

NOTE

Analysis of Substitution Events in HIV-1 *vif* Gene of the Korean Clade

Chan Seung Park¹, Mi Sook Kim¹, Hyun Ah Yi¹, Dong Hun Lee¹, Keon Myung Lee³, and Chan Hee Lee^{1,*}

¹School of Life Sciences, and ³School of Electrical and Computer Engineering, Chungbuk National University, Cheongju, Republic of Korea

(Received December 11, 2006 / Accepted January 12, 2007)

Nucleotide and amino acid substitution pattern in *vif* gene of the Korean clade of HIV-1 isolated from Koreans were analyzed using consensus sequences. At nucleotide level, transition/transversion substitution ratio was 1.88, and nonsynonymous/synonymous substitution ratio was 2.67, suggesting a divergent evolution in the Korean clade. At amino acid level, there were 17 substitutions and G→E substitution at position 37 may be responsible for change in predicted secondary structure.

Keywords: HIV-1, *vif*, Korean clade, substitution ratio

Human immunodeficiency virus (HIV) was detected in Korea for the first time in 1985 and since then the number of HIV-infected persons has accumulated to 4,227 by June 2006. Phylogenetic studies based on nucleotide and amino acid sequences revealed an interesting feature that a majority of the HIV-1 isolated from Koreans clustered to the exclusion of foreign sequences, forming a Korean clade within subtype B (Kang *et al.*, 1998; Kim *et al.*, 1999; Sung *et al.*, 2001; Lee *et al.*, 2003, Park *et al.*, 2006). The presence of the Korean clade is highly unique since, to our knowledge, no other national or ethnic clade has been clearly reported elsewhere. Although the proportion of the Korean clade to the HIVs isolated from Koreans varies depending on the number of sequences or the type of genes analyzed, studies with the most of the gene sequences registered in NCBI GenBank database suggested approximately 60 to 80% of the Korean isolates could be classified as the Korean clade (Lee *et al.*, 2003; Park *et al.*, 2006). Considering the importance of the Korean clade in terms of majority, characterization of the Korean clade is most warranted.

Recent studies suggested that the genetic sequences of the Korean clade differ considerably from those of non-Korean clade (Lee *et al.*, 2003; Park *et al.*, 2006). The sequence difference does not involve functionally important regions, hence the function of the proteins from the Korean clade appears to be unaltered. Nonetheless analysis of the *vif* gene sequence difference identifies several amino acid residues by which the Korean clade is defined and characterized (Park *et al.*, 2006). These signature amino acids are generated by mutation of the preexisting non-Korean clade subtype B. In this study, we attempted to understand how the Korean clade evolved by analyzing the pattern of nucleotide and amino acid substitutions in *vif* gene.

As described previously, Korean clade was identified when a total of 342 *vif* nucleotide sequences from Korean and foreign isolates were analyzed (Park *et al.*, 2006). Of the 233 *vif* nucleotide sequences from Koreans, 172 sequences belonged to the Korean clade. For easier and simple analysis, a consensus sequence (KcB-con) was generated from these 172 sequences of the Korean clade and compared with subtype B consensus sequence (SubB-con) which was obtained from HIV Sequence Database (<http://hiv-web.lanl.gov/content/hiv-db/mainpage.html>). Sequences of the Korean clade were multiple-aligned with CLUSTAL X and the output files in Clustal format (.aln) were used to calculate the frequencies of the nucleotides at each position using SeqAid program developed by the authors. Then the most frequent nucleotides at each position were collected to generate the consensus nucleotide sequence. The consensus amino acid sequence of the Korean clade was generated by translating the consensus nucleotide sequence. The consensus nucleotide and amino acid sequences of the Korean clade are shown in Fig. 1 together with those of SubB-con.

There are several models for nucleotide substitution and sequence evolution. The simplest is Jukes-Cantor (JC) model which assumes equal frequencies of the four bases and equal probabilities for all substitutions. Kimura 2 parameter (K2P) model also assumes equal frequencies of the four bases, but incorporates the generally observed and accepted notion that transitions and transversions have different substitution rates (Kimura, 1980). More complicated model such as HKY85 developed by Hasegawa, Kishino and Yano allows unequal base frequencies (Hasegawa *et al.*, 1985). When the nucleotide sequences of the SubB-con and KcB-con were compared, there were 23 substitutions in KcB, of which 15 were transition and 8 were transversion (Table 1). The observed transition to transversion ratio was calculated to be 1.88 (15/8), supporting K2P model for *vif* gene evolution.

