The Journal of Microbiology, December 2006, p. 655-659
Vol. 44, No. 6

Molecular Phylogenetic Analysis of HIV-1 *vif* Gene from Korean Isolates

Chan Seung Park[1], Mi Sook Kim[1], Sung Duk Lee[2], Sung Soo Kim[3], Keon Myung Lee[3], and Chan-Hee Lee[1,*]

[1]School of Life Sciences, [2]Department of Statistics, and [3]School of Electrical and Computer Engineering, Chungbuk National University, Cheongju, Chungbuk 361-763, Republic of Korea

(Received August 24, 2006 / Accepted September 29, 2006)

Phylogenetic studies of *nef*, *pol*, and *env* gene sequences of HIV-1 isolated from Koreans suggested the presence of a Korean clade in which Korean sequences are clustered to the exclusion of foreign sequences. We attempted to identify and characterize the Korean clade using all *vif* gene sequences isolated from Koreans registered in the NCBI GenBank database (n = 233). Most (77%) of the Korean isolates belonged to the Korean clade as a large subcluster in subtype B, designated the Korean clade subtype B (K_CB). K_CB sequences were relatively homogenous compared to Korean subtype B sequences that did not belong to the K_CB (non-Korean clade subtype B; NK_CB). Comparison of amino acid frequencies of K_CB and NK_CB sequences revealed several positions where the amino acid frequencies were significantly different. These amino acid residues were critical in separating K_CB from NK_CB or from foreign sequences, since substitution of these amino acids in K_CB with the NK_CB amino acids relocated the K_CB sequences to NK_CB, and vice versa. Further analyses of K_CB will help us to understand the origin and evolutionary history of K_CB.

Keywords: HIV-I, *vif*, phylogeny, Korean clade

Human immunodeficiency virus (HIV) is the causative agent for acquired immunodeficiency syndrome (AIDS), first identified in 1981 in North America. Since then AIDS has become one of the leading causes of mortality worldwide. There are two types of HIV, HIV type 1 (HIV-1) and type 2 (HIV-2). While HIV-2 infection is largely confined to Africa, HIV-1 is found all over the world. HIV-1 and HIV-2 are genetically distinct. The HIV-1 genome contains the *vpu* gene, which is not found in the HIV-2 genome. HIV-2 and simian immunodeficiency viruses (SIVs) contain the *vpx* gene which is not found in HIV-1 and in fact HIV-2 is genetically more closely related to SIVs than to HIV-1 (Levy, 1998). HIV-1 is composed of 3 groups, M, N, and O, and group M is further divided into at least 10 subtypes and circulating recombinant forms (Fauquet *et al.*, 2005). Subtype distribution of HIV-1 is somewhat geographical, subtype B being the major subtype in the western hemisphere and subtype C being prevalent in Africa.

In 1985, HIV was first detected in Korea, and as of June 2005 a total of 3,468 people have been identified as being infected with HIV in Korea (press release from the Korea Center for Disease Control and Prevention, July 15, 2005). Molecular phylogenetic studies based on the sequences of the *nef* (Kang *et al.*, 1998; Lee *et al.*, 2003), *pol* (Sung *et al.*, 2001), and *env* genes (Kim *et al.*, 1999a; Kim *et al.*, 1999b) have suggested that subtype B is the major subtype in Korea. Furthermore, the majority of the Korean subtype B isolates seems to be clustered together, forming a "Korean clade" in which foreign isolates are excluded. In this study, we analyzed the nucleotide and amino acid sequences of *vif*, one of the accessory genes of HIV-1 needed for virion assembly and/or viral DNA synthesis (Cullen, 1998; Rose *et al.*, 2004).

Materials and Methods

Obtaining the vif nucleotide and amino acid sequences

A total of 342 *vif* nucleotide sequences registered at the NCBI GenBank (as of March, 2006) were analyzed. These sequences included 233 *vif* nucleotide sequences isolated from Koreans and obtained by searching for "hiv-1 AND vif AND Korea" in the GenBank database (http://www.ncbi.nlm.nih.gov/entrez/). Also included were 44 HIV-1 *vif* sequences of reference strains from the HIV Sequence Database (http://hiv-web.lanl.gov/content

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

/hiv-db/mainpage.html). Foreign *vif* sequences were selected by choosing the most closely matching sequences to the Korean sequences using Blast (http://www.ncbi.nlm.nih.gov/BLAST/). The NCBI accession numbers of the *vif* nucleotide and amino acid sequences analyzed in this study are available in the Supplemental Material.

