Estimating the Origin and Evolution Characteristics for Korean HIV Type 1 Subtype B Using Bayesian Phylogenetic Analysis

Gab Jung Kim,* Mi-Ran Yun,* Min Jee Koo, Bo-Gyeong Shin, Joo-Shil Lee, and Sung Soon Kim

Abstract

The majority of Korean human immunodeficiency virus type 1 (HIV-1) isolates are composed of the Korean clade B strain that is distinct from the subtype B prevalent in North America and Europe. However, it is still not clear how HIV-1 was introduced, transmitted, and evolved within the Korean population. To identify the evolutionary characteristics of Korean HIV-1, we estimate the molecular epidemic history of HIV-1 subtype B gp120 *env* in Korea in comparison with sequences isolated from other geographic locations. A Bayesian Markov chain Monte Carlo (MCMC) statistical inference was used to estimate the time of divergence of subtype B. The estimated time of divergence of subtype B and the distinct monophyletic Korean B cluster was estimated to be in the early and mid-1960s, respectively. Substitution rates were estimated at 7.3×10^{-3} and 8.0×10^{-3} substitutions per site per year for HIV-1 subtype B and Korean clade B, respectively. The demographic dynamics of two Korean data sets showed that the effective number of infections in Korea increased rapidly until the early 1980s, and then the rate only slowly increased until the mid-1990s when the population growth approached a steady-state. These results suggest that the growth rate of prevalent HIV-1 strains in Korea was lower than in other countries, suggesting that the evolution of HIV-1 Korean clade B was relatively slow. Furthermore, the limited transmission of HIV-1 within the Korean population likely led to the independent evolution of this virus to form the HIV-1 Korean clade B.

Introduction

HUMAN IMMUNODEFICIENCY VIRUS type 1 (HIV-1) subtype B is the predominant subtype circulating globally and is prevalent in North America, Europe, and East Asia, including Korea. The first cases of HIV-1 were observed in men who have sex with men and Haitian immigrants in the United States in the early 1980s.^{1,2} Since then, HIV-1 subtype B has spread extremely rapidly within specific risk groups, such as homosexuals and injection drug users.³ There was a period of rapid, exponential growth during the early stages of the pandemic, followed by a declining rate since the 1980s.⁴⁻⁶

HIV-1 subtype B sequences from Western Europe, North America, and Australia had a weak geographic structure within the phylogeny of the subtype B lineage of HIV-1, but some geographic clusters have been assembled using sequences from Southeast Asia (Thai-B or B'),^{7,8} Brazil,⁹ and South Korea.¹⁰ In the case of HIV-1 infection in Korea, through 2010, 7656 HIV-1 infections have been reported since

the first identification of AIDS in 1985.¹¹ Previous studies have demonstrated the prevalence of HIV-1 B in Korea and the existence of a unique Korean clade B strain. However, it remains unknown how subtype B was introduced into the Korean population and, more importantly, how the virus has subsequently been transmitted and evolved.

The construction of phylogenetic relationships using sampled gene sequences reveals information about the history of a viral population and the course of a viral epidemic over time.¹² Many phylogenetic studies using the coalescent theory of population genetics¹³ have been previously employed to investigate the demographic history and the growth pattern of HIV populations.^{4-6,12,14,15} These studies have typically introduced the coalescent-based skyline plot^{4,15} that generates a plot of the effective number of infections over time.

In the present study, the molecular epidemiology and epidemic history of HIV-1 subtype B circulating in the Korean population and the Korean clade B distinctive monophyletic clustering were explored by conducting phylogenetic and

Division of AIDS, Department of Immunology and Pathology, National Institute of Health South Korea, Cheongwon-gun, Chungcheongbuk-do, Korea.

*These authors contributed equally to this work.

population genetic analyses of *env* gene sequences. These analyses were based on coalescent theory to estimate the age of HIV-1 subtype B strains and the historical rates at which these strains spread through the Korean population.

