Analysis of Virus Evolution Based on Inhomogeneous Markov Evolution Model

Fengrong Ren rencom@mri.tmd.ac.jp
Hiroshi Tanaka tanaka@cim.tmd.ac.jp
Medical Research Institute, Tokyo Medical and Dental University
Yushima 1-5-45, Bunkyo, Tokyo 113, Japan

Abstract

We have been studying the reconstruction method of molecular phylogenