The Journal of Microbiology, February 2008, p. 88-94
DOI 10.1007/s12275-007-0156-x

Vol. 46, No. 1

Characterization and Signature Pattern Analysis of Korean Clade HIV-1 Using *nef* Gene Sequences

Chan Seung Park[1], Dong Hun Lee[1], Keon Myung Lee[2], and Chan-Hee Lee[1]*

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

(Received September 5, 2007 / Accepted January 3, 2008)

Phylogenetic studies of the HIV-1 gene sequences isolated from Korean patients have suggested that most of Korean isolates belong to the subtype B strain. This study aims to characterize the Korean clade by molecular phylogenetic analysis using all of the Korean *nef* gene sequences registered in the NCBI GenBank (N=422), in addition to 41 reference strains and 94 foreign isolates. Through phylogenetic analyses, we verified that most of the Korean isolates belonged to the subtype B, where 78.8% are clustered exclusively of foreign isolates. This cluster has been named the Korean clade subtype B (K_CB) in order to distinguish it from other subtype B clusters. Genetic distance analysis suggested that the K_CB cluster was more homogeneous and clearly distinctive from the non-Korean clade subtype B (NK_CB). Comparison of consensus amino acid sequences from K_CB and NK_CB revealed that characteristic K_CB signature amino acid patterns composed of 11 amino acid residues, whose frequencies in the K_CB were significantly higher than in the NK_CB. The K_CB signature amino acid residues were critical in identifying K_CB from NK_CB, since substitution of the NK_CB sequences with K_CB signature amino acids relocated them to the Koran clade, and vice versa.

***Keywords*:** HIV-1, *nef*, signature

In 1985, the first official case of HIV-1 infection was recorded in Korea. At last official count, the total number of HIV infections were reported to be 4,956 individuals (Press release from Korea Center for Disease Control and Prevention, July 19, 2007). Although this number is much smaller than those in other developed countries, Korean health authorities are concerned that the infection rate has been increasing rapidly, from less than 200 per year in 1999 to more than 750 by 2006. Despite the relatively small number of HIV-infected people in Korea, an interesting feature of the so-called "Korean clade" has been reported. Korean clade was first proposed based on the molecular phylogenetic studies of *nef* gene sequences of HIV-1 isolated from Korean patients (Kang *et al*., 1998). Korean clade by definition is an exclusive cluster of Korean sequences where no other foreign sequences are included. Subsequent studies with *env* (Kim *et al*., 1999a, 1999b) and *pol* (Sung *et al*., 2001) gene sequences of Korean HIV-1 isolates supported the presence of the Korean clade, followed by confirmation from more comprehensive studies (Lee *et al*., 2003; Park *et al*., 2006a).

The Korean clade is a sub-cluster of subtype B, of which 80% are composed of Korean HIV-1 isolates. Since the majority (more than 80%) of the HIV-1 isolated from Korean patients belong to subtype B, almost 2/3 of the Korean HIV-1 isolates form the Korean clade.

Unlike the heterogeneous nature of some Asian ethnic groups (e.g., in South-east Asia), Koreans are ethnically highly pure, and the presence of a large and undisputed clade specific for ethnicity is unprecedented. Thus, characterization of the Korean clade is merited and will help to establish a regimen to control localized HIV infection. In this study, using all the Korean *nef* sequences registered in GenBank database, we confirmed the presence of the Korean clade, and proceeded to identify and characterize its defining signature amino acid residues.

Materials and Methods

Obtaining the *nef* sequences

In total, 557 *nef* nucleotide sequences registered at NCBI GenBank were analyzed. These sequences included all of the nucleotide sequences isolated from Koreans (n=422) and obtained by searching for "hiv-1 AND nef AND Korea" in the GenBank database. Also included were 41 HIV-1 *nef* sequences of the reference strains from HIV SEQUENCE DATABASE (http://hiv-web.lanl.gov/content/hiv-db/mainpage.html). Foreign *nef* sequences were selected by choosing the most closely matching sequences with the Korean sequences using the Basic Local Alignment Search Tool (BLAST). Since many foreign sequences were selected redundantly by the BLAST program, we handpicked 94 unique sequences from the selection. The list and NCBI accession numbers of the *nef* nucleotide sequences analyzed in this study are listed in Table 1.