Construction of phylogenetic trees

The 342 *vif* nucleotide sequences were aligned using the CLUSTALX program (Thomson *et al.*, 1997), and the resulting alignments were confirmed by manual editing. Neighbor-joining (NJ) trees were made with PHYLIP version 3.6a (Felsenstein, 1993) using SEQBOOT, DNADIST, PRODIST, NEIGHBOR, and CONSENSE programs. Protein distance calculations were based on the Jones-Taylor-Thornton protein weight matrix with 1,000 bootstrap (BS) replicates. All other variables were set as default. Group O sequences were used as outgroups. Constructed phylogenetic trees were confirmed using the TREEVIEW program (Page, 1996). Maximum likelihood (ML) trees were generated with DNAML and PROML programs in PHYLIP. Consensus trees were generated by majority rule and viewed with TREEVIEW. The multiple sequence alignments and phylogenetic trees are readily available upon request.

Similarity test

Nucleotide and amino acid sequences were multiple-aligned with CLUSTALX. Aligned sequences were compared with SeqAid (ver 0.91) to get similarity values among sequences belonging to different groups. Mean and Standard Deviation (SD) values for each pair of comparisons were calculated. Statistical analysis was performed when needed with SPSS (ver. 1.0) in order to obtain the statistical significance of the data.

Results and Discussion

Identification of the Korean clade

NJ tree of 342 HIV-1 *vif* nucleotide sequences revealed that the *vif* sequences could be divided into three groups specific for group O, group N and group M (Fig. 1). Group M was further divided into several subtype specific clusters. Subtype B was the largest and accounted for 93.1% of all Korean isolates (217/233). Subtype A and D accounted for 4.3% and 2.6% respectively of the Korean isolates. No other subtypes were found in the *vif* gene phylogeny. Examination of subtype B revealed that a large cluster comprised of Korean isolates only was separate from another cluster in which Korean, foreign and reference sequences coexisted (Fig. 1). The separation was firm and strongly supported by a BS value of 889 out of 1,000 in the NJ tree. This cluster of Korean isolates exclusive of foreign sequences is by definition the Korean clade. The Korean clade numbered 180 and this is 82% of subtype B of Korean isolates and 77% of the total of the Korean isolates.

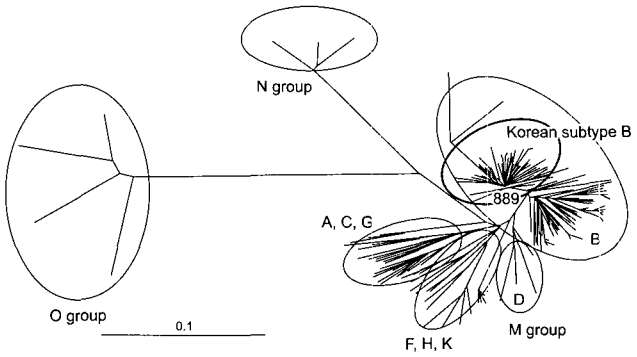

Fig. 1. Unrooted phylogenetic tree of 342 *vif* nucleotide sequences. Sequences obtained from the NCBI GenBank database were aligned with CLUSTALX, and a phylogenetic tree based on neighbor-joining method was constructed using the PHYLIP package. Group O sequences were used as an outgroup, and subtype clusters were indicated within group M. Subtype B cluster was further separated into two subclusters: the Korean clade subtype B free from foreign sequences and other subtype B.

Table 1. *Vif* sequence diversity among subtype B of Korean isolates

Cluster	Mean ± SD (%)	
	Nucleotide	Amino acid
K_CB - K_CB	4.0 ± 1.5	8.5 ± 2.3
NK_CB - NK_CB	6.5 ± 2.1	10.0 ± 3.2
K_CB - NK_CB	7.8 ± 1.2	13.5 ± 1.7