When the 23 substitutions were analyzed according to three positions of codons, 7 substitutions were observed in each

* To whom correspondence should be addressed.
(Tel) 82-43-261-2304; (Fax) 82-43-273-2451
(E-mail) chlee@chungbuk.ac.kr

(A)	SuB-con	AUGGAAAACAGAUGGCAGGUGAUGAUJUGUGGCAAGUAGACAGGAUGAGGAUUAGAACA	60
	KcB-con	AUGGAAAACAGAUGGCAGGUGAUGAUJUGUGGCAAGUAGACAGGAUGAGGAUUAGAACA	
	SuB-con	UGGAAAAGUUUAGUAAAACACCAUAUGUAUUAUUUCAAGGAAAGCUAAGGGAUGGUUUUAU	120
	KcB-con	UGGAAAAGUUUAGUAAAACACCAUAUGUAUUAUUUCAAGGAAAGCUAAGGGAUGGUUUUAU	
	SuB-con	AGACAUCACUAUGAAAGCACUCAUCCAAGAUAAGUUCAGAAGUACACAUCCCACUAGGG	180
	KcB-con	AGACAUCACUAUGAAAGCACUCAUCCAAGAUAAGUUCAGAAGUACACAUCCCACUAGGG	
	SuB-con	GAUGCUJAGAUUGGUAAUAACAACAUAUJUGGGGUCUGCAUACAGGAGAAAGAGACUGGCAU	240
	KcB-con	GAUGCUAAAUAAGUAUAACAACAUAUJUGGGGUCUGCAUACAGGAGAAAGAGAAUGGCAU	
	SuB-con	UUGGGUCAGGGAGUCUCCAUAAGAUGGAGGAAAAAGAGAUUAGCACACAAGUAGACCCU	300
	KcB-con	CUGGGUCAGGGAGUCUCCAUAAGAUGGAGGAAAAAGAGAUUAAACACACAAGUAGACCCU	
SuB-con	GACCUJAGCAGACCAACUAUUAUCUUGUAUUACUUGAUUUUUUUCAGAAUCUGCUAUA	360	
KcB-con	GACCUJAGCAGACCAACUAUUAUCUUGUAUUACUUGAUUUUUUUCAGAAUCUGCUAUA		
SuB-con	AGAAAUGCCAUUAUJAGGACAUUAJAGUJAGUCCUJAGGUGAGAAUAUCAAGCAGGACUAAC	420	
KcB-con	AGACAUGCCAUUAUJAGGACGUUJAGUJAGUCCUJAGGUGAGAAUAUCAAGCAGGACUAAC		
SuB-con	AAGGUJAGGACUCUJACAGUACUUGGCACUJAGCAGCAUUAUAACACCAAAAAAGAUAAAG	480	
KcB-con	AAGGUJAGGACUCUJACAGUACUUGGCACUJAGCAGCAUUAUAACACCAAAAAAGAUAAAG		
SuB-con	CCACCUUUGCCUJAGUJUAACGAAACUGACAGAGGAUJAGUUGAACAAGCCCCAGAAGACC	540	
KcB-con	CCACCUUUGCCUJAGUJUAACGAAACUGACAGAGGAUJAGUUGAACAAGCCCCAGAAGACC		
SuB-con	AAGGGCCACAGAGGGAGCCAUACAUAUGAAUGGACACUAG	579	
KcB-con	AAGGGCCACAGAGGGAGCCAUACAUAUGAAUGGACACUAG		
(B)	SuB-con	MENRWQVM VWQVDRMR RTWKSLVKHHMY SRKAKGWFYRHHYESTHPR SSEVHI PLG	60
	KcB-con	MENRWQVM VWQVDRMR RTWKSLVKHHMY SVKKAKEWVYRHHYESTHPR SSEVHI PLG	
	SuB-con	DARLV TTYWGLHTGERDWHLGQGV EWRKKRYSTQVDPDLADQL HLHYFDCFSESA	120
	KcB-con	DAKLV TTYWGLHTGEREWHLGQGV EWRKKRYNTQVDPDLADKL HLHYFDCFSDSA	
	SuB-con	RNA LGH VSPRCEYQAGHNKVGSLQYLALAL TPCK KPPLPSVTKL TEDRWNPQKT	180
KcB-con	RHA LGRLVRPKCEYQAGHNKVGSLQYLALAL TPCK KPPLPSVRKL TEDRWNPQKT		
SuB-con	KGHRGSHTMNGH	192	
KcB-con	KGHRGSHTMNGH		

