53.7)	9	(22.0)		34	(82.9)	7	(17.1)	
Others	30	(7.9)	3	(10.0)	6	(20.0)	21	(70.0)		26	(86.7)	4	(13.3)	

Korean B, Korea subtype B; global B, global subtype B; Others, other subtypes. C-C chemokine receptor type 5; CCR5, C-X-C chemokine receptor type 4; CXCR4

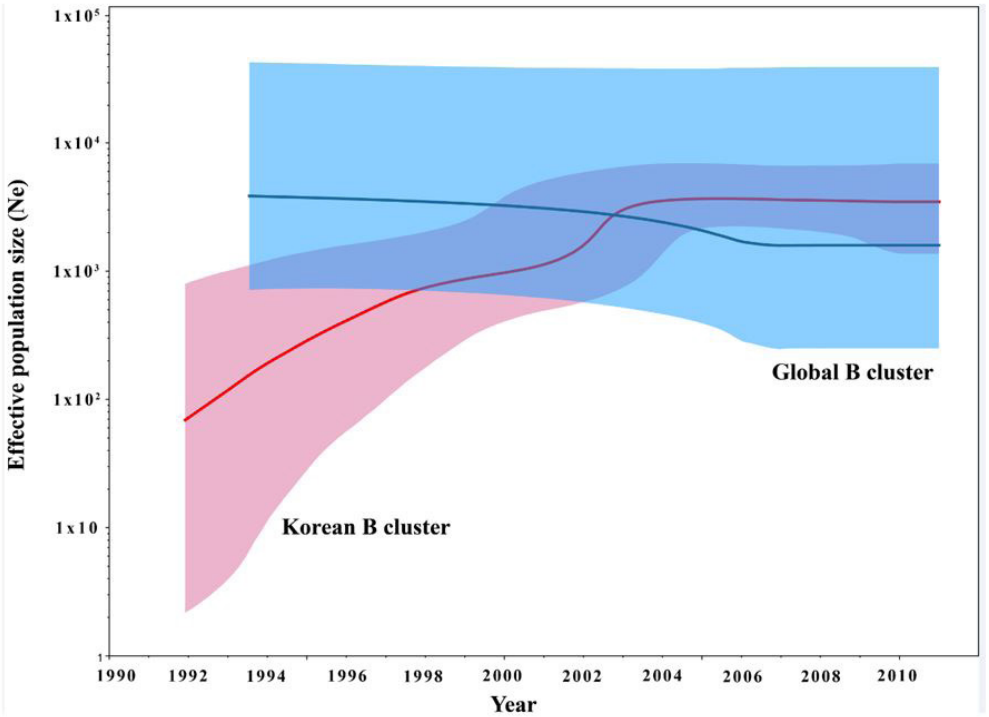
Evolutionary Dynamics of env C2V3

The Tempest analysis of molecular clock structure revealed a correlation in the C2V3 gene (R^2 = 0.160). Residual mean squared values was 9.44 x 10^{-4}. Genomic substitution rate estimates of *env* C2V3 genes (n=418) were obtained by Bayesian MCMC coalescent analyses, as implemented in BEAST. The best fitting nucleotide substitution model was identified as the TVM+F+R6 model (AIC value (6298.17) and BIC value (66427.8)). The lowest AICM values were obtained with the strict clock and coalescent constant size tree models (3×10^8 chain length) that were used to generate the MCC tree. The Bayesian phylogenetic tree were reconstructed employing the best fit model as maximum clade credibility tree *in Supplementary* File S1. By Maximum likelihood analysis, the one major Korean B cluster (including global B isolates) was analyzed. Genomic substitution rate for the C2V3 genes was estimated to be 4.53 x 10^{-3} substitutions/site/year (95% highest posterior density (HPD):3.94 x 10^{-3}–5.16 x 10^{-3} substitutions/site/year).