Sequence diversity was calculated. Korean sequences belonging to subtype B were separated into 2 groups: Korean clade subtype B (K_CB, n = 180) and non-Korean clade subtype B (NK_CB, n = 37). The nucleotide sequence diversities among K_CB and NK_CB sequences were 4.0 ± 1.5% (0.0 - 16.8%) and 6.5 ± 2.1% (0.0 - 10.2%) respectively (Table 1). Inter-group diversity between K_CB and NK_CB sequences was 7.8 ± 1.2% (4.8 - 16.8%), which was greater than intra-group diversities. The intra-group amino acid sequence diversity of K_CB (8.5 ± 2.3%, 0.0 - 19.7%) was smaller than that of NK_CB (10.0 ± 3.2, 0.0 - 16.5%), and the inter-group amino acid sequence diversity between K_CB and NK_CB was 13.5 ± 1.7% (8.7 - 19.7%). Thus, the Korean clade appeared to be more homogeneous than other subtype B isolates.

Previous studies with *nef, pol,* and *env* sequences reported the presence of the Korean clade in subtype B with proportions ranging from 76% to 81% of the subtype B of Korean isolates (Sung *et al.*, 2001; Kim

and Cho, 2003; Lee *et al.*, 2003; Sung *et al.*, 2003). Thus, our result showing that the Korean clade accounted for 82% of the subtype B of Korean isolates is compatible with the previous reports. The genetic distances of *vif* sequences among the Korean isolates were similar to those of *nef* reported by Lee *et al.* (2003). Our data also suggest that the Korean clade is relatively homogeneous compared to other subtype B and that it is also distinguishable from other subtype B. Our observations are similar to that of Kim *et al.* (1999a) and Lee *et al.* (2003) in that the diversities of *env* or *nef* nucleotide sequences were lower in the Korean clade than in other subtype B.

Comparison of the nucleotide and amino acid sequences between K_CB and NK_CB

The open reading frame of the *vif* gene is composed of 579 nucleotides or 193 amino acids. At most positions, the nucleotide or amino acid frequencies were the same or almost invariant for both K_CB and NK_CB sequences. At some positions, however, the frequencies were quite different. For example at position 105 of the amino acid sequence, K was found in 91% of the K_CB sequences while Q was found in 100% of the NK_CB sequences. Table 2 lists 11 positions where the frequencies of the amino acids between K_CB and NK_CB appeared highly different. Interestingly, sequence variation between K_CB and NK_CB was found in the middle of the amino acid sequence. In other words, the N-terminal or C-terminal sequences were relatively conserved both in K_CB and in NK_CB.

Vif is a phosphoprotein that is required for HIV-1 replication in nonpermissive cells, but that is irrelevant for viral replication in permissive cells (Rose *et al.*, 2004). The mechanism for nonpermissiveness to HIV-1 replication appears to involve a cellular protein called APOBEC3G, which is expressed in nonpermissive cell lines and absent or undetectable in permissive cell lines (Sheehy *et al.*, 2002). Vif binds to APOBEC3G and neutralizes its function. Mutagenesis studies indicated that Vif contains at least two functionally crucial domains: an N-terminal region (first 11 amino acids) that is important for binding to APOBEC3G (Luo *et al.*, 2004; Rose *et al.*, 2004) and a C-terminal region with a conserved SLQ(Y/F)LAΦΦΦΦ *motif* (Φ being a hydrophobic amino acid), which is unnecessary for binding but is essential for degradation (Marin *et al.*, 2003; Yu *et al.*, 2003). These functionally important N-terminal and C-terminal regions were almost 100% conserved in K_CB, despite rather extensive mutation in some other regions. Thus, the fundamental function of Vif protein seems to be unaltered in the Korean clade.

Signature pattern analysis

Comparison of the amino acid sequences of K_CB and NK_CB identified 11 positions where the frequencies of the amino acids between K_CB and NK_CB appeared highly different as described above. At these positions, the most frequently found amino acids in K_CB are different from those in NK_CB. The collection of the amino acid residues at these positions can be referred to as a signature pattern (Ou *et al.*, 1992). The frequencies of the K_CB and NK_CB signature amino acid residues are shown in Table 3. Signature amino acid residues are presumed to differentiate the Korean clade from other subtype B. This possibility was tested by statistical analysis. Frequencies of the 11 amino acids that define the signature pattern in K_CB and in NK_CB were subjected to a Wilcoxon's rank-sum test in order to determine whether the two populations are different or not. The p-value of K_CB and NK_CB or foreign subtype B (FB) pairs were very low ($p<10^{-5}$), while those of NK_CB and FB pairs were high, $p = 0.40$ for the NK_CB signature and $p = 0.56$ for the K_CB signature (Table 3). Thus, the Korean clade can be differentiated from both the non-Korean clade subtype B