Fig. 1. Consensus sequences of the Korean clade and subtype B. Sequences aligned with CLUSTAL X were imported into SeqAid program and the proportions of the nucleotides or amino acids at each position were calculated. Then the most frequent nucleotides or amino acids at each position were collected to generate the consensus sequences. KcB-con, consensus sequence of the Korean clade; SubB-con, consensus sequence of the subtype B. (A), consensus nucleotide sequences; (B) consensus amino acid sequences.

of the first and the second positions, while 9 substitutions were found in the third positions (Table 1). All of the substitutions at the second codon positions resulted in nonsynonymous substitution, and 6 out of 7 substitutions at the first codon positions also resulted in nonsynonymous substitutions. The only synonymous substitution resulting from substitutions at the first codon positions was observed at the nucleotide position 241 (nt-241). On the other hand, 9 substitutions at the third codon positions resulted in 3 nonsynonymous and 6 synonymous substitutions. Interestingly, all the nonsynonymous substitutions were due to transversions, and 5 of the 6 synonymous substitutions were related with transitions. The only exception was at the nt-117 corresponding to the codon number 39, where nonsynonymous substitution resulted from 2 substitutions at the first (nt-115) and the third (nt-117) position of the codon (Table 1). If there was no substitution at nt-115, substitution at nt-117 (which is transition) would result in synonymous substitutions: GUU/GUC (V/V) or UUU/UUC (F/F). Therefore, it is safe to say that all the transitional substitutions at the third codon positions of *vif* gene resulted in synonymous substitutions.

The number of nonsynonymous substitutions was 16 and this number is somewhat smaller than the number of substitutions at the first and second codon positions (7+7=14) as expected (Nei and Kumar, 2000). The number of synonymous substitution was 6 and the ratio of nonsynonymous substitution rate (d_N) to synonymous substitution rate (d_S) was 2.67 (16/6), much greater than 1, suggesting a positive Darwinian selection (for example, Hughes and Nei, 1988) has worked during the evolution of *vif* gene in the Korean clade. In most cases mutations will be deleterious and so disadvantageous to survive. These mutations will be removed from the population by purifying or negative selection. Therefore, d_N is generally much lower than that of d_S due to purifying selection (Kimura, 1983). Positive selection might be imposed if the mutation will be more advantageous to survive. Perhaps HIV-1 of the Korean clade has undergone a positive Darwinian evolution in order to ensure better survival in Korean people.

Nonsynonymous substitution of nucleotide causes amino acid substitution. We found 16 substitutions at amino acid level corresponding to 16 nonsynonymous substitutions of

Table 1. Substitution analysis of nucleotides and amino acids of the Korean clade