Construction of phylogenetic trees

Phylogenetic trees were constructed as described in Park *et al*. (2006a). Briefly, the *nef* nucleotide sequences were aligned

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

Table 1. The list and GenBank accession numbers of the sequences used in this study

Sequence	Accession number
Korean isolates	AF063908~AF063933, AF224507, AF238268~AF238270, AF238273~AF238278, AF272003, AF272005, AF286239, AF462693~AF462709, AF462712~AF462733, AF462736, AF462737, AF462739~AF462742, AF462744~AF462745, AF462746, AF462748~AF462751, AF462753~AF462771, AF462773, AF462774, AF462776~AF462780, AF462782, AF462783, AF462785, AF462786, AF462788~AF462793, AY121440~AY121455, AY121457~AY121463, AY121465, AY121467~AY121480, AY221648, AY221649, AY221652~AY221660, AY221662~AY221676, AY221681, AY221682, AY221684, AY221685, AY221687~AY221689, AY221692~AY221700, AY221702~AY221711, AY221713, AY221714, AY221716~AY221719, AY260770~AY260780, AY260782~AY260792, AY260794, AY260796~AY260798, AY260803~AY260812, AY363309~AY363312, AY363314~AY363318, AY363320, AY363324~AY363330, AY363333, AY363334, AY363337, AY363338, AY363340, AY363341, AY363343, AY363345~AY363361, AY363363, AY363365, AY363367~AY363369, AY584754~AY584760, AY584762, AY584764, AY584766~AY584771, AY584773~AY584800, AY584802~AY584804, AY584806~AY584808, AY839827, AY899339~AY899346, AY899350~AY899362, AY899365~AY899379, AY899381~AY899385, AY899387, AY899388~AY899391, AY899394, AY899397~AY899403, Z98019~Z98034
Foreign isolates	AB032741, AF063224, AF069673, AF082376, AF107770, AF120802, AF120811, AF120887, AF120890, AF120907, AF120912, AF129366, AF129367, AF203120, AF203152, AF203159, AF203198, AF219714, AF219819, AF252913, AF252920, AF252925, AF259954, AF286224, AF361873, AF425882, AF425885, AF457061, AF457070, AF457081, AF457084, AF484478, AF484482, AF484503, AF484509, AF490974, AF538642, AF538644, AF538646, AF538668, AF538673, AJ232957, AJ232974, AJ232983, AJ232984, AJ251056, AJ291720, AJ508597, AY064814, AY093604, AY116802, AY125894, AY167123, AY265062, AY265069, AY265084, AY265097, AY265111, AY271690, AY314060, AY444304, AY444312, AY536901, K02011, K02013, K02083, M19921, M74426, U16920, U16940, U16950, U24455, U24460, U24462, U26107, U26942, U34603, U34604, U44443, U44453, U44460, U48905, U48921, U48927, U48931, U48932, U48933, U51188, U51189, U54771, U80225, Z11530 , NC_001802, U16915
References	AF004885, AF005494, AF005496, AF061641, AF061642, AF067155, AF069670, AF075703, AF077336, AF082394, AF082395, AF084936, AF110967, AF190127, AF190128, AF286237, AF286238, AJ006022, AJ249235, AJ249236, AJ249237, AJ249238, AJ249239, AJ271370, AJ302646, AJ302647, K03454, K03455, L20571, L20587, M17451, M27323, M62320, U21135, U46016, U51190, U52953, U63632, U88822, U88824, U88826

using CLUSTAL X, and the resulting alignments were confirmed by manual editing. Neighbor-joining (NJ), maximum-parsimony (MP), and maximum likelihood (ML) trees were constructed with PHYLIP version 3.6b using SEQBOOT, DNADIST, PROTDIST, NEIGHBOR, PROTPARS, DNAPARS, CONSENSE, DNAML, and PROML programs (Felsenstein, 2004). Evolutionary distance calculations for NJ trees were based on the Jukes-Cantor method and Jones-Taylor-Thornton protein weight matrix (Jukes and Cantor, 1969; Jones *et al*., 1992). All other variables were set as default value according to individual programs. The topology of phylogenetic trees was evaluated by a bootstrap resampling method based on 1,000 replicates. Constructed phylogenetic trees were viewed using the TreeView program.