Table 2. Distribution of amino acids at positions where the frequencies of amino acids of K_CB are significantly different from those of NK_CB

Position	NK_CB (%)	K_CB (%)
33	R(42), K(35), G(23)	K(100)
37	G(77), R(16), K(6)	E(46), G(43), K(8), D(3), N(0.6)
63	R(55), K(45)	K(95), E(3), R(2)
78	D(94), E(6)	E(86), D(14)
95	S(55), N(45)	N(80), S(10), I(5), T(2), K, R(1)
105	Q(100)	K(91), Q(5), R(3), T(1)
110	Y(100)	H(92), Y(8)
117	E(77), D(23)	D(90), E(9), N(0.6)
127	H(97), Q(3)	R(74), H(18), N(8)
128	I(68), R(23), K(10)	L(53), V(19), I(16), S(8), R(2), P, T(1)
130	S(100)	R(96), T(2), K, S(1)

Table 3. Frequencies of amino acids that define a signature pattern in the Korean clade

Group	Frequencies (%)										
	Reference signature (Subtype B consensus sequence)										
	33R	37G	63R	78D	95S	105Q	110Y	117E	127H	128I	130S
K_CB	0	43	2	14	10	5	8	9	18	16	1
NK_CB	42	77	55	94	55	100	100	77	97	68	100
FB	47	81	47	93	89	93	86	81	68	67	95
	Wilcoxon's rank sum test: p1 = 2.8×10^{-6}, p2 = 2.8×10^{-6}, p3 = 0.40										
	Korean clade signature (Korean clade consensus sequence)										
	33K	37E	63K	78E	95N	105K	110H	117D	127R	128L	130R
K_CB	100	46	95	86	80	91	92	90	74	53	96
NK_CB	35	0	45	6	45	0	0	23	0	0	0
FB	26	4	46	5	4	0	14	19	23	5	4
	Wilcoxon's rank sum test: p1 = 5.7×10^{-6}, p2 = 2.3×10^{-6}, p3 = 0.56										

FB: Foreign subtype B
p1: K_CB - NK_CB, p2: K_CB - FB, p3: NK_CB - FB

Table 4. List of K_CB and NK_CB sequences used for amino acid substitution experiments

Sequence	Original location	Amino acid substitution (No. amino acids substituted)	New location
AAZ06003	NK_CB	Substitution with K_CB signature 33:R->K, 37:G->E, 78:D->E, 105:Q->K, 110:Y->H, 112:H->R, 128:I->L, 130:S->R (8)	K_CB
		Substitution with K_CB non-signature 31:I->V, 46:N->S, 47:I->T, 51:V->I, 56:Y->H, 66:V->I, 93:G->R, 101:T->D, 109:R->L, 122:K->H, 132:S->K, 158:R->K, 159:T->I, 167:A->R (14)	NK_CB
AAS93515	K_CB	Substitution with NK_CB signature 33:K->R, 63:K->R, 95:N->S, 105:K->Q, 110:H->Y, 117:D->E, 127:R->H, 130:R->S (8)	NK_CB
		Substitution with NK_CB non-signature 22:N->K, 31:V->I, 39:A->V, 41:K->R, 48:N->H, 92:Q->R, 122:N->K, 123:T->A, 124:L->I, 132:K->S, 167:K->A, 182:D->G, 190:H->N (13)	K_CB

and the foreign subtype B on the basis of the signature amino acid residues, with strong statistical support ($p<0.0001$). In contrast, NK_CB sequences do not seem to be different from the foreign sequences. The difference in signature pattern between the K_CB and NK_CB might be explained by different evolutionary processes. In fact, synonymous/nonsynonymous substitution analysis using PAML suggested that the K_CB had undergone relatively positive evolutionary process compared with the NK_CB (data not shown).