The estimated evolutionary rates of Korean B and the global B were 4.29×10^{-3} (95% HPD: 3.10×10^{-3} – 5.49×10^{-3}) and 7.46×10^{-4} (95% HPD: 1.39×10^{-7} – 2.42×10^{-3}) substitutions/site/year using the coalescent bayesian skyline tree models, respectively (P < 0.0001). The mean MRCA of the Korean B subtype was 1985.2 (95% HPD: 1977.8–1991.9), as determined using a Bayesian model.

Estimating demographic history

To analyze evolutionary rates for 296 Korean B variants and 40 global B variants, we performed Bayesian skyline plot analysis with Uncorrelated relaxed clock model and Bayesian Skyline Tree model (Chain length, 3×10^8) (Figure 3). Our results showed that the effective population size (number of infected individuals) of HIV-1 virus experienced four major stage since 1992. This analysis for effective population size showed that the size of the Korean B cluster increased steadily until approximately 2000, followed by an exponential growth from 2000 to 2002 and a stationary phase up to 2011 (Figure 3). The population history of global B clusters in Korea was estimated to approach a stationary phase.



We reconstructed the demographic histories of Korean B and global B. Effective population size of HIV-1 Korean B cluster (red) and the global B cluster (blue). The mean and 95% central posterior density intervals of the effective population size are shown on a log scale

Figure 3: Effective population size of HIV-1 through Bayesian skyline plot analysis based on the *env* C2V3 sequence

Discussion

We characterized the HIV-1 *env* C2V3 region of isolates from newly diagnosed Koreans with HIV infection from 2006 to 2011 in Korea. According to phylogenetic analyses to evaluate the molecular epidemiological characteristics of HIV-1 infections, the proportion of subtype B was 91.8%, and Korean B in subtype B was predominant (86.7%). CRF07_BC and CRF02_AG were newly detected in this study. Our previous molecular analysis of HIV-1 from 1985 to 2005 revealed the diversity of HIV-1 in Korea, including subtypes B (80.0%), A (8.1%), CRF01_AE (7.0%), G (3.0%), and C (1.4%), with subtypes D, F, and H detected in each isolate. The proportion of Korean B in this study was similar to that previously reported (87.3%) [5]. The Korean B strain had been the predominant strain for 26 years since 1985 in Korea.

The first report regarding HIV phylogenetics in Korea was published in 1998 [23], 13 years after the first HIV case was identified. This study analyzed the *nef* gene in 46 HIV-1-infected Koreans and identified that Korean B isolates formed a distinct monophyletic clade within HIV-1 subtype B, which was not related to any of the sequences reported from other countries available in the Los Alamos Database or GenBank. Korean B presumably originated from strains in the USA through a founder effect [12,24,25]. Subsequently, several studies consistently reported this unique Korean clade B (KCB, Korean B) by analyzing various genes such as *nef* [23,26], *env* [4,24], *vif* [26], and *pol* [27,28].

Variability in the V3 loop tip motif is potentially associated with HIV-1 co-receptor usage and disease progression. GPGQ is the most common tetrameric tip motif in the V3 loop among all HIV-1 subtypes, and most subtype B isolates have a high proportion of GPGR, and some have an alternate motif such as GWGR [6,29]. The Korean B was distinguished by the high proportion of GPGS (41.2%). Furthermore, most Korean B isolates were predicted as R5 variants (98.1%). R5 virus infections are predominant in individuals positive for HIV-1 subtype B or who are in the early disease stages of HIV infection (80–90%) [10,30].

X4 viruses emerging in later stages of HIV-1 infection are associated with more rapid depletion of CD4+ T cells [31]. However, in the present study, there were no differences not only among CD4+ cell counts but also subtypes, transmission routes, and age in X4 virus-infected cases. The GPGR motif was present in most of the X4 virus-infected cases [32]; therefore, we specu-

lated that the GPGS variant of Korean B is less pathogenic than other HIV-1 isolates. HIV diversity impacts most aspects of the HIV pandemic, including diagnosis, pathogenesis, transmission, clinical treatment, and vaccine development. Although variation has been observed in different genes and in different regions of the same gene, the variability within a subtype is 8–17% in comparison with 17–35% between subtypes [33,34]. Sequence diversity allows for more rapid adaptation to a changing environment and contributes to the evolution of HIV-1.