Genetic distance analysis

Nucleotide and amino acid sequences of the *nef* gene obtained from the NCBI GenBank were multiple-aligned with CLUSTAL X. Aligned sequences were compared with SeqAid program developed by the second author to ascertain genetic distances among sequences belonging to different groups according to Jukes and Cantor's method (Jukes and Cantor, 1969).

Subsequently, mean and standard deviation values for each comparative pairs were calculated. Statistical analysis was performed with SPSS (ver. 10, SPSS Inc., USA) in order to ascertain statistical significance (i.e., $P<0.05$).

Generation of consensus sequences

Consensus sequences were obtained using the SeqAid program. Each base in the consensus sequence corresponds to the base most frequently occurring at that position in the aligned sequences.

Paired t-test

The paired t-test is used to determine the significance of a difference in paired measurements, where the null hypothesis is that there is no difference in the paired measurements. This test was performed with SigmaPlot program (Ver. 8) in order to examine statistical significance.

Wilcoxon's rank-sum test

When there is no assurance of normal distribution of the population and sample size is less than 30, Wilcoxon's rank-sum test can be used instead of a two-sample t-test to determine whether the two populations are different or not. This requires independent random samples of sizes n1 and n2. The test is very simple and consists of combining the two samples into one sample of size n1+n2, sorting the result, assigning ranks to the sorted values (giving the average rank to any 'tied' observations), and then letting T be the sum of the ranks for the observations in the first sample. If the two populations have the same distribution then the sum of the ranks of the first sample and those in the second sample should be close to the same value. Stataquest returns a p value for the null hypothesis to ascertain if the two distributions are the same. This test was performed with SPSS (Ver. 10) in order to gauge statistical significance.

Results

Identification of the Korean clade

Phylogenetic trees of 557 HIV-1 *nef* nucleotide sequences from 422 Korean isolates, 41 HIV-1 reference strains, and 94 foreign isolates were constructed using maximum-likelihood, maximum-parsimony and neighbor-joining methods. The *nef* nucleotide sequences from subtype O and subtype N reference strains were used as outgroups. All three methods generated identical tree topology as reported previously by Lee *et al*. (2003). Analyses of the phylogenetic trees indicated that subtype B was the biggest group and accounted for 79.4% of all Korean isolates (data not shown). Examination of the subtype B revealed that a big cluster comprised of only Korean isolates separable from the other subtype B cluster where reference, Korean and foreign sequences co-existed. This cluster of Korean isolates exclusive of foreign sequences is by definition the Korean clade. The size of the Korean clade is 264, which equates to 78.8% of subtype B isolates or 62.6% of total Korean isolates.

Generation and comparison of consensus sequences

In order to identify distinguishable characteristics of the Korean clade from the other subtype B sequences, consensus sequences were deduced. Consensus nucleotide and amino acid sequences were established using the SeqAid program from 264 *nef* sequences belonging to the Korean clade (K_CB-con) and 71 Korean *nef* sequences belonging to subtype B, but not to Korean clade (NK_CB-con). The amino acid sequences of K_CB-con and NK_CB-con are shown in Fig. 1 together with the consensus sequence of HIV-1 subtype B *nef* (SubB-con) obtained from HIV SEQUENCE

```
                  *  *              *           **    *    *
Sub B-con   MGGKWSKRSVVGWPTVRERMRRAEPAADGVGAVSRDLEKHGAITSSNTAA
NKcB-con    ----------A---A-----------------------------------
KcB-Con     ----------P--N-I------T----E----------QR----T----S
                                                             50

Sub B-con   NNADCAWLEAQEEEEVGFPVRPQVPLRPMTYKGALDLSHFLKEKGGLEGL
NKcB-con    ------------------------------------V-------------
KcB-Con     ---A-----------------------------------------------
                                                            100

Sub B-con   IYSQKRQDILDLWVYHTQGYFPDWQNYTPGPGIRYPLTFGWCFKLVPVEP
NKcB-con    ------------------------------------------------
KcB-Con     ---------------------------------T-F--------------
                                                            150
                  **          *                            *
Sub B-con   EKVEEANEGENNSLLHPMSLHGMDDPEREVLVWKFDSRLAFHHMARELHP
NKcB-con    ------------C------Q----------------------V------
KcB-Con     ------TV----C-----N---------------------------K--

                                                            200
Sub B-con   EYYKDC
NKcB-con    ----N-
KcB-Con     ------
```