Materials and Methods

Study population

We collected blood samples and epidemiological data for the newly diagnosed HIV individuals from local public health centers in Korea. Plasma and peripheral blood mononuclear cells (PBMCs) were isolated from blood samples by centrifugation and preserved at -70°C . Each year, 10–15% of newly diagnosed HIV/AIDS infected individuals were selected by random sampling for the genetic surveillance study. The dataset consisted of 560 HIV-1 subtype B 1.2 kb *env* (gp120) sequences with known sampling date between 1990 and 2005 in Korea. These sequences were deposited in an internal sequence database in the Korea National Institute of Health. In addition, we collected 2347 worldwide subtype B gp120 *env* sequences, which were composed of 1563 North American sequences (1516 from the United States), 532 European, 90 Asian, and 162 others, available in the GenBank database and the Los Alamos HIV Sequence Database (<http://www.hiv.lanl.gov>). These sequences were used to examine the evolutionary structure of subtype B worldwide.

Sequencing and phylogenetic analysis of Korean HIV-1 isolates

To identify HIV-1 lineage derived from single independent introductions of the virus into the Korean population, we used 560 HIV-1 subtype B *env* gp120 sequences and 2347 worldwide subtype B *env* gp120 sequences.

Genomic DNA was extracted from the isolated and preserved PBMCs of HIV/AIDS-infected individuals using a QIAamp DNA blood mini kit (Qiagen, Valencia, CA). After amplifying the 1.2 kb V1–V5 region of the HIV-1 *env* gene using nested polymerase chain reaction (PCR) with a previously described primer set (ed3/ed14, ed5/ed12) (19), the PCR product was directly sequenced with primer ed5 (6556–6581), ed12 (7822–7792), ed31 (6816–6844), ed33 (7359–7380), r25 (6873–6857), and v3f (7314–7334). The sequencing reaction for the *env* region used the ABI Prism Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Applied Biosystems, Foster City, CA) in an automated ABI prism 3730 DNA sequencer (Perkin-Elmer, Norwalk, CT).

Nucleotide sequences were aligned using Clustal X¹⁶ and MegAlign (DNASTar, WI). A neighbor-joining phylogenetic tree was constructed using PAUP* (Dave L. Swofford, Sinauer Associates) with the Hasegawa-Kishino-Yano (HKY85) model of nucleotide substitution.¹⁷ The statistical robustness of the neighbor-joining phylogenetic topologies was assessed by bootstrapping with 1000 replicates.

Estimation of evolutionary rates and dates

The evolutionary rate and the date of origin were estimated using Bayesian Markov chain Monte Carlo (MCMC) as implemented in BEAST.¹⁸ The analyses were estimated using an uncorrelated lognormal relaxed molecular clock under the general time-reversible (GTR) of nucleotide substitution model with gamma distribution (Γ) and a proportion of in-

variable sites (I), as selected by the program JmodelTest.¹⁹ Several independent runs were performed with a chain length 5.0×10^7 for HIV-1 subtype B. Effective sampling sizes (ESSs) of the statistical estimates for all runs were >100 . The results were visualized with Tracer²⁰ using combined estimates of the independent runs.

Estimation of population history

The past population histories of HIV-1 subtype B and monophyletic Korean clade B were estimated from the gp120 heterochronous *env* gene sequences using a Bayesian skyline plot (BSP) implemented in BEAST. The BSP method used the MCMC procedure to estimate the distribution of generalized skyline plots and to create 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 result of the Bayesian MCMC was used to calculate a marginal posterior distribution of the demographic inference, and estimated parameters included the effective population size at the most recent time of sampling, N_e (the effective number of prevalent infections). The posterior samples were analyzed using the Tracer program.²⁰

Results

Phylogenetic analyses

A neighbor-joining phylogenetic tree was constructed from 560 Korean subtype B and 2347 worldwide subtype B *env* sequences (Fig. 1). This global phylogenetic tree demonstrated that American and European sequences were widely dispersed throughout the tree and were not clearly phylogenetically distinguishable. Among the 560 HIV-1 subtype B isolates from Koreans, phylogenetic tree analysis indicated that 491 (88%) belonged to the distinct Korean clade B (75% bootstrap value), which was discriminated from other countries. However, no specific demographic or epidemiologic feature, including age, sex, or the date of isolates, was significantly associated with clustering in this analysis.

This tree showed a distinct monophyletic Korean cluster that does not include sequences from North American and European subtype B, as identified in the previous study.¹⁰ Namely, among the 560 HIV-1 subtype B isolates, phylogenetic analysis indicated that 491 sequences belonged to the distinct monophyletic Korean clade B group.