Our results estimated that HIV-1 Korean B and global subthype B have evolution rates 4.29×10^{-3} and 7.46×10^{-4} substitutions/site/year, respectively. These results indicated that Korean B and global B have evolved independently in Korea. Furthermore, Korean B of 2006-2011 in this study evolved more slowly than Korean B of 1990-2005 in the previous study [2]. In subtype B, the evolutionary rate of the *env* V3 region was estimated as $2.3–6.7 \times 10^{-3}$ nucleotide substitutions/site/year, while that for the entire *env* gene was approximately $1.0–1.7 \times 10^{-3}$ nucleotide substitutions/site/year [35]. This discrepancy is probably related to the exact gene region examined herein, alignment, substitution model, homogeneity of datasets, population size, and transmission route of HIV-1 epidemic. In such cases, the evolutionary rate depends on the intra-host evolution rate or bottleneck effects, and antiretroviral treatment influences the evolution of subtype B viruses [36,37]. Although HIV transmission in Korea primarily occurs through sexual contacts, the virus is also transmitted through networks with high rates of partner exchanges, including injection drug use in Europe and the Americas [38,39]. In Korea, global B variants emerged approximately 10 years before Korean B variants, and their prevalence increased between the 1960s and 1990s, having stabilized since the late 1990s [2]. We suggest that global B was more heterogeneously disseminated and less efficient than Korean B.

In the phylogenetic tree analysis of the *env* C2V3 gene, Korean B formed a distinct monophyletic clade within HIV-1 subtype B. Furthermore, no association with risk factors of HIV transmission or geographic dissemination was observed in the epidemic clusters. Even in newly diagnosed individuals, it was often difficult to accurately determine the time and duration of infection, which is problematic because infection time affects genetic distance and, consequently, the evolutionary rate of HIV-1.

This study has two limitations. First, the transmission route was determined based on self-reported sexual behavior, which may not always be accurate. In particular, as the present study population comprised newly diagnosed cases, we expected a low proportion of individuals in the MSM group because some individuals in this group did not wish to reveal their sexual orientation [40]. The Korea HIV/AIDS cohort study reported a higher percentage for an individual in the MSM group than the present study [41,42]. Therefore, our results should be interpreted with caution. The second limitation is the partial genome-based HIV-1 analysis. To date, many epidemiological studies have been conducted using phylogenetic analysis of small portions of the genome sequence, however, recombination and sequence diversity of a complex genome cannot be completely characterized when partial genome sequences are used for HIV-1 subtyping. Although viral diversity in southeast Asia is quite complex owing to the constant traveling or migration between countries in this region, CRF or URF strains may be present, we did not identify any recombinant strains in this study. We recently reported recombinant form of nearly full-length HIV-1 sequence genome [43]. Such events affect most aspects of the HIV pandemic; therefore, further studies are needed to improve the resolution of the HIV-1 genomic diversity and transmission dynamics.

Conclusions

Korean B accounts for most nationwide cases of HIV-1 in Korea. The present results provide novel insights to further the current understanding of the current characteristics and dynamics of the HIV-1 epidemic in Korea. Further continuous studies are required to monitor the spread of the virus in Korea.

Acknowledgments

This research was supported by an intramural grant of the Korea National Institute of Health (2017-NI51003-). S.R. and M-K.K conceived and designed the study, analyzed and interpreted the data, and wrote the original manuscript. Y-S.J. contributed to study conception. M.Y. analyzed and formatted the data. K.K. and S.K. interpreted the results, and critically reviewed the manuscript. All authors read and approved the final article.

References

1. Jeon Y, cha J, kim T, Shim E (2020) HIV/AIDS notifications in Korea, 2019. Public Health Weekly Report, KCDC 13.

2. Kim GJ, Yun MR, Koo MJ, Shin BG, Lee JS, et al. (2012) Estimating the Origin and Evolution Characteristics for Korean HIV Type 1 Subtype B Using Bayesian Phylogenetic Analysis. Aids Res Human Retrovirus 28: 880-4.

