Phylogenetic Analysis of the HIV-1 *nef* Gene from Korean Isolates

Dong-Hun Lee[1], Yeup Yoon[2], and Chan-Hee Lee[1,*]

[1]*Division of Life Sciences and Biotechnology Research Institute, Chungbuk National University, Cheongju 361-763, Korea*
[2]*Mogam Biotechnology Research Institute, Yongin 449-913, Korea*

Previous phylogenetic studies on human immunodeficiency virus type 1 (HIV-1) isolated from Korean patients suggest that the major subtype of Korean isolate is subtype B. In this subtype, some of the Korean isolates seem to be clustered exclusively of foreign isolates. Presence of this so-called "Korean clade" among Korean isolates is unique but needs verification since the number of Korean isolates used in previous studies was limited. This study aimed to identify the presence of the "Korean clade" by molecular phylogenetic analysis using all the Korean *nef* gene sequences registered in the NCBI GenBank (N=243) together with 32 reference strains and 77 foreign isolates. Extensive analysis of the *nef* gene nucleotide sequences by neighbor-joining method revealed the following. Most (83.1%) of the Korean isolates belonged to subtype B, and 81.2% of subtype B were clustered together and excluded foreign isolates (bootstrap value=91.9%). Within Korean subtype B cluster, no characteristic subcluster formation was evident since the bootstrap values for the subcluster were very low. Due to limited information, the phylogenetic analysis failed to identify the epidemiological linkage among specific groups such as homosexuals and hemophiliacs within the Korean subtype B cluster. Detailed analysis and epidemiological information are needed to clarify the origin and significance of the Korean subtype B cluster.

Key words: HIV-1, phylogeny, *nef*

Since the first identification of AIDS (acquired immunodeficiency syndrome) in 1981, AIDS has become the leading cause of mortality around the world. Human immunodeficiency virus (HIV) is the causative agent and destroys the immune system by infecting CD4+ helper T cells. In Korea, the first identification of AIDS was made in 1985, and it was reported that 2,122 individuals have been infected with HIV as of March 2003. Of these HIV-infected people, 347 developed AIDS (Press release from Korean National Institute of Health, April 15, 2003). Although this figure is lower than that in the United States of America, Europe, Southeastern Asia or Africa, the number of HIV-infected and/or AIDS patients in Korea continues to increase. Furthermore, the recent mode of HIV transmission tends to be different from the mode of transmission in the early days.

Molecular phylogenetic analysis has become a valuable tool for the study of the evolutionary and systematic relationships among various organisms or genes of interest. In the case of HIV-1, the analysis of different HIV-1 genes illustrates the distinctive aspects of the molecular evolution of the HIV-1 viruses. The *pol* gene encoding reverse transcriptase is highly expressed in virions (Layne *et al.*, 1992) and is immu-

nogenic in the early response to HIV-1 (Hosmalin *et al.*, 1990). From the analysis of a large population, polymorphisms in the *pol* gene were demonstrated to be evident at sites of least functional or structural constraint and were frequently associated with particular host HLA class I alleles, suggesting the selective evolution of HIV-1 at a population level (Moore *et al.*, 2002). The *env* gene has been shown to show the most extensive evolutionary convergence, possibly because of its primary role in cell infection (Alkhatib *et al.*, 1996), cytotropism (Chesebro *et al.*, 1996), and because it is a recognition site for neutralizing antibody and cytotoxic T lymphocytes (Goudsmit *et al.*, 1988; Tsubota *et al.*, 1989). In fact, phylogenetic analysis of the *env* gene has been frequently used for molecular epidemiology. Although the *nef* gene encodes the transactivating factor (p27), it is also believed to be involved in the pathogenicity of HIV-1 (Kirchhoff *et al.*, 1995). Zanotto *et al.* (1999) presented evidence for the strong positive selection in the *nef* gene from genealogical analysis of a hemophiliac patient over 30 month of infection using the maximum likelihood method.

Attempts have been made to understand the molecular phylogenetic relationships among the HIV isolated from Korean patients by analyzing the nucleotide or amino acid sequences of the *nef* (Kang *et al.*, 1998), *pol* (Sung *et al.*, 2001), and *env* genes (Kim *et al.*, 1999a; Kim *et al.*, 1999b). In analyses of Korean HIV isolates, these workers were able to conclude that HIV-2 is rare in Korea and that

most of the HIV-1 isolated from Koreans belongs to subtype B. Moreover, a majority of the Korean subtype B isolates seems to be clustered together, forming a "Korean clade", which contains no foreign isolates. However, previous approaches used a limited number of Korean HIV isolates and did not include appropriate outgroups for phylogenetic analysis. In this study, we undertook to analyze all Korean HIV-1 isolates reported to the NCBI GeneBank in an effort to verify the existence and meaning of the Korean subcluster in subtype B.