Estimation of evolutionary characteristics of HIV-1 subtype B *env* gene sequences

Rates of evolution, measured as the number of nucleotide substitutions per site per year, were estimated using a Bayesian MCMC method (Table 1). The mean estimated substitution rates were 8.7×10^{-3} substitutions per site per year using a constant size model and 7.3×10^{-3} substitutions per site per year using BSP for HIV-1 subtype B sequences. The estimates for Korean clade B were 9.4×10^{-3} and 8.0×10^{-3} substitutions per site per year for constant size and BSP models, respectively. Using the same approach, the estimated time of most recent common ancestor (T_{MRCA}) is related to the age of the sampled genetic diversity (Table 1). The mean T_{MRCA} was estimated to be around 1961 and 1967 for subtype B and Korean clade B, respectively, with a Bayesian model. The Korean clade B was estimated to have emerged less than a

FIG. 1. Neighbor-joining phylogenetic relationship generated from 2907 HIV-1 subtype B gp120 *env* sequences. Korean clade B (in red) formed a distinct cluster from worldwide HIV-1 subtype B. This tree showed the U.S. and European strains in different colors (Spain in cyan, France in blue, Great Britain in violet). U.S. and European sequences are scattered widely through tree.

decade after the introduction of subtype B into the Korean population, after which the Korean clade B was formed in the years since.

Demographic history of Korean HIV-1 subtype B

To infer HIV-1 dynamics in Korea, a Bayesian skyline coalescent model was employed that provides a piecewise graphic representation of changes in genetic diversity (population size) over time with the nonparametric estimates for the Korean cluster. This analysis showed that the effective numbers of infections increased rapidly until the early 1980s, followed by a slower increase until the mid-1990s, and thereafter reached a steady-state (Fig. 2). This pattern was observed for both HIV-1 subtype B and Korean clade B analyses.

Using BSP analysis results, a plot of growth rate was also generated, showing that the mean growth rates were estimated at 0.15 and 0.17 per year for subtype B and Korean clade B (Table 1). These estimates were approximately three to five times lower than that reported for other countries.^{5,21}

Discussion

The purpose of this study was to investigate the evolution and demographic history of the HIV-1 subtype B epidemic in Korea. HIV-1 subtype B accounts for approximately 80% of HIV-1 infections in Korea and the Korean strain forms a distinct monophyletic cluster, called Korean clade B. We attempted to estimate the time of origin for HIV-1 subtype B and Korean clade B in this study. As shown in Table 1, estimates revealed that the HIV-1 subtype B prevailing in Korea, was emerged in the early 1960s and then evolved and formed the distinct Korean clade B in the mid-1960s, less than a decade after its emergence.

The T_{MRCA} for HIV-1 subtype B in high-income countries, including the United States and Western Europe, has been estimated to be around the mid-1950s to late 1960s.^{5,22,23} Subtype B epidemics in United States, Western Europe, and Brazil were also estimated to have started around this time, which is consistent with the concurrent detection of the first acquired immune deficiency syndrome (AIDS) cases in those regions during the early 1980s.²¹ In addition, the T_{MRCA} of HIV-1 Thai-B was also estimated to be in the mid-1980s for the *env* gene.⁸

The time of origin or evolutionary rate often varies within a study group. These differences may be associated with the target gene location, the multiple sequence alignment method used, the substitution model, and/or the degree of homogeneity in the data set used for estimation.⁵ For example, the T_{MRCA} for HIV-1 subtype B in the United States was 1954–1955 under a variable rate of evolution and a Bayesian approach, and around 1967 with strict the molecular clock model.²² A comparison of our findings with this U.S. study, using the same analytical model, estimated that the origin of HIV-1 subtype B in Korea was 6–7 years after its origin in the United States. Also, according to the phylogenetic tree, dominant HIV-1 strains in early branching lineages after first HIV-1 infection in Korea were mingled with HIV-1 subtype B sequences from the United State and Europe (Fig. 1). Thus, we assumed that HIV-1 subtype B was transmitted from major subtype B epidemic regions, such as the Americas and Europe, and the occurrence of Korean clade B evolved since it was first introduced into Korea.