3. Lee JH, Kim GJ, Choi BS, Hong KJ, Heo MK, et al.(2010) Increasing late diagnosis in HIV infection in South Korea: 2000-2007. Bmc Public Health 10.

4. Kim YB, Cho YK, Lee HJ, Kim CK, Kim YK, et al. (1999) Molecular phylogenetic analysis of human immunodeficiency virus type 1 strains obtained from Korean patients: env gene sequences. AIDS Res Hum Retroviruses 15: 303-7.

5. Kim GJ, Nam JG, Shin BG, Kee MK, Kim EJ, et al. (2008) 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 48: 127-32.

6. Tomasini-Grotto RM, Montes B, Triglia D, Torres-Bracconi C, Aliano-Block J, et al. (2010) Variability of the conserved v3 loop tip motif in hiv-1 subtype B isolates collected from brazilian and French patients. Braz J Microbiol 41: 720-8.

7. Tamamis P, Floudas CA (2014) Molecular recognition of CCR5 by an HIV-1 gp120 V3 loop. PLoS One 9: e95767.

8. Tamamis P, Floudas CA (2013) Molecular recognition of CXCR4 by a dual tropic HIV-1 gp120 V3 loop. Biophys J 105: 1502-14.

9. Kitawi RC, Hunja CW, Aman R, Ogutu BR, Muigai AW, et al. (2017) Partial HIV C2V3 envelope sequence analysis reveals association of coreceptor tropism, envelope glycosylation and viral genotypic variability among Kenyan patients on HAART. Virol J 14: 29.

10. Panos G, Watson DC (2015) Effect of HIV-1 subtype and tropism on treatment with chemokine coreceptor entry inhibitors; overview of viral entry inhibition. Crit Rev Microbiol 41: 473-87.

11. Castley A, Sawleshwarkar S, Varma R, Herring B, Thapa K et al. (2017) A national study of the molecular epidemiology of HIV-1 in Australia 2005-2012. PLoS One 12: e0170601.

12. Junqueira DM, Almeida SE (2016) HIV-1 subtype B: Traces of a pandemic. Virology 495: 173-84.

13. Sato H, Shiino T, Kodaka N, Taniguchi K, Tomita Y, et al. (1999) Evolution and biological characterization of human immunodeficiency virus type 1 subtype E gp120 V3 sequences following horizontal and vertical virus transmission in a single family. J Virol 73: 3551-9.

14. Mutenherwa F, Wassenaar DR, de Oliveira T (2019) Ethical issues associated with HIV molecular epidemiology: a qualitative exploratory study using inductive analytic approaches. BMC Med Ethics 20: 67.

15. KCDC (2011) Guideline for HIV/AIDS KCDC.

16. Delwart EL, Herring B, Rodrigo AG, Mullins JI (1995) Genetic subtyping of human immunodeficiency virus using a heteroduplex mobility assay. PCR Methods Appl 4: S202-16.

17. Rambaut A, Lam TT, Max Carvalho L, Pybus OG (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2: vew007.

18. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

19. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32: 268-74.

20. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, et al. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol 29: 2157-67.

21. Baele G, Lemey P, Vansteelandt S (2013) Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. BMC Bioinformatics 14: 85.

22. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol 67: 901-4.

23. Kang MR, Cho YK, Chun J, Kim YB, Lee I, et al. (1998) Phylogenetic analysis of the nef gene reveals a distinctive monophyletic clade in Korean HIV-1 cases. J Acquir Immune Defic Syndr Hum Retrovirol 17: 58-68.

24. Daniels RS, Kang C, Patel D, Xiang Z, Douglas NW, et al.(2003) An HIV type 1 subtype B founder effect in Korea: gp160 signature patterns infer circulation of CTL-escape strains at the population level. AIDS Res Hum Retroviruses 19: 631-41.

25. Cho YK, Kim JE, Foley BT (2019) Genetic Analysis of the Full-Length gag Gene from the Earliest Korean Subclade B of HIV-1: An Outbreak among Korean Hemophiliacs. Viruses 11.

26. Park CS, Kim MS, Lee SD, Kim SS, Lee KM, et al. (2006) Molecular phylogenetic analysis of HIV-1 vif gene from Korean isolates. J Microbiol 44: 655-9.