Materials and Methods

Obtaining the nef nucleotide sequences

Total of 352 *nef* nucleotide sequences were analyzed in this study. Two hundred and sixty seven HIV-1 *nef* nucleotide sequences from Korean isolates are registered at NCBI GenBank. The 267 *nef* nucleotide sequences were obtained by searching for "hiv-1 AND nef AND Korea" in the GenBank database (http://www.ncbi.nlm.nih.gov/entrez/). Of the 267 *nef* nucleotide sequences, 24 sequences were too short for analysis and 243 sequences with more than 500 nucleotides were used for phylogenetic analysis. Foreign *nef* sequences were selected by choosing the corresponding most closely matching sequences using Blast (http://www.ncbi.nlm.nih.gov/BLAST/). Since many foreign sequences were selected redundantly, the number of unique foreign sequences was 127. Of these 50 sequences were excluded since they were from the same patients or had identical sequences, thus 77 foreign sequences were included in this study. Also included in this study were 32 *nef* nucleotide sequences from HIV-1 reference strains: 92UG037, 97CDKTB48 and U455 for subtype A; HXB2, JRFL, WEAU160, SF162, RF and OYI for subtype B; 92BR025 and

ETH2220 for subtype C; ELI and NDK for subtype D; CM240 for subtype E; 93br020 and 95CM-MP255 for subtype F; SE6165 and 92NG083 for subtype G; 90CR056 for subtype H; SE9280 for subtype J; 96CM-MP535 for subtype K; YBF30 for subtype N; MVP5180 for subtype O; 94CY032-3, 96TZ-BF061, 97CN001, ARMA159, BFP90, DJ263, GR17, KAL 153 and VI1310 for recombinant subtypes. The *nef* sequences of the reference strains were identified using the HIV SEQUENCE DATABASE (http:hiv-web.lanl.gov/content/hiv-db/mainpage.html). The NCBI accession numbers of the *nef* nucleotide sequences analyzed in this study are listed in Table 1.

Construction of phylogenetic trees

The 352 *nef* nucleotide sequences were aligned using ClustalX program (Thomson *et al.*, 1997), and the resulting alignments were confirmed by manual editing. Similarity tests on the *nef* nucleotide sequences and phylogenetic tree construction were conducted using a DNASIST program and the NEIGHBOR program included in the PHYLIP package (Felsenstein *et al.*, 1989; 1993). The distance matrix was obtained by using Jukes-Cantor and Kimuras method. And, cluster analysis was performed by Neighbor-joining, placing subtype O as an outgroup. Constructed phylogenetic trees were confirmed using the TreeView program (Page, 1996). The significance of the phylogenetic trees was verified by bootstrap analysis. Phylogenetic trees were constructed from one thousand replicates generated by the SEQBOOT program and the consensus tree was identified using the CONSENSE program. The resulting trees were too large to be seen clearly in an off-line print version of this manuscript, and computer files in Microsoft Powerpoint or Adobe Acrobat Reader format will be available upon request by e-mail to the corresponding author (mailto:chlee@chungbuk.ac.kr).
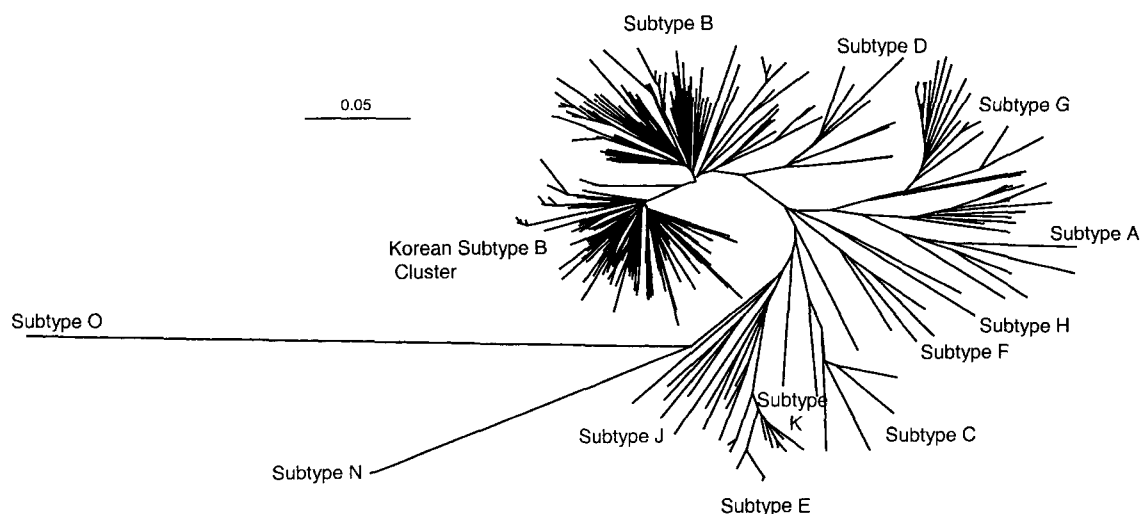
Fig. 1. Unrooted tree of 352 *nef* nucleotide sequences from 243 Korean isolates, 77 foreign isolates and 32 reference strains. All sequences were obtained from the NCBI GeneBank database. The tree was constructed using Jukes-Cantors model and the Neighbor-joining method in the PHYLIP package.