If the 11 amino acid residue signature differentiates K_CB from NK_CB, substituting the corresponding amino acids of NK_CB sequences with the K_CB signature amino acids may place them in the Korean clade. Hence a new sequence with the signature 11 amino acid residues was generated from a randomly selected NK_CB sequence (AAZ06003). This new sequence was subjected to phylogenetic analysis and we found that the new sequence was now in the K_CB region (Table 4). If amino acid substitution was made in non-signature residues, the new sequence remained outside the Korean clade despite rather extensive amino acid substitution (Table 4). Similar results were obtained with 3 other NK_CB sequences (AAZ06018, AAZ06020, AAZ06064; data not shown). Substitution experiments were also done with K_CB sequence (AAS93515). If amino acid replacement was made at signature positions, the new sequence relocated to the NK_CB region. On the other hand, there was no relocation if the amino acid substitutions were made at non-signature positions (Table

4). Therefore, it can be concluded that the K_CB signature amino acid residues play a pivotal role in clustering of the Korean clade.

We identified 11 amino acid residues whose frequencies in the K_CB were significantly different from those in the NK_CB and these are the signature amino acid residues. A signature amino acid pattern is a collection of amino acid residues that define a specific group separating from another group (Ou *et al.*, 1992). The importance of the K_CB signature amino acid residues in the function of Vif protein is yet to be investigated. Nonetheless, clustering of the K_CB and separation of K_CB from the rest of subtype B appears strictly dependent on the signature amino acid residues. Substitution of the NK_CB amino acids with the K_CB signature amino acids relocated the new sequences from outside the Korean clade into the Korean clade, while substitution with non-signature amino acid residues did not cause relocation. Previously, Kim and Cho (2003) suggested 6 amino acids that appear at very high frequencies within the Korean clade compared to other subtype B. They also showed that JP<2> strain, one of the NK_CB isolates, could be relocated into the K_CB region by substituting these amino acids.

Acknowledgements

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

References

Cullen, B.R. 1998. HIV-1 auxiliary proteins: making connections in a dying cell. *Cell* 93, 685-692.

Fauquet, C.M., M.A. Mayo, J. Maniloff, U. Desselberger, and L.A. Ball. 2005. Virus Taxonomy VIII. Elsevier Academic Press, San Diego, London, UK

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, USA

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.

Kim, E.Y., Y.S. Cho, S.H. Maeng, C. Kang, J.G. Nam, and J.S. Lee. 1999a. 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.

Kim, Y.B. and Y.K. Cho. 2003. Monophyletic clade of HIV-1 subtype B in Korea: evolutionary pressure or single introduction? *AIDS Res. Hum. Retroviruses* 19, 619-623.

Kim, Y.B., Y.K. Cho, H.J. Lee, C.K. Kim, Y.K. Kim, and J.M. Yang. 1999b. Molecular phylogenetic analysis of human immunodeficiency virus type 1 strains obtained from Korean patients: *env* gene sequences. *AIDS Res. Hum. Retroviruses* 15, 303-307.

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.

Levy, J.A. 1998. HIV and the Pathogenesis of AIDS. 2nd ed. ASM Press, Washington DC, USA

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.

Ou, C.Y., C.A. Ciesielski, G. Myers, C.I. Bandea, C.C. Luo, B.T. Korber, J.I. Mullins, G. Schochetman, R.L. Berkelman, and A.N. Economou. 1992. Molecular epidemiology of HIV transmission in a dental practice. *Science* 256, 1165-1171.

Page, R.D. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12, 357-358.

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.

Sheehy, A.M., N.C. Gaddis, J.D. Choi, and M.H. Malim. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418, 646-650.

Sung, H., B.T. Foley, I.G. Bae, H.S. Chi, and Y.K. Cho. 2001. Phylogenetic analysis of reverse transcriptase in antiretroviral drug-naive Korean HIV type 1 patients. *AIDS Res. Hum. Retroviruses* 17, 1549-1554.

Sung, H., B.T. Foley, S.H. Ahn, Y.B. Kim, J.D. Chae, Y.O. Shin, H.I. Kang, and Y.K. Cho. 2003. Natural polymorphism o protease in protease inhibitor-naïve HIV-1 patients in Korea: A novel L63M in subtype B. *AIDS Res. Hum. Retroviruses* 19, 525-530.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876-4882.

Yu, X., Y. Yu, B. Liu, K. Luo, W. Kong, P. Mao, and X.F. Yu. 2003. Induction of APOBEC3G Ubiquitination and Degradation by an HIV-1 Vif-Cul5-SCF Complex. *Science* 302, 1056-1060.