Position in sequence ^a	Codon position	Nucleotides			Amino acids		
		Substitution ^b	Ts/Tv ^c	Syn/Non-syn ^d	Position in sequence ^a	Substitution ^b	Class change ^e
98	2	AGG/AAG	Ts	Non-syn	33	R/K	4/4
110	2	GGA/GAA	Ts	Non-syn	37	G/E	2/3
115/117	1,3	UUU/GUC	Ts/Tv	Non-syn	39	F/V	1/1
188	2	AGA/AAA	Ts	Non-syn	63	R/K	4/4
192	3	UUG/UUA	Ts	Syn			
234	3	GAC/GAA	Tv	Non-syn	78	D/E	3/3
241	1	UUG/CUG	Ts	Syn			
284	2	AGC/AAC	Ts	Non-syn	95	S/N	2/2
313	1	CAA/AAA	Tv	Non-syn	105	Q/K	2/4
324	3	CAU/CAC	Ts	Syn			
328	1	UAU/CAU	Ts	Non-syn	110	Y/H	2/4
333	3	UAC/UAU	Ts	Syn			
351	3	GAA/GAC	Tv	Non-syn	117	E/D	3/3
364	1	AAU/CAU	Tv	Non-syn	122	N/H	2/4
380	2	CAU/CGU	Ts	Non-syn	127	H/R	4/4
382	1	AUA/UUA	Tv	Non-syn	128	I/L	1/1
390	3	AGU/AGG	Tv	Non-syn	130	S/R	2/4
395	2	AGG/AAG	Ts	Non-syn	132	R/K	4/4
429	3	GGA/GGG	Ts	Syn			
451	1	GCA/ACA	Ts	Non-syn	151	A/T	1/2
500	2	ACG/AGG	Tv	Non-syn	167	T/R	2/4
507	3	CUG/CUA	Ts	Syn			

^aPosition of the nucleotides or amino acids in *vif* sequence

^bNucleotides or amino acids in SubB-con are in the left and those in KcB-con are in the right side of slash (/)

^cTs: transitional substitutions, Tv: transversal substitutions

^dSyn: synonymous substitutions, Non-syn: nonsynonymous substitutions

^eAmino acids are classified into 4 groups: Group1, hydrophobic (A, V, L, I, P, M, F, W); Group 2, Polar uncharged (G, S, T, C, Y, N, Q); Group 3, Acidic (D, E); Group 4, Basic (K, R, H).

nucleotides. Of the 16 amino acid substitutions, 9 were within same functional group, for example R to K or D to E change (Table 1). These within-group substitutions may not lead to dramatic changes in regional structure and function. On the other hand, inter-group substitutions such as G to E change at position 37 or Q to K change at position 105 may affect the regional structure. Thus, we attempted to compare the secondary structures of SubB-con and KcB-con predicted by Hierarchical Neural Network (HNN) or Secondary Consensus Prediction (SCP) programs provided by Pole BioInformatique Lyonnais (PBIL: <http://pbil.ibcp.fr/htm/index.php>). HNN identified one major difference in the secondary structure between SubB-con and KcB-con: random coil and extended strand structure found in amino acid positions between 34 and 42 in SubB-con disappeared and replaced by α -helix in KcB-con (Fig. 2). SCP also predicted region between aa-31 and aa-37 as the major difference: random coil structure in SubB-con was replaced by α -helix in KcB-con (data not shown). In these regions (between aa-31 and aa-42), 4 amino acid substitutions were identified and only one (aa-37) was inter-group change.

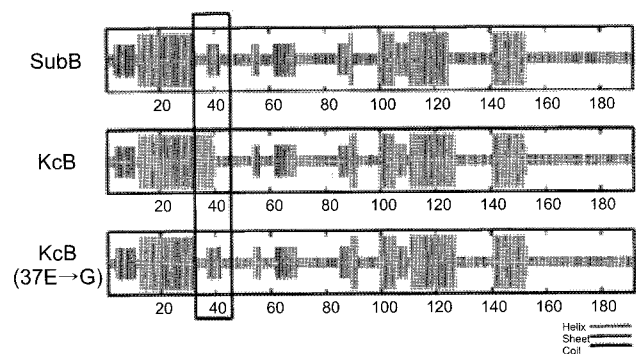


Fig. 2. Predicted secondary structures. Secondary structure was predicted on web site of Pole Bioinformatique Lyonnais (<http://pbil.ibcp.fr/htm/index.php>). The region where the predicted secondary structures are significantly different is boxed for easier identification.

Thus, it was hypothesized that substitution of aa-37 played some role in the secondary structural change of the Korean clade *vif* protein. The E residue in KcB-con was replaced by

G which was found in SubB-con and the resulting secondary structure of KcB-con/37E→G became very similar to that of the SubB-con (Fig. 2). Therefore G→E substitution at aa-37 appears to be critical in the secondary structure of the Korean clade vif protein.