Table 1. HIV-1 isolates and their GeneBank accession numbers analyzed in this study

Subtype		Reference	Foreign isolates	Korean isolates
A		HIVU455A, AF193276, AF286238, HIVU51190	AF286237, AF457084, U86780	AF286239, AF462696, AF462724, AF462725, AF462736, AY121458, AY121459, AY121463, AY221652
D		HIVELICG, HIVNDK	AF425873, AF425892, AF484516	AF063920, AF462775, AF462776, AF462777, AY121473, AY221703
G		AF061642, AF063223, U88826	AB049811, AF107770, AJ232974, AJ251057	AF063908, AF063930, AF238275, AF462695, AF462698, AF462699, AF462702, AF462723, AF462770, AY121440, AY121443, AY121444, AY121445
Others		HIVMVP5180, HIV6022, HIVU54771, AF082394, AF005494, AF005496, AF179368, AF193253, HIVU52953, AF286226, AF385936, HIM249236, HIM249239, AF049337, AF289548, AF064699, HIVU46016	AF192135, AF197338, AJ232968, AJ245481, AJ291718, M15896, U48905, U48906, U48908, U48933	AF238270, AF462700, AF462717, AF462744, AF462780, AF462790, AY121446, AY121452, AY121478, AY221653, AY221659, AY221719, Z98033
B	Korean subtype B cluster			AF063909, AF063910, AF063911, AF063912, AF063913, AF063914, AF063915, AF063916, AF063918, AF063919, AF063921, AF063923, AF063924, AF063925, AF063927, AF063928, AF063929, AF063931, AF063932, AF063933, AF224507, AF238273, AF238274, AF238276, AF238277, AF238278, AF272003, AF462701, AF462703, AF462704, AF462705, AF462706, AF462707, AF462709, AF462710, AF462712, AF462713, AF462714, AF462715, AF462716, AF462718, AF462719, AF462720, AF462727, AF462728, AF462729, AF462730, AF462731, AF462741, AF462742, AF462745, AF462746, AF462748, AF462749, AF462750, AF462751, AF462752, AF462753, AF462754, AF462755, AF462756, AF462757, AF462758, AF462759, AF462760, AF462761, AF462762, AF462763, AF462764, AF462765, AF462766, AF462767, AF462768, AF462769, AF462771, AF462773, AF462774, AF462778, AF462779, AF462782, AF462783, AF462785, AF462786, AF462787, AF462789, AF462791, AF462793, AY121447, AY121448, AY121449, AY121450, AY121451, AY121455, AY121460, AY121461, AY121462, AY121465, AY121467, AY121468, AY121469, AY121470, AY121471, AY121472, AY121474, AY121475, AY121476, AY121477, AY121480, AY221654, AY221655, AY221656, AY221657, AY221660, AY221662, AY221663, AY221664, AY221665, AY221667, AY221668, AY221669, AY221671, AY221672, AY221673, AY221674, AY221675, AY221676, AY221681, AY221682, AY221684, AY221685, AY221687, AY221688, AY221689, AY221692, AY221693, AY221694, AY221695, AY221696, AY221697, AY221698, AY221699, AY221700, AY221702, AY221704, AY221705, AY221706, AY221708, AY221709, AY221710, AY221711, AY221713, AY221717, Z98019, Z98020, Z98021, Z98022, Z98023, Z98024, Z98026, Z98027, Z98028, Z98029, Z98030
	Other subtype B	HIVRF, HIVOYI, HIVS-FAAA, U63632, U21135, K03455	AB012816, AB012818, AF011472, AF011479, AF011483, AF047080, AF047081, AF082352, AF082375, AF082389, AF120773, AF120801, AF120825, AF120859, AF120899, AF129333, AF129335, AF129346, AF129358, AF129366, AF129367, AF129384, AF129385, AF203116, AF203120, AF203121, AF203147, AF219774, AF219800, AF219819, AF219838, AF219851, AF219856, AJ233015, AY064797, AY064804, L15489, M19921, U03303, U16875, U16877, U16878, U16898, U16921, U16940, U23487, U24455, U26060, U26079, U26104, U26111, U26942, U43096, U52499, U61790, U69591, U80225	AF063917, AF063922, AF063926, AF238268, AF238269, AF238271, AF272005, AF462693, AF462694, AF462697, AF462708, AF462721, AF462722, AF462726, AF462732, AF462733, AF462737, AF462739, AF462788, AF462792, AY121441, AY121442, AY121453, AY121454, AY121457, AY121479, AY221648, AY221649, AY221658, AY221666, AY221670, AY221714, AY221716, AY221718, Z98025, Z98031, Z98032, Z98034