27. Chin BS, Lee SH, Kim GJ, Kee MK, Suh SD, el al. (2007) Early identification of seronegative human immunodeficiency virus type 1 infection with severe presentation. J Clin Microbiol 45: 1659-62.

28. Sung H, Foley BT, Ahn SH, Kim YB, Chae JD, et al. (2003) Natural polymorphisms of protease in protease inhibitor-naive HIV-1 infected patients in Korea: a novel L63M in subtype B. AIDS Res Hum Retroviruses 19: 525-30.

29. Kim EY, Cho YS, Maeng SH, Kang C, Nam JG, et al. (1999) Characterization of V3 loop sequences from HIV type 1 subtype B in South Korea: predominance of the GPGS motif. AIDS Res Hum Retroviruses 15: 681-6.

30. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, et al. (2008) Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105: 7552-7.

31. Ho SH, Shek L, Gettie A, Blanchard J, Cheng-Mayer C (2005) V3 loop-determined coreceptor preference dictates the dynamics of CD4+-T-cell loss in simian-human immunodeficiency virus-infected macaques. J Virol 79: 12296-303.

32. Zhou HZ, Xu HF, Xin XM, Guan XR, Zhou (2011) Position 22 of the V3 Loop is Associated with Co-Receptor Usage and Disease Progression in HIV-1 Subtype B Isolates. Curr Hiv Res 9: 636-41.

33. Hemelaar J (2013) Implications of HIV diversity for the HIV-1 pandemic. J Infect 66: 391-400.

34. Hemelaar J (2012) The origin and diversity of the HIV-1 pandemic. Trends Mol Med 18: 182-92.

35. Maljkovic Berry I, Ribeiro R, Kothari M, Athreya G, Daniels M, et al. (2007) Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. J Virol 81: 10625-35.

36. Li X, Zhu K, Xue Y, Wei F, Gao R, et al (2017) Multiple introductions and onward transmission of HIV-1 subtype B strains in Shanghai, China. J Infect 75: 160-8.

37. Nijhuis M, Boucher CA, Schipper P, Leitner T, Schuurman R, et al. (1998) Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. Proc Natl Acad Sci U S A 95:14441-6.

38. Walker PR, Pybus OG, Rambaut A, Holmes EC (2005) Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. Infect Genet Evol 5: 199-208.

39. Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimaraes ML, Chequer-Fernandez SL et al. (2007) Demographic history of HIV-1 subtypes B and F in Brazil. Infect Genet Evol 7: 263-70.

40. Kee MK, Park CM, Chang CG, Go UY (2004) Sexual Behavioral Characteristics and the Knowledge of HIV/AIDS among Men who have Sex with Men in Republic of Korea. J Prev Med Public Health 37: 5.

41. Oh DH, Ahn JY, Kim SI, Kim MJ, Woo JH, et al. (2017) Metabolic Complications among Korean Patients with HIV Infection: The Korea HIV/AIDS Cohort Study. J Korean Med Sci 32: 1268-74.

42. Kim EJ, Ahn JY, Kim YJ, Wie SH, Park DW, et al. (2017) The Prevalence and Risk Factors of Renal Insufficiency among Korean HIV-Infected Patients: The Korea HIV/AIDS Cohort Study. Infect Chemother 49: 194-204.

43. Ryou S, Yoo M, Kim K, Kim S, Kim SI, et al. (2021) Characterization of HIV-1 recombinant and subtype B near full-length genome among men who have sex with men in South Korea. Sci Rep 11: 4122.

44. Nguyen LT, Schmidt HA, von Haeseler A (2015) Minh BQ: IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 2015, 32: 268-274.

45. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

Supplementary information

Table S1: Representative reference sequences of HIV-1 subtypes for the Maximum Likelihood analysis in this study

Subtype Reference	GenBank accession	Sampling Country	Sampling Year
A1	AB253429	UG	1992
A1	AB253421	RW	1992
A1	FJ641684	KE	1994
A1	DQ676872	AU	2003
A1	AY175007	KE	-
A1	FJ368	RW	2005
A1	FJ187475	RW	2005
A2	AF286238	CD	1997
A1A2	AY174913	KE	-
B	K03455	FR	1983
B	M38431	US	1984
B	M93258	US	1986
B	AY173951	TH	1990
B	KC690411	KR	1996
B	AY331295	-	1998
B	KC690513	KR	2000
B	KC690583	KR	2001
B	KC690638	US	2002
B	JQ429433.2	KR	2002
B	KC690711	KR	2003
B	KC690775	KR	2004
B	KY820528.1	KR	2015
C	U52953	BR	1992
C	AF067155	IN	1995
C	KC863154	MN	1996
C	AY162225	ZA	1998
C	AY772699	ZA	2004
C	JQ061131	ZA	2007
D	K03454	CD	1983
D	U88824	UG	1994
D	AY736831	CM	2000
D	AY371157	CM	2001
G	AF061641	KE	1993
G	DQ168573	NG	2001
G	EU786670	ES	2005
AE	U54771	TH	1990
AE	GU564221	CN	2005

AE	GQ477441	AF	2007
AE	KM217974	CN	2007
AG	AF063223	FR	1991
AG	AJ389769	NG	1996
AG	AY271690	CM	1999
AG	DQ168577	NG	2001
AG	JX500705	RU	2012
BC	AY727526	BR	2004
BC	HQ292449	CN	2004
BC	HQ669162	CN	2006
BC	HQ6691526	CN	2006
BC	HM067748	CN	2006
BC	EU363834	CN	2007
cpx	KF241229	GW	2010

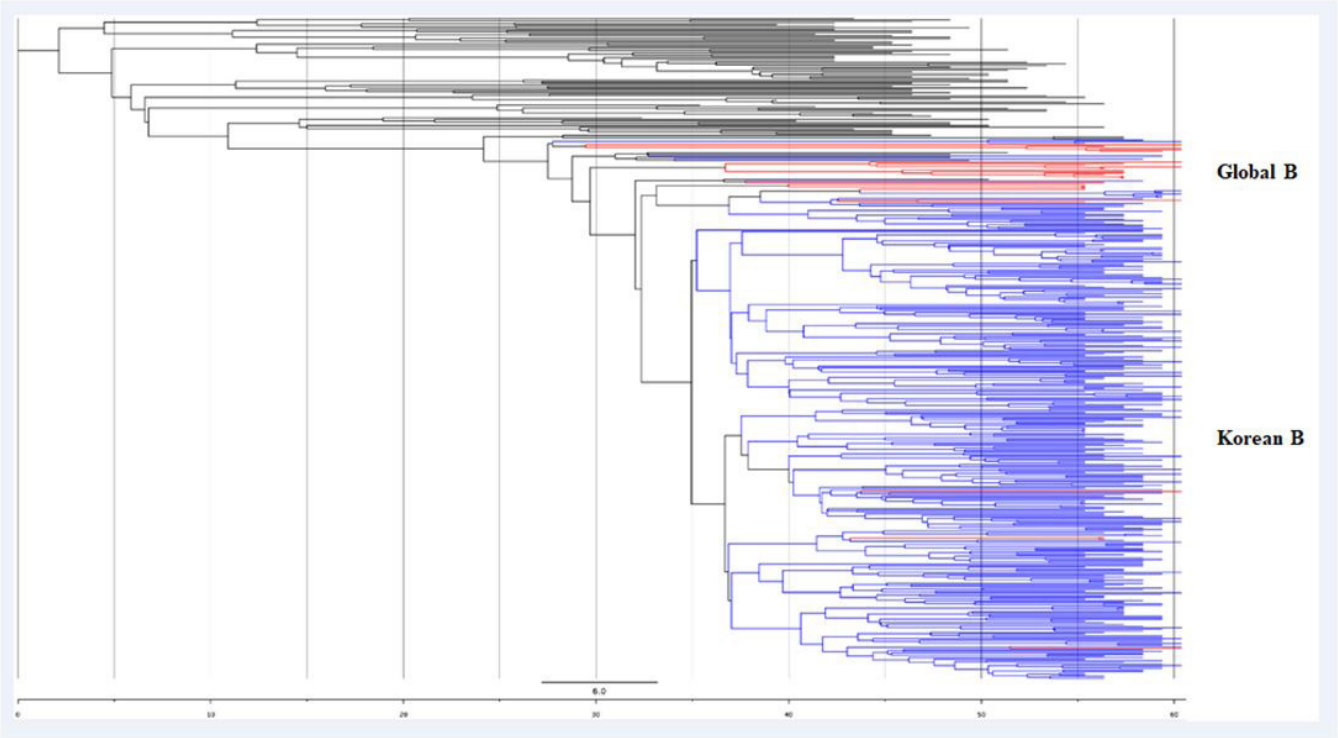
*cpx, complex recombination of several subtypes

