

| S10-4 |

Unity in HIV-1 Sequence Diversity: Identification and Characterization of Korean Clade in HIV-1 Isolated from Korean

Chan Hee Lee

Division of Life Sciences, College of Natural Sciences, Chungbuk National University

Abstract

Through molecular phylogenetic analysis using the *nef* gene sequences of HIV-1 isolated from Korean registered in the NCBI GenBank together with 41 reference strains and 94 foreign isolates, we verified that most (~80%) of Korean isolates belonged to subtype B and 78% of subtype B were clustered together exclusively of foreign isolates, and this cluster was named Korean clade subtype B (K_CB). Similarity study suggested that the K_CB cluster was more homogeneous than and clearly distinctive from the non-Korean subtype B (NK_CB). Comparison of the consensus amino acid sequences of the K_CB or NK_CB revealed characteristic K_CB signature amino acid pattern comprised of 13 amino acid residues. The K_CB signature amino acid residues were critical in separating the K_CB from the NK_CB, since substitution of the NK_CB sequences with K_CB signature amino acids relocated them to the Korean clade, and vice versa. Synonymous and nonsynonymous substitution rate study suggested positive selection event for the K_CB.

Introduction

Since the first detection in 1985 of HIV-1 infection in Korea, total of 3,829 have been identified as being infected with HIV in Korea as of December 2005. Although this number is much smaller compared with that in other countries, Korean health authorities worries that the infection rate have been increasing. Despite the small number of HIV-infected people in Korea, an interesting feature of the so-called "Korean clade" has been reported (Kang *et al.*, 1998; Kim *et al.*, 1999a; Kim *et al.*, 1999b; Sung *et al.*, 2001; Lee *et al.*, 2003). Korean clade was first proposed based on the molecular phylogenetic studies of *nef* gene sequences of HIV-1 isolated from Koreans (Kang *et al.*, 1998). Korean clade by definition is a cluster of Korean sequences only and no other foreign sequences are included. Later studies with *env* (Kim *et al.*, 1999a; Kim *et al.*, 1999b) and *pol* (Sung *et al.*, 2001) gene sequences of Korean HIV-1 isolates supported the presence of Korean clade. A more comprehensive confirmed the presence of

Korean clade (Lee *et al.*, 2003). In this study, using all the Korean *nef* sequences registered in GenBank database, we confirmed the presence of the Korean clade, identified and characterized the signature amino acid residues defining the Korean clade.

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 was constructed. Subtype B was the biggest group and accounted for 79.4% of all Korean isolates. Examination of the subtype B revealed that a big cluster comprised of Korean isolates only was separated from the other cluster where Korean and foreign including reference sequences. This cluster of Korean isolates exclusive of foreign sequences is by definition the Korean clade. The size of the Korean clade is 78.8% of subtype B of Korean isolates or 62.6% of the total Korean isolates.

Generation and comparison of consensus sequences of Korean subtype B clade and of other subtype B

Comparison of the consensus amino acid sequences revealed that there were four substitutions between SubB-con and NK_CB-con; V11A, T15A and S163C. On the other hands, there were 14 substitutions between SubB-con and K_CB-con: V11P, P14N, V16I, A23T, D28E, K39Q, H40R, S45T, A50S, D54A, N157T, S163C, S169N, and L198K.

Distances of each sequence from consensus sequences were calculated. The nucleotide sequences distances of K_CB from K_CB-con were calculated to be $3.8 \pm 1.3\%$ and those of NK_CB to K_CB-con were much greater, $10.0 \pm 1.2\%$. The distances from NK_CB-con to the sequences of K_CB or NK_CB were $7.9 \pm 1.1\%$ and $6.2 \pm 1.2\%$, respectively. The distances from SubB-con to the sequences of K_CB or NK_CB were $7.7 \pm 1.1\%$ and $6.5 \pm 1.3\%$, respectively.

Signature pattern analysis of the Korean clade subtype B

Comparison of SubB-con, K_CB-con and NK_CB-con identified 13 amino acid residues unique to KB-con. These unique amino acid residues are referred as signature pattern. Overall, the K_CB signature amino acid residues occurred more than 80%, except for N14 (58%), in K_CB. In contrast, the frequency of the K_CB signature amino acid residues was at most 23% in NK_CB. There are amino acid residues most frequently found in all subtype B isolates worldwide. These amino acid residues, however, occurred very infrequently in K_CB, ranging from 16% D at position 54 to 0% A, P, K or A at position 11, 14, 39 or 50, respectively. On the other hands, these amino acids were the most frequently found residues in NK_CB. Korean clade can be differentiated from non-Korean clade subtype B or foreign subtype B on the basis of the signature 13 amino acid residues, with strong statistical support.

If the signature 13 amino acid residues differentiate K_CB from N K_CB, substituting the corresponding amino acids of NK_CB or subtype B reference sequences with the signature 13 amino acids may place them in the Korean clade. Hence new sequences with the signature 13 amino acid residues were generated from 4 randomly selected NK_CB sequences from 4 different patients and 3 subtype B reference sequences. These new sequences were subjected to phylogenetic analysis together with the subtype B sequences from Korean isolates and subtype B reference sequences. The new sequences generated from subtype B were located in the Korean clade, while their original sequences were located outside the Korean clade. 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 (14 to 24 amino acid residues), NK_CB sequences remained outside the Korean clade. Therefore, it could be concluded that the signature 13 amino acid residues were responsible for clustering of the Korean clade.

References

1. Kang MR, Cho YK, Chun J, Kim YB, Lee I, Lee HJ, Kim SH, Kim YK, Yoon K, Yang JM, Kim JM, Shin YO, Kang C, Lee JS, Choi KW, Kim DG, Fitch WM, Kim S. 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.
2. Kim YB, Cho YK. 2003. Monophyletic clade of HIV-1 subtype B in Korea: evolutionary pressure or single introduction? *AIDS Res Hum Retroviruses.* 19:619-623.
3. Kim EY, Cho YS, Maeng SH, Kang C, Nam JG, Lee JS. 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.
4. Kim YB, Cho YK, Lee HJ, Kim CK, Kim YK, Yang JM. 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.
5. Lee DH, Yoon Y, Lee CH. 2003. Phylogenetic analysis of the HIV-1 nef gene from Korean isolates. *J Microbiol.* 41: 232-238.
6. Sung H, Foley BT, Bae IG, Chi HS, Cho YK. 2001. Phylogenetic analysis of reverse transcriptase in antiretroviral drug-naive Korean HIV type 1 patients. *AIDS Res Hum Retroviruses.* 17:1549-1554.