Results and Discussion

Analysis of the phylogenetic tree based on nef nucleotide sequences: presence of Korean subtype B cluster

An unrooted phylogenetic tree of 352 HIV-1 *nef* nucleotide sequences from 32 HIV-1 reference strains (Kuiken *et al.,* 2001), 243 Korean isolates and 77 foreign isolates was constructed using Jukes-Cantors model and the neighbor-joining method (Fig. 1). The *nef* nucleotide sequences from subtype O and subtype N reference strains were used as outgroups. The reliability of the tree was estimated by bootstrap analysis. The 352 *nef* nucleotide sequences were divisible into five groups containing subtype B (bootstrap value=96.2%), subtype D (bootstrap value=98%), subtype A (bootstrap value=74.4%), subtype G (bootstrap value=83.8%) and other subtypes. Each group was comprised of subtype-specific reference strains. Korean isolates and foreign isolates were included in each of these groups. The biggest group was identified to be subtype B. Of the 243 Korean isolates, 202 belonged to subtype B (83.1%), 6 to subtype D (2.5%), 9 to subtype A (3.7%), 13 to subtype G (5.3%), and 13 to other subtypes (5.3%). In order to find the subtype for each isolate,
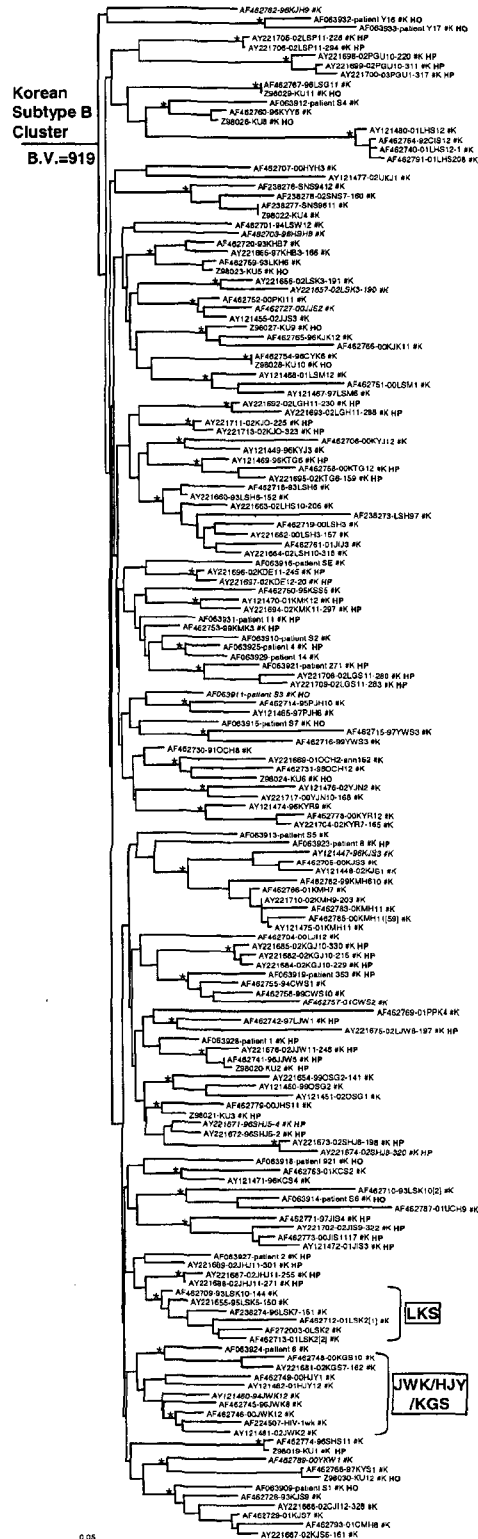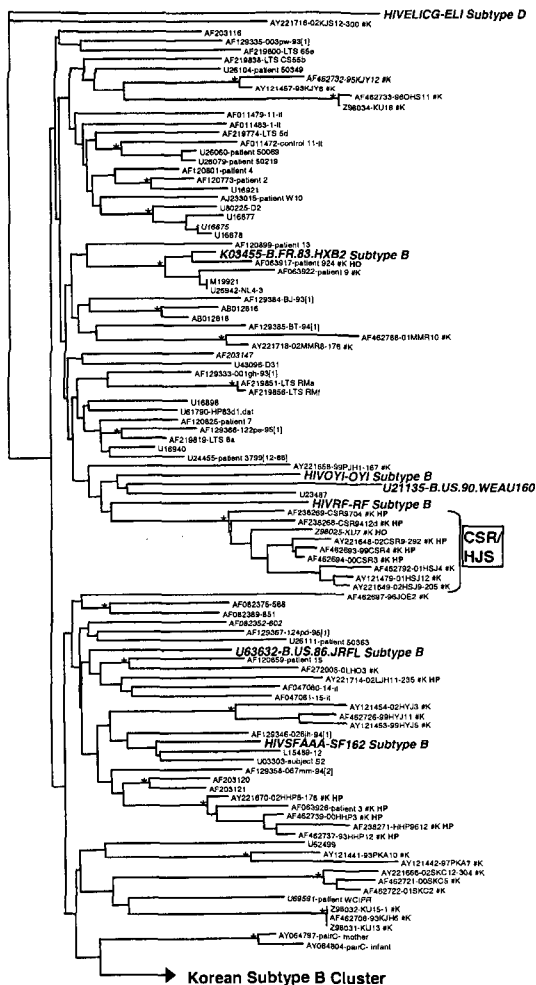
Fig. 2. Phylogenetic analysis of the subtype B *nef* nucleotide sequences. The tree includes 265 *nef* nucleotide sequences belonging to subtype B, of which 202 are from Korean isolates. Only bootstrap values greater than 50% are asterisked (*) in the tree. For easy identification, suffices are attached to the GeneBank accession numbers of each isolate from each patient described in the GeneBank database; HO, homosexuals, and HP, hemophiliacs. Korean isolates are indicated by #K.

please refer to Table 1.

As reported previously (Kang *et al.*, 1998; Kim *et al.*, 1999a), the majority of Korean isolates were found to belong to subtype B. This subtype B group included 202 Korean isolates and 73 foreign isolates, and some of the widely used subtype B reference strains such as, HXB2, JRFL, WEAU160, SF162 and RF. The bootstrap value for subtype B was 96.2%, indicating that the subtype B group is distinct from the other subtypes. A careful look at the subtype B group revealed a big subcluster, containing 164 Korean isolates exclusively of any foreign isolates or reference strains. This subcluster of Korean isolates-only appeared to be similar to the "Korean clade" or "Korean cluster" mentioned by other investigators (Kang *et al.*, 1998; Sung *et al.*, 2001). We named this subcluster the "Korean subtype B cluster" rather than the "Korean clade" since the term clade is defined as a group that shares a common ancestor that is not shared by any other group outside the clade (Li and Grauer, 1991), and the basis of this Korean subtype B cluster is not clear at this time. The Korean subtype B cluster appeared to be very tightly clustered with a bootstrap value of 91.9%, and to consist of 81.2% of Korean isolates of subtype B. This figure also means that the Korean subtype B cluster accounts for 67.5% of all Korean HIV-1 isolates.