The biological meaning of the amino acid substitutions and change in the secondary structure is yet to be elucidated, since functionally important N-terminal and C-terminal regions in interaction with vif protein's major target APOBEC3G protein (Marin *et al.*, 2003; Luo *et al.*, 2004; Rose *et al.*, 2004) were almost 100% conserved in the Korean clade (Park *et al.*, 2006). At this time it is only conjectured that change in the secondary structure may be related with survival or pathogenesis of the Korean clade HIV-1 in Korean people. Previous study with HIV-1 gp160 gene sequences suggested that the amino acid substitutions such as A344T/S/A, N365E, Q388/H/R/K/P, and R453M/L found in the Korean clade may be responsible for immune escape and host adaptation (Daniels *et al.*, 2003), important factors in viral pathogenesis and evolution, which are influenced by genetic polymorphism in population level. Recent study suggested a possible association of interferon gamma gene polymorphism with HIV-1 infection in the Korean population (Kang *et al.*, 2006). Thus, it would be interesting to investigate in the future correlation between the structure/functions of Korean clade vif protein and polymorphism of human genes involved in HIV-1 pathogenesis.

In conclusion, the *vif* gene of the Korean clade appears to have undergone a positive Darwinian evolution. And the nonsynonymous substitutions at nucleotide level were responsible for the change in the secondary structure of vif protein. Future studies are merited to understand the structure-function relationship of vif protein of the Korean clade and to look into the potential correlation between the substitution pattern of *vif* gene with the Korean-specific polymorphisms occurring in vif-interactant proteins, for example, APOBEC3G.

Acknowledgements

This work was supported by the Regional Research Centers Program of the Ministry of Education & Human Resources Development in Korea.

References

Daniels, R.S., C. Kang, D. Patel, Z. Xiang, N.W. Douglas, N.N. Zheng,

- H.W. Cho, and J.S. Lee. 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, 631-641.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160-174.
- Hughes, A.L. and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex I loci reveals overdominant selection. *Nature* 335, 167-170.
- Kang, M.R., Y.K. Cho, J. Chun, Y.B. Kim, I. Lee, H.J. Lee, S.H. Kim, Y.K. Kim, K. Yoon, J.M. Yang, J.M. Kim, Y.O. Shin, C. Kang, J.S. Lee, K.W. Choi, D.G. Kim, W.M. Fitch, and S. Kim. 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.
- Kang, M.W., C.W. Pyo, S.H. Wie, H.B. Choi, S.Y. Kim, Y.K. Kim, J.S. Lee, and T.G. Kim. 2006. Associations of IFN-gamma polymorphism with HIV-1 infection in the Korean population. *AIDS Res. Hum. Retroviruses* 22, 297-299.
- Kim, E.Y., Y.S. Cho, S.H. Maeng, C. Kang, J.G. Nam, and J.S. Lee. 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, 681-686.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111-120.
- Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK
- Lee, D.H., Y. Yoon, and C.H. Lee. 2003. Phylogenetic analysis of the HIV-1 *nef* gene from Korean isolates. *J. Microbiol.* 41, 232-238.
- Luo, K., B. Liu, Z. Xiao, Y. Yu, X. Yu, R. Gorelick, and X.F. Yu. 2004. Amino-terminal region of the human immunodeficiency virus type 1 nucleocapsid is required for human APOBEC3G packaging. *J. Virol.* 79, 9579-9587.
- Marin, M., K.M. Rose, S.L. Kozak, and D. Kabat. 2003. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat. Med.* 9, 1398-1403.
- Nei, M. and S. Kumar. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, New York, USA
- Park, C.S., M.S. Kim, D.H. Lee, S.D. Lee, S.S. Kim, K.M. Lee, and C.H. Lee. 2006. Molecular phylogenetic analysis of HIV-1 *vif* gene from Korean isolates. *J. Microbiol.* 44, 655-659.
- Rose, K.M., M. Marin, S.L. Kozak, and D. Kabat. 2004. The viral infectivity factor (Vif) of HIV-1 unveiled. *Trends Mol Med.* 10, 291-297.
- Sung, H., B.T. Foley, I.G. Bae, H.S. Chi, and Y.K. Cho. 2001. Phylogenetic analysis of reverse transcriptase in antiretroviral drug-naïve Korean HIV type 1 patients. *AIDS Res. Hum. Retroviruses* 17, 1549-1554.