Table S2: Details of subtype reference sequences downloaded from the Los Alamos National Laboratory HIV Database for Bayesian phylogenetic analysis in this study

Subtype reference	GenBank accession	Sampling country	Sampling year
A1	DQ676872	AU	2003
AU	FM877777	CD	1997
AU	FM877780	CD	2002
AU	FM877782	CD	2002
AG	AY271690	CM	1999
AE	GU564221	CN	2005
A2D	AF457051	KE	1999
A2D	AF457072	KE	1999
A2D	AY945737	KE	1991
BC	AY727527	BR	2004
BC	AF286230	CN	1998
BC	AY008715	CN	1997
BF	EU581827	BO	2002
BF	AF385936	AR	1999
BF	AF408629	AR	1997
BF	AF408630	AR	1997
BF	AY771590	BR	2002
BF	DQ085873	BR	1999
BF	DQ085876	BR	2001
BF	DQ358801	BR	2001
BF	DQ358802	BR	2001
BF	EF091932	BR	2002
BF	EU735534	BR	2003
BF	EU735536	BR	2003

BF	EU735539	BR	2005
BF	EU735540	BR	2004
BF	FJ358521	CL	2000
BF	FJ670529	ES	2008
BF	GQ372987	ES	2008
BG	AY586545	CU	1999
C	U52953	BR	1992
C	U46016	ET	1986
C	AF067155	IN	1995
D	K03454	CD	1983
D	AY371157	CM	2001
DF	AF07699	BE	1993
DF	AY227107	ES	1999
F1	AF075703	FI	1993
F1	AJ249238	FR	1996
F1	AF077336	BE	1993
F1	AF005494	BR	1993
F2	AF377956	CM	1997
F2	AJ249236	CM	1995
F2	AJ249237	CM	1995
F2	AY371158	CM	2002
G	AY586549	CU	1999
G	AF084936	BE	1996
G	AF061641	KE	1993
H	AF190127	BE	1993
H	AF190128	BE	1993
H	AF005496	CF	1990
H	FJ711703	GB	2000
J	EF614151	CD	1997
K	AJ249235	CD	1997
K	AJ249239	CM	1996
cpx	AF064699	AU	1996
cpx	AJ404325	CD	1997
cpx	AF377959	CM	1997
cpx	AF460972	CM	1996
cpx	AJ291718	CM	1997
cpx	FN392876	CM	1997
cpx	AY586541	CU	1999
cpx	AY588971	CU	1999
cpx	AF049337	CY	1994
cpx	AM851091	FR	2004
cpx	HQ385478	GM	1997

* cpx, complex recombination of several subtypes



Bayesian inference analysis was carried out using the Markov Chain Monte Carlo algorithm. The analysis involved 418 *env* sequences (339 Korean isolates and 79 reference sequences). Korean samples are shown in blue (Korean B) and red (global B). The transversion model (TVM) and FreeRate model with six rate categories (R6) among empirical base frequencies (F) (TVM+F+R6) were selected using Model Finder of IQ-TREE v.1.6.12. [44] The strict clock and coalescent constant size tree models were selected with a chain length of 3×10^8 in BEAST v.1.10.4.[45]

Supplementary Figure 1: Maximum clade credibility (MCC) tree of HIV-1 *env* C2V3 regions