Analysis of Korean subtype B cluster: absence of meaningful subclusters within

Since the Korean subtype B cluster forms a large group, the possibility of the existence of a subcluster of moderate size within the Korean subtype B cluster was investigated. Although some isolates appeared to form a small subcluster, the presence of moderately-sized subclusters could not be confirmed (Fig. 2). In fact, with a lower-than-normal threshold bootstrap value of 50%, meaningful subclusters were composed of 2 to 9 isolates. Most (80%) of the subclusters consisted of 2 or 3 isolates and only 5 subclusters contained more than 5 isolates. Moreover, a majority (24 out of 45 subclusters with bootstrap values higher than 50%) of the subclusters were found to be composed of HIV-1 isolates from the same patients taken at different times as judged by the codes of the isolates. For example, a subcluster of 6 isolates from the same patient who had an initial of LSK included AF462709 (93LSK10-144), AY221655 (95LSK5-150), AF238274 (96LSK7-151), AF272003 (0LSK2), AF-

462712 (01LSK2[1]), and AF462713 (01LSK2[2]). In some cases, the codes of isolates did not have any information on the name of the patients (for example, KU1, etc), and we classified these isolates as different sources, although they may in fact have been from the same patient (see below).

Another characteristic of Korean subtype B cluster is that the middle branches with very short and similar lengths to each subcluster from the branching point, which separates Korean subtype B cluster from the other B subtype. The presence of the middle branches is rather ambiguous. On the other hand, the terminal branches to each isolates are relatively long and diverse in length. The short middle branches may account for the difficulty of identifying the presence of the distinctive meaningful subclusters within Korean subtype B cluster. According to the phylogenetic analysis of the p19 *gag* and V3 region of *env* from 9 Swedish patients with known infection routes and time (Leitner *et al.*, 1996), similar phylogenetic trees were constructed by applying different analytical methods, such as Neighbor-joining, Fitch-Margoliash, and Maximum-likelihood. However, all of these methods resulted in an overestimation of short branches and an underestimation of long branches. Application of a rather complex base-substitution model did not improve the results. Therefore, the short middle branches and long terminal branches might be due to errors inherent in the analytical method.

Analysis of amino acid sequence

The amino acid sequences of the *nef* gene were also analyzed using the Neighbor-joining and Weighed-joining methods with bootstrap calculations. The overall structure of the phylogenetic tree appears to be very similar to that of the nucleotide sequence-based tree (data not shown). As with the nucleotide analysis, the Korean cluster seemed to be present in subtype B, and again there are no distinctive subclusters within this Korean subtype B cluster. The bootstrap values for the overall cluster were much lower than those of the nucleotides; probably due to the shorter length of the amino acid sequence. Although the Korean subtype B cluster showed a relatively low bootstrap value of 12.6%, the individual isolates in this cluster coincided with those grouped as the Korean subtype B cluster from the nucleotide analysis, without exception, further confirming the existence of the Korean subtype B cluster.

Table 2. *Nef* nucleotide sequence diversity among subtype B from Korean isolates

Clusters	Mean±SD[a] (Minimum-Maximum)	Mean ± SD[b] (Minimum-Maximum)
Subtype B, total	10.8 ± 3.9% (0-36.1%)	10.5 ± 3.4% (0-30.6%)
Korean subtype B cluster	8.4 ± 3.6% (0-36.1%)	8.0 ± 2.6% (0-23.6%)
Other subtype B	10.6 ± 3.1% (0-21%)	10.6 ± 3.1% (0-21%)
Korean subtype B cluster: Other subtype B	12.7 ± 3.2% (6.6-33.7%)	12.5 ± 2.7% (6.6-30.6%)

[a]Differences in *nef* nucleotide sequences from 202 Korean HIV-1 isolates
[b]Differences in *nef* nucleotide sequences from 200 Korean HIV-1 isolates, after deleting AF462783 and AF462785

Nef nucleotide sequence diversity among subtype B from Korean isolates

The relativities of isolates in a cluster could be determined by measuring the genetic distance within a cluster or inter-clusters. Differences in *nef* nucleotide sequences in 202 subtype B, 164 Korean subtype B cluster, and 38 other subtype B were calculated (Table 2). Difference between Korean subtype B cluster and other subtype B was also calculated. The *nef* nucleotide sequence diversity was lowest within the 164 Korean subtype B cluster (8.4±3.6%), while the *nef* nucleotide sequences of the Korean subtype B cluster differed from the other subtype B by 12.7±3.2%. The relatively higher standard deviation observed in the Korean subtype B cluster was due to two isolates (AF462783 and AF462785) that were extraordinarily outstanding from the other isolates. After omitting these two isolates, the standard deviation decreased as shown in Table 2. Our data suggest that the Korean subtype B cluster is relatively homogeneous as compared with the other subtype B, and that the Korean subtype B cluster is distinguishable from the other subtype B. Our observations in this respect, are similar to those of Kim *et al.* (1999a) in that the *env* nucleotide sequences differ less in the Korean subtype B cluster than in the *env* nucleotide sequence of the other subtype B.