TABLE 1. ESTIMATED SUBSTITUTION PARAMETERS FOR SUBTYPE B AND KOREAN CLADE B

Dataset	Substitution model	Clock model	Tree model	Date of origin	Rate of evolution ^a	Growth rate ^b
Subtype B	GTR + I + G	Relaxed (uncorrelated lognormal)	Const	1958 (1944.0–1967.0)	0.0087 (0.0078–0.0097)	0.15
			BSP	1961 (1954.6–1966.6)	0.0073 (0.0063–0.0082)	
Korean Clade B			Const	1966.4 (1960.6–1972.0)	0.0094 (0.0084–0.0104)	0.17
			BSP	1966.6 (1961.4–1971.5)	0.0080 (0.0070–0.0091)	

^aUnit of rate of evolution is nucleotide substitutions per site per year. 95% HPD confidence limits are shown in parentheses.

^bUnit of growth rate is year⁻¹.

GTR=General Time-Reversible model; GTR + I + G=GTR model with gamma distribution and proportion of invariable site; Const, constant size; BSP=Bayesian Skyline Plot.

Several studies have shown HIV evolutionary rate estimates of $1\text{--}17 \times 10^{-3}$ substitutions per site per year.^{22,24} After HIV-1 subtype B was introduced into Korea, the nucleotide substitution rates of subtype B were found to be similar to those in other countries. Our results estimated that HIV-1 subtype B and Korean clade B have nucleotide substitution

rates in the range of $7.3\text{--}8.7 \times 10^{-3}$ and of $8.0\text{--}9.4 \times 10^{-3}$ substitutions per site per year, respectively. These estimates are similar to evolutionary rates estimated for HIV-1 subtype B in North America and Europe.²² These findings indicate that Korean HIV-1 did not evolve at an extraordinary rate but perhaps evolved independently to form the distinct Korean clade B population currently observed in Korea.

The population history of HIV-1 subtype B showed that virus infections rapidly increased until the early 1980s from the introduction of HIV-1 into the Korean population. The HIV-1 population then only slowly increased until the mid-1990s, when a steady-state was achieved. The Korean demographic trend is similar to that described by Walker *et al.*, which showed an early rapid expansion of infection cases in the United States, followed by a more recent easing of the increase, which then reached a plateau.²⁴ The growth rate estimate of subtype B in Korea is significantly lower than recently described for the subtype B epidemic in North America.⁶ It has been suggested that such differences in the growth rate of HIV-1 are more likely to reflect differences in transmission networks⁴ and the features of the transmission pool strongly influenced HIV-1 transmission.²¹ The rapid spread of subtype B in North America and Europe was caused by a transmission network with a high partner exchange rate between injection drug user or male homosexual populations.²¹ However, the method of transmission of HIV-1 in Korea is via homosexual or heterosexual contact¹⁰ and the growth rate is approximately three times lower than that of high-income countries because of the limitation of Korean transmission to sexual contacts and of rare partner exchange. This population dynamic reflects the fact that the rate of increase in new HIV-1 infections was steadily maintained in recent years in Korea.¹¹

In summary, this study demonstrated that a unique molecular evolution and population history existed for HIV-1 subtype B in Korea. The time of origin for HIV-1 subtype B was estimated to be slightly later than that of the major global subtype B circulating in areas such as North America and Western Europe. The distinct Korean clade B then emerged as a monophyletic cluster less than a decade after its introduction. The number of HIV-1 infections in Korea increased exponentially after its introduction until the early 1980s and then the epidemic slowed to eventually reach a steady-state level after the mid-1990s. It has previously been

FIG. 2. Demographic history estimated from all subtype B samples in the *env* gp120 region: (A) HIV-1 subtype B and (B) Korean HIV-1 clade B. The demographic histories were estimated by Bayesian Markov chain Monte Carlo (MCMC) inference with the Skyline plot and show changes in the effective number of infections through time (time scale in year). The solid lines are the median estimate and the dashed lines indicate the 95% upper and lower highest posterior density intervals.

shown that differences in increases in the infection rate reflect large differences in transmission networks.⁶ Specific transmission routes such as sexual contact and injection drug use could be responsible for the transmission of viruses within a population.²⁵ Therefore, the small increase in the number of HIV-1 infections in the Korean population could demonstrate that the Korean transmission network was limited and could account for the development of the unique clade of HIV subtype B.