Fig. 1. Comparison of the consensus amino acid sequences. Consensus sequence of the Korean clade subtype B (K_CB-con) was generated using SeqAid (ver. 0.91) program from 264 Korean sequences belonging to the K_CB. Consensus sequence of the non-Korean clade subtype B (NK_CB-con) was generated from 71 Korean *nef* subtype B sequences that did not belong to the Korean clade. Consensus sequence of subtype B (SubB-con) was obtained from Las Alamos HIV SEQUENCE DATABASE. All the consensus sequences were aligned to the *nef* amino acid sequence of HIV-1 strain HXB2. Asterisks (*) indicate the location of the signature amino acid residues.

Table 2. Distance of Korean isolates from the consensus sequences

Consensus	Distance, Mean±SD (min~max)	
	K_CB[a]	NK_CB[b]
	Nucleotide	
K_CB-con	3.8±1.3 (1.5~10.2)%	9.9±1.1 (7.6~12.4)%
NK_CB-con	7.9±1.1 (5.9~14.5)%	6.3±1.2 (4.1~9.9)%
SubB-con	7.7±1.1 (5.5~14.0)%	6.4±1.5 (3.3~9.0)%
	Amino acid	
K_CB-con	7.4±2.1 (2.6~14.3)%	16.8±2.1 (12.0~22.8)%
NK_CB-con	14.7±1.6 (11.0~20.4)%	11.3±1.7 (8.9~17.5)%
SubB-con	13.8±1.7 (9.4~18.8)%	12.1±2.0 (8.8~17.8)%

[a] K_CB, Korean isolates belonging to the Korean clade
[b] NK_CB, Korean isolates that did not belong to the Korean clade

DATABASE and aligned to the *nef* sequence of HIV-1 strain HXB2. Comparisons of the consensus amino acid sequences revealed that there were seven differences between SubB-con and NK_CB-con; 11:V/A, 15:T/A, 85:D/V, 163:S/C, 170:L/Q, 194:M/V, and 205:D/N (Fig. 1), while there were 17 differences between SubB-con and K_CB-con: 11:V/P, 14:P/N, 16:V/I, 23:A/T, 28:D/E, 39:K/Q, 40:H/R, 45:S/T, 50:A/S, 54:D/A, 133:I/T, 135:Y/F, 157:N/T, 158:E/V, 163:S/C, 169:S/N, and 198:L/K. Of these, only 163:S/C substitution was common to NK_CB-con and K_CB-con. There were 21 differences between NK_CB-con and K_CB-con.

Distances of each sequence from consensus sequences were calculated. The nucleotide sequence distances of K_CB from K_CB-con were estimated to be 3.8±1.3%, approximately half of those from NK_CB-con or SubB-con (Table 2). The sequence of SubB-con and NK_CB-con differed only 1.8% in nucleotide (11/621 nt) and 3.4% in amino acid (7/206 aa), and we expected the distance from SubB-con to K_CB or NK_CB sequences to be similar to that from NK_CB-con. The data shown in Table 1 supported this hypothesis. The distances of K_CB from SubB-con or NK_CB were 7.7±1.1% and 7.9±1.1%, respectively. Thus, NK_CB sequences were, in general, almost indistinguishable from foreign subtype B sequences, while K_CB sequences were clearly distinguishable. The data obtained with amino acid sequences were almost paralleled with the nucleotide data, except that the distances were greater in amino acid than in nucleotide sequences (Table 2).