While analyzing the diversity of the *nef* nucleotide sequences, it was noted that some isolates were 100% identical to each other in terms of nucleotide sequence. This is the reason for the above observation that the minimum nucleotide sequence difference was 0%, which is quite unusual. For example, in Korean subtype B cluster, Z98022 was identical to AF238277, Z98028 to AF-462754, and Z98029 to AF462767. In the other subtype B, Z98031, Z98032 and AF462708 were identical. Thus, it is presumed that these isolates are actually the same isolates with different GenBank accession numbers.

A relatively homogeneous population, as observed in the Korean subtype B cluster, has been reported in the case of restricted epidemiology, best exampled by the HIV subtype C transmission in India from a specific founder virus (Shankarappa *et al.*, 2001). However, it is difficult to imagine that HIV transmission has occurred from a single origin during the short period of HIV-1 transmission in Korea, and to a majority of infected people. In addition, we were not able to find any founder virus or genealogically significant subcluster within the Korean subtype B cluster, perhaps due to the high substitution rate of the *nef* gene (2.3×10^{-5} to 7×10^{-6} per genome replication) and its rapid generation time (2.6 days) (Temin, 1993). Therefore, it is presumed that certain selective processes might work in the transmission of HIV-1 in Korea, as suggested previously (Kang *et al.*, 1998).

Distribution of the patient groups in the phylogenetic tree

Some of the Korean HIV-1 isolates were annotated in the

GenBank database and in Kang *et al.* (1998) as being related to disease or to the potential contraction routes. These include 58 hemophiliacs (or clotting factor deficient) and 17 homosexuals (Fig. 2). Of the 17 homosexuals, 14 belonged to the Korean subtype B cluster, 2 to the other subtype B, and 1 to subtype G. In the case of 58 isolates from Korean hemophiliacs, all were located to subtype B, and of these 47 (81%) belonged to the Korean subtype B cluster. As mentioned earlier, many isolates were actually taken from the same patients at different times. When this redundancy was taken into account, we identified 28 hemophiliac patients who had contracted HIV-1 as shown in the phylogenetic tree of Fig. 2. This number seems to be exaggerated since the same patient may have been registered using two different stain codes (see above). Simple calculation indicated that 85.7% (24/28) of the Korean hemophiliacs belonged to Korean subtype B. The distribution of most of the homosexuals and hemophiliacs within the Korean subtype B cluster appeared random with low bootstrap values, although some isolates looked clustered. Careful examination of these apparent hemophiliac subclusters, however, revealed that they were formed by multiple isolates from the same people. Thus, it could be concluded that the distribution of a group of patients such as homosexuals and hemophiliacs appeared as random as that of the other general population.

According to GenBank information about the patients, there were two cases with known transmission routes, a hemophiliac and his wife (CSR, and HSJ) and a blood transfusion (donor JWK to recipients KGS and HJY). The *nef* nucleotide sequences of the HIV-1 isolates from CSR and HSJ formed a subcluster with a bootstrap value of 50.8%, somewhat supporting the known epidemiology. In contrast, the blood transfusion group of JWK, HJY and KGS did not form any meaningful subcluster since the bootstrap value for the apparent subcluster was as low as 10.1%.

In order to use phylogenetic analysis as a tool for molecular epidemiology, many aspects have to be considered such as sampling time, target population, target genes, sample size, and the quality control of laboratory data (Slattery, 2002). The most successful demonstration of likelihood relatedness was provided by the case of a gastroenterologist who was charged with second-degree murder by injecting his former girlfriend with blood obtained from a HIV-1-infected patient, which resulted in the first use of phylogenetic analyses in a criminal court case in the United States (Metzker *et al.*, 2002). From the phylogenetic analyses of HIV-1 reverse transcriptase and *env* DNA sequences isolated from the victim, the patient, and a local population sample of HIV-1-positive individuals, they demonstrated that the sequence of the victim was closely related to that of a patient with a direction of transmission from the patient to the victim. In contrast,

our analysis did not appear to support the existing transmission route with reliable confidence, except for multiple samples from the same individuals.