Acknowledgments

This research was supported by an intramural grant of the National Institute of Health, Korea (4842-302-210-13 and 2010-N51002-00).

Author Disclosure Statement

No competing financial interests exist.

References

- Gottlieb MS, Schroff R, Schanker HM, *et al.*: Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: Evidence of a new acquired cellular immunodeficiency. *N Engl J Med* 1981;305:1425–1431.
- Laverdiere M, Tremblay J, Lavallee R, *et al.*: AIDS in Haitian immigrants and in a Caucasian woman closely associated with Haitians. *Can Med Assoc J* 1983;129:1209–1212.
- Selik RM, Haverkos HW, and Curran JW: Acquired immune deficiency syndrome (AIDS) trends in the United States (1978–1982). *Am J Med* 1984;76:493–500.
- Pybus OG, Rambaut A, and Harvey PH: An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 2000;155:1429–1437.
- Robbins KE, Lemey P, Pybus OG, *et al.*: U.S. Human immunodeficiency virus type 1 epidemic: Date of origin, population history, and characterization of early strains. *J Virol* 2003;77:6359–6366.
- Walker PR, Pybus OG, Rambaut A, and Holmes EC: Comparative population dynamics of HIV-1 subtypes B and C: Subtype-specific differences in patterns of epidemic growth. *Infect Genet Evol* 2005;5:199–208.
- Ou CY, Takebe Y, Luo CC, *et al.*: Wide distribution of two subtypes of HIV-1 in Thailand. *AIDS Res Hum Retroviruses* 1992;8:1471–1472.
- Deng X, Liu H, Shao Y, *et al.*: The epidemic origin and molecular properties of B': A founder strain of the HIV-1 transmission in Asia. *AIDS* 2008;22:1851–1858.
- Leal E and Villanova FE: Diversity of HIV-1 subtype B: Implications to the origin of BF recombinants. *PLoS One* 2010;5(7):e11833.
- Kim GJ, Nam JG, Shin BG, *et al.*: Limited genetic variation of Korean HIV-1 Clade B within the population of Korean men who have sex with men. *J AIDS* 2008;48:127–132.
- KCDC: Periodical report of KCDC, 2010 December.
- Holmes EC, Nee S, Rambaut A, *et al.*: Revealing the history of infectious disease epidemics through phylogenetic trees. *Philos Trans R Soc Lond B Biol Sci* 1995;349:33–40.
- Kingman JF: Origins of the coalescent 1974–1982. *Genetics* 2000;156:1461–1463.
- Salemi M, de Oliveira T, Soares MA, *et al.*: Different epidemic potentials of the HIV-1B and C subtypes. *J Mol Evol* 2005;60:598–605.
- Strimmer K and Pybus OG: Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol Biol Evol* 2001;18:2298–2305.
- Larkin MA, Blackshields G, Brown NP, *et al.*: Clustal W and Clustal X version 2.0. *Bioinformatics* 2007;23(21):2947–2948.
- Saitou N and Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425.
- Drummond AJ and Rambaut A: BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7:214.
- Posada D: jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 2008;25(7):1253–1256.
- Rambaut A and Drummond AJ: Tracer: MCMC trace analysis tool. University of Oxford, 2003.
- Bello G, Eyer-Silva WA, Couto-Fernandez JC, *et al.*: Demographic history of HIV-1 subtypes B and F in Brazil. *Infect Genet Evol* 2007;7:263–270.
- Korber B, Muldoon M, Theiler J, *et al.*: Timing the ancestor of the HIV-1 pandemic strains. *Science* 2000;288:1789–1796.
- Gilbert MT, Rambaut A, Wlasiuk G, *et al.*: The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci USA* 2007;104:18566–18570.
- Berry M, Ribeiro R, Kothari M, *et al.*: 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* 2007;81:10625–10635.
- Pybus OG, Charleston MA, Gupta S, *et al.*: The epidemic behavior of the hepatitis C virus. *Science* 2001;292:2323–2325.

Address correspondence to:

Gab Jung Kim

Division of AIDS

Korea National Institute of Health

Osong Health Technology Administration Complex

#187 Osongsaengmyeong2(i)-ro

Gangoe-myeon, Cheongwon-gun,

Chungcheongbuk-do 363-951

Korea

E-mail: gabjkim@korea.kr