Signature pattern analysis of the Korean clade subtype B

Comparison of SubB-con, K_CB-con and NK_CB-con identified 16 amino acid residues unique to K_CB-con. Of these, we found 11 amino acid residues where the frequencies of K_CB-con amino acids were very low in NK_CB, and the frequencies of NK_CB-con amino acids were very low in K_CB. For example, at position 50, the frequency of A (NK_CB-con) was 94% in NK_CB and 0% in K_CB, while the frequency of S (K_CB-con) was 0% in NK_CB and 89% in K_CB. These were defined to be the signature amino acid residues. Not all isolates in the Korean clade contained the K_CB signature amino acid residues. For example, only 58% of the isolates had N at position 14 and 71% had V at position 158. Overall, most of the K_CB signature amino acid residues occurred more than 80% in K_CB. In contrast, the frequency of the K_CB signature amino acid residues was at most 23% in NK_CB (see Table 3). Table 3 also presents amino acid residues at corresponding positions found in SubB-con, most frequently found in all subtype B isolates worldwide. These amino acid residues, however, occurred very infrequently in K_CB, ranging from 10% E at position 158 to 0% V, P, K or A at position 11, 14, 39 or 50, respectively. In contrast, these amino acids were the most frequently found residues in NK_CB (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 amino acids that define signature patterns in K_CB and NK_CB were subjected to a paired t-test. Analysis of the K_CB signature amino acid residues revealed that the p value of K_CB and NK_CB or foreign subtype B (FB) pair was statistically significant ($P<10^{-9}$), while the *p* value of NK_CB and FB pair was insignificant (P=0.157) (Table 3). Similar results were obtained when the subtype B signature amino acid residues were analyzed.

The results of the Wilcoxon's rank-sum test using the frequencies of the signature amino acid resides are shown in Table 3. The p value of K_CB and NK_CB or FB pairs were very low ($p=6.7\sim7.0\times10^{-5}$), while those of NK_CB and FB pairs were high (p=0.178) for reference set and p=0.767 for the Korean clade set (Table 3). Thus, Korean clade can be differentiated from non-Korean clade subtype B or foreign subtype B on the basis of the signature amino acid residues, with strong statistical support.

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

Group[a]	Frequencies (%)										
	Reference signature (Subtype B-con)										
	11V	14P	28D	39K	40H	45S	50A	157N	158E	169S	198L
K_CB	0.0	0.0	0.8	0.0	6.6	1.6	0.0	0.4	10	0.4	8.2
NK_CB	19	70	58	67	68	90	94	77	64	88	93
FB	28	85	74	83	96	98	81	91	68	94	85
	Paired t-test: p1[b]=8.5×10^{-7}, p2=1.3×10^{-7}, p3 = 0.032										
	Wilcoxon's rank sum test: p1=6.7×10^{-5}, p2=6.7×10^{-5}, p3=0.178										
	Korean clade signature (K_CB-con)										
	11P	14N	28E	39Q	40R	45T	50S	157T	158V	169N	198K
K_CB	100	58	96	96	93	81	89	97	71	98	87
NK_CB	8.7	0.0	20	2.9	4.4	1.5	0.0	23	1.5	8.7	10
FB	6.4	0.0	11	6.4	4.3	2.1	2.1	8.5	2.3	4.3	6.4
	Paired t-test: p1=3.0×10^{-10}, p2=2.0×10^{-10}, p3=0.157										
	Wilcoxon's rank sum test: p1=7.0×10^{-5}, p2=6.9×10^{-5}, p3=0.767										

[a] K_CB, Korean isolates belonging to the Korean clade; NK_CB, Korean isolates that did not belong to the Korean clade; FB, foreign isolates
[b] p1, p value of K_CB/NK_CB pair; p2, p value of K_CB/FB pair; p3, p value of NK_CB/FB pair