In conclusion, the results of the current phylogenetic analysis of the Korean *nef* gene sequences show the presence of a distinctive Korean subtype B cluster exclusive of any foreign sequences. However, resolution within this cluster was too low to make possible interpersonal likelihood relationships and did not provide evidence for the presence of a single or a small number of sources responsible for the formation of the Korean subtype B cluster. More extensive analysis linked with epidemiological information of well-designed HIV-1 samples under tightly controlled analytical conditions are prerequisite for the accurate characterization of the possible causes and significance for the Korean subtype B cluster.

Acknowledgment

References

Alkhatib, G., C. Combadiere, C.C. Boder, Y. Feng, P.E. Kennedy, P.M. Murphy, and E.A. Berger. 1996. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 272, 1955-1998.

Chesebro, B., K. Wehrly, J. Nishio, and S. Perryman. 1996. Mapping of independent V3 envelope determinants of human immunodeficiency virus type 1 macrophage tropism and syncytium formation in lymphocytes. *J. Virol.* 70, 9055-9059.

Felsenstein, J. 1989. PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164-166.

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

Goudsmit, J., C. Debouck, R.H. Meloen, L. Smit, M. Bakker, D.M. Asher, A.V. Wolff, C.J. Gibbs Jr, and D.C. Gajdusek. 1988. Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. *Proc. Natl. Acad. Sci. USA* 85, 4478-4482.

Hosmalin, A., M. Clerici, R. Houghten, C.D. Pendleton, C. Flexner, D.R. Lucey, B. Moss, R.N. Germain, G.M. Shearer, and J.A. Berzofsky. 1990. An epitope in human immunodeficiency virus 1 reverse transcriptase recognized by both mouse and human cytotoxic T lymphocytes. *Proc. Natl. Acad. Sci. USA* 87, 2344-2348.

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., 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.

Kirchhoff, F., T.C. Greenough, D.B. Brettler, J.L. Sullivan, and R.C. Desrosiers. 1995. Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. *N. Engl. J. Med.* 332, 228-232.

Kuiken, C.L., B. Foley, B. Hahn, B. Korber, P.A. Marx, F. McCutchan, J.W. Mellors, and S. Wolinksy. 2001. HIV Sequence Compendium 2001. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico.

Layne, S.P., M.J. Merges, M. Dembo, J.L. Spouge, S.R. Conley, J.P. Moore, J.L. Raina, H. Renz, H.R. Gelderblom, and P.L. Nara. 1992. Factors underlying spontaneous inactivation and susceptibility to neutralization of human immunodeficiency virus. *Virology* 189, 695-714.

Leitner, T., D. Escanilla, C. Franzen, M. Uheln, and J. Albert. 1996. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc. Natl. Acad. Sci. USA* 93, 10864-10869.

Li, W.-H. and D. Graur. 1991. Fundamentals of Molecular Evolution. Sinauer, Sunderland, Massachusetts.

Metzker, M.L., D.P. Mindell, X.M. Liu, R.G. Ptak, R.A. Gibbs, and D.M. Hillis. 2002. Molecular evidence of HIV-1 transmission in a criminal case. *Proc. Natl. Acad. Sci. USA* 99, 14292-14297.

Moore, C.B., M. John, I.R. James, F.T. Christiansen, C.S. Witt, and S.A. Mallal. 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 296, 1439-1443.

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

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.

Slattery, M.L. 2002. The science and art of molecular epidemiology. *J. Epidemiol. Community Health* 56, 728-729.

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.

Temin, H.M. 1993. The high rate of retrovirus variation results in rapid evolution. p219-233. *In* S.E. Morse (ed), Emerging viruses, Oxford University Press, New York, Oxford.

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.

Tsubota, H., C.I. Lord, D.I. Watkins, C. Morimoto, and N.L. Letvin. 1989. A cytotoxic T lymphocyte inhibits acquired immunodeficiency syndrome virus replication in peripheral blood lymphocytes. *J. Exp. Med.* 169, 1421-134.

Zanotto, P.M., E.G. Kallas, R.F. de Souza, and E.C. Holmes. 1999. Genealogical evidence for positive selection in the nef gene of HIV-1. *Genetics* 153, 1077-1089.