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
		Substitution with Korean clade signature	
AF063922	NK_CB	I11P, P14N, D28E, K39Q, H40R, S45T, A50S, N157T, K158V, S169N, L198K (11)	K_CB
AF462788	NK_CB	P11P, P14N, D28E, K39Q, L40R, S45T, P50S, N157T, E158V, S169N, L198K (11)	K_CB
		Substitution with Korean clade non-signature	
AF063922	NK_CB	S8R, A15T, V16I, A23T, R27A, A71R, A83G, V85L, R93E, H102Y, R105K, I114V, V133T, Y135F, Y143F, D151E, T162N, S163C, V168M, G177E, E182V, R183W, R184K, V194M, S206C (25)	NK_CB
AF462788	NK_CB	A10V, S15T, V16I, D24E, P32A, D33V, Q35R, P42A, L43I, T46S, I47N, T51N, M52N, D54A, D63E, Q71R, A83G, V85L, V101I, H102Y, Q105K, I114V, K129P, I168M, Q170L, R172G, V194M, F203Y, N205D (29)	NK_CB
		Substitution with Reference signature	
AY221693	K_CB	P11A, D14P, E28D, R39K, R40H, T45S, S50A, T157N, V158E, N169S, K198L (11)	NK_CB
AF462766	K_CB	P11A, D14P, E28D, Q39K, H40H, T45S, E50A, T157N, V158E, N169S, L198L (11)	NK_CB
		Substitution with Reference non-signature	
AY221693	K_CB	R9S, T15A, I16V, T23A, A33V, S60A, H73Q, L85V, H86D, D98E, H102Y, Q105K, I114V, V133I, F135Y, K178R, R184K (17)	K_CB
AF462766	K_CB	G5W, C8R, G9S, T15A, I16V, A33V, A54D, L85V, N86D, N88S, V101I, F102Y, N116H, D151E, A168M, M170Q, R184K, K194V (18)	K_CB

Substitution of the signature residues

If the signature amino acid residues differentiate K_CB from NK_CB, substituting the corresponding amino acids of NK_CB or subtype B reference sequences with the signature amino acids may place them in the Korean clade. Hence new sequences with the signature 11 amino acid residues replaced were generated from 2 randomly selected NK_CB sequences (AF063922, AF462788; Table 4). These new sequences were subjected to phylogenetic analysis together with the subtype B sequences from Korean isolates and subtype B reference sequences using subtype D reference sequences as an outgroup. The new sequences generated from subtype B were re-located in the Korean clade, while their original sequences were located outside the Korean clade (Table 4). If amino acids were selected from non-signature residues and used to generate new sequences by substituting with corresponding amino acids of K_CB-con (25 or 29 amino acid residues, Table 4), NK_CB sequences still remained within the non-Korean clade despite rather extensive amino acid substitution (Table 4). Similar results were obtained with 3 subtype B reference sequences (HXB2, RF, TRFL). Parallel experiments of amino acid substitution were done with K_CB sequences (AY221693, AF462766). If the K_CB sequences at signature position were replaced with subB-con, the new sequences were relocated to NK_CB. On the other hand, new sequences remained within the Korean clade if amino acid substitutions were made at non-signature positions (Table 4). Therefore, it could be concluded that the K_CB signature amino acid residues were responsible for clustering of the Korean clade.

Discussion

In this study, we confirmed the presence of the Korean clade of HIV-1, suggested first by Kang *et al*. (1998). Although they and other investigators (Kim *et al*., 1999a, 1999b; Sung *et al*., 2001) have identified the Korean clade using *nef*, *pol*, or *env* sequences of HIV-1 isolated from Korean patients, their data suffer from the use of limited number of sequences. We used all the *nef* sequences from Korean HIV-1 isolates available within the NCBI GenBank as of May 15, 2005. A further 376 more Korean *nef* sequences were added to the GenBank database by October 2006. Focusing on this newly added data set, we did not find significantly different results from the May 2005 data set. We also included foreign sequences by choosing the most similar sequences with each of the Korean *nef* sequences using the BLAST program. The inclusion of a number of foreign sequences is critical for conclusive proof of Korean clade since the use of Korean sequences with a limited number of reference or foreign sequences might force some Korean sequences to form a cluster free of reference or foreign sequences by chance.

Comparison of the sequences from the K_CB with those of NK_CB or foreign subtype B revealed that the frequencies of amino acids at several positions were significantly different. Despite the changes in amino acid residues at several positions, residue at N terminal end (01-10) and middle (55-155) did not change significantly in the K_CB. The sequences at N-terminal myristoylation site (01-07: MGGKWSK), (PxxP)3 motif (69-78: PVxPQVPLRP), PKC recognition site (77-82: RPMTYK), beta-turn motif (130-132: GPG), and the second PxxP motif (147-150: PVEP) were well conserved in K_CB, and these sites are known to be functionally important (Harris and Neil, 1994; Saksela *et al*., 1995; Asamitsu *et al*., 1999; Fackler *et al*., 1999; Geyer *et al*., 1999). Thus, the fundamental function of *nef* protein of the K_CB appears to be unaltered.

The presence of the Korean clade is highly unique since, to our knowledge, no other national or ethnic clade has been

clearly reported elsewhere. There have been reports on a clustering of HIV-1 according to countries; Spain (Casado *et al*., 2000), Trinidad Tobago (Cleghorn *et al*., 2000), India (Shankarappa *et al*., 2001), and Congo (Taniguchi *et al*., 2002). There are, however, two major drawbacks in proposing the existence of a national clade in these reports. First, is the use of a limited number of isolates, which may result in accidental clustering of some sequences as discussed here. Second and more importantly, is the lack of pure clustering of the isolates from a specific country. For example, several foreign sequences are included in the proposed Indian clade (Shankarappa *et al*., 2001) or Spainish clade (Casado *et al*., 2000). Thus, the Korean clade appears to be the only genuine national or ethnic clade consisting of the majority of HIV-1 isolates.

This poses the following obvious question in relation to the origin of this unprecedented unique Korean clade. The most probable answer is the single introduction of a common ancestor in Korea as suggested by other Korean investigators (Kim *et al*., 1999b; Sung *et al*., 2001; Kim and Cho, 2003). The basis for this conclusion includes the lack of distinct sub-clusters, that does not allow easy cause-and-effect relation according to the risk group among the infected persons within the Korean clade. The star-like topology of the phylogenetic tree has been interpreted as a result of a founder effect and used as evidence supporting the single introduction of HIV-1 in Spain (Casado *et al*., 2000) and India (Shankarappa *et al*., 2001). Korean clade in phylogenetic tree portrays a somewhat star-like topology. Thus, the notion that HIV-1 in the Korean clade might have descended from a single origin is a distinct possibility. A star-like topology of the phylogenetic tree are, however, often seen in HIV-1 subtype B (Anderson *et al*., 2001). The separation of the K_CB from the NK_CB may be viewed as a speciation event. Regardless of its origin, the K_CB appears to have undergone a different evolutionary pathway from the subtype B, and the hallmark of the evolutionary pathway seems to be the K_CB-specific signature amino acid pattern.

The potential application of the consensus sequence in relation to the Korean clade needs to be emphasized. Currently, HIV vaccines are derived from isolates, with the hope that they will be sufficiently cross-reactive to most circulating viruses. However, considering the huge diversity of HIV-1 isolates, country-specific, consensus or ancestral sequences could be more useful in vaccine design to minimize the genetic differences between vaccine strains and contemporary isolates (Gaschen *et al*., 2002). Subtype C consensus approach for vaccine development has been proposed to overcome the high genetic diversity of HIV-1 subtype C (Novitsky *et al*., 2002). In this regard, if we were to develop a HIV vaccine customized for Koreans, Korean clade consensus sequence, rather than subtype B consensus sequence, would be the most appropriate avenue to direct research effort, since the majority (about 2/3 of all and 80% of subtype B) of the Korean HIV isolates belong to the Korean clade, as previously reported by the authors of this study (Park *et al*., 2006b).

Acknowledgements

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (The Regional Research Universities Program/Chungbuk BIT Research-Oriented University Consortium).

References

Anderson, J.P., A.G. Rodrigo, G.H. Learn, Y. Wang, H. Weinstock, M.L. Kalish, K.E. Robbins, L. Hood, and J.I. Mullins. 2001. Substitution model of sequence evolution for the human immunodeficiency virus type 1 subtype B gp120 gene over the C2-V5 region. *J. Mol. Evol.* 53, 55-62.

Asamitsu, K., T. Morishima, H. Tsuchie, T. Kurimura, and T. Okamoto. 1999. Conservation of the central proline-rich (PxxP) motifs of human immunodeficiency virus type 1 Nef protein during the disease progression in two hemophiliac patients. *FEBS Lett.* 459, 399-404.

Casado, C., I. Urtasun, M.V. Martin-Walther, S. Garcia, C. Rodriguez, J. Del Romero, and C. Lopez-Galindez. 2000. Genetic analysis of HIV-1 samples from Spain. *J. Acquir. Immune Defic. Syndr.* 23, 68-74.

Cleghorn, F.R., N. Jack, J.K. Carr, J. Edwards, B. Mahabir, A. Sill, C.B McDanal, S.M. Connolly, D. Goodman, R.Q. Bennetts, T.R. O'Brien, K.J. Weinhold, C. Bartholomew, W.A. Blattner, and M.L. Greenberg. 2000. A distinctive clade B HIV type 1 is heterosexually transmitted in Trinidad and Tobago. *Proc. Natl. Acad. Sci. USA* 97, 10532-10537.

Fackler, O.T., W. Luo, M. Geyer, A.S. Alberts, and B.M. Peterlin. 1999. Activation of Vav by Nef induces cytoskeletal rearrangements and downstream effector functions. *Mol. Cell* 3, 729-739.

Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package) version 3.6b. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, USA.

Gaschen, B., J. Taylor, K. Yusim, B. Foley, F. Gao, D. Lang, V. Novitsky, B. Haynes, B.H. Hahn, T. Bhattacharya, and B. Korber. 2002. Diversity considerations in HIV-1 vaccine selection. *Science* 296, 2354-2360.

Geyer, M., C.E. Munte, J. Schorr, R. Kellner, and H.R. Kalbitzer. 1999. Structure of the anchor-domain of myristoylated and non-myristoylated HIV-1 Nef protein. *J. Mol. Biol.* 289, 123-138.

Harris, M.P. and J.C. Neil. 1994. Myristoylation-dependent binding of HIV-1 Nef to CD4. *J. Mol. Biol.* 241, 136-142.

Jones, D.T., W.R. Taylor, and J.M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *Comp. Appl. Biosci. (CABIOS)* 8, 275-282.

Jukes, T.H. and C.R. Cantor. 1969. Evolution of protein molecules, p. 21-132. *In* H.N. Munro (ed.). Mammalian protein metabolism, Academic Press, New York, NY, 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, Y.B., Y.K. Cho, H.J. Lee, C.K. Kim, Y.K. Kim, and J.M. Yang. 1999a. Molecular phylogenetic analysis of human immunodeficiency virus type 1 strains obtained from Korean patients: *env* gene sequences. *AIDS Res. Hum. Retroviruses* 15, 303-307.

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

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.

Novitsky, V., U.R. Smith, P. Gilbert, M.F. McLane, P. Chigwedere, C. Williamson, T. Ndung'u, I. Klein, S.Y. Chang, T. Peter, I. Thior, B.T. Foley, S. Gaolekwe, N. Rybak, S. Gaseitsiwe, F. Vannberg, R. Marlink, T.H. Lee, and M. Essex. 2002. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? *J. Virol.* 76, 5435-5451.

Park, C.S., M.S. Kim, S.D. Lee, S.S. Kim, K.M. Lee, and C.H. Lee. 2006a. Molecular phylogenetic analysis of HIV-1 *vif* gene from Korean isolates. *J. Microbiol.* 44, 655-659.

Park, C.S., M.S. Kim, H.A. Yi, K.M. Lee, S.S. Kim, S.D. Lee, and C.H. Lee. 2006b. Comparison of consequence sequences and ancestor sequences of HIV-1 *nef* and *vif* genes for developement of Korean-specific vaccine. *J. Bacteriol. Virol.* 36, 287-292.

Saksela, K., G. Cheng, and D. Baltimore. 1995. Proline-rich (PxxP) motifs in HIV-1 Nef bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of Nef^+ viruses but not for down-regulation of CD4. *EMBO J.* 14, 484-491.

Shankarappa, R., R. Chatterjee, G.H. Learn, D. Neogi, M. Ding, P. Roy, A. Ghosh, L. Kingsley, L. Harrison, J.I. Mullins, and P. Gupta. 2001. Human immunodeficiency virus type 1 *env* sequences from Calcutta in eastern India: identification of features that distinguish subtype C sequences in India from other subtype C sequences. *J. Virol*. 75, 10479-10487.

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.

Taniguchi, Y., J. Takehisa, B. Bikandou, I. Mboudjeka, M.Y. N′Doundou-N′Kodia, Obengui, M. M′Pandi, P. M′Pele, Y. Harada, E. Ido, M. Hayami, H. Ichimura, and H.J. Parra. 2002. Genetic subtypes of HIV type 1 based on the *vpu*/*env* sequences in the Republic of Congo. *AIDS Res. Hum. Retroviruses* 18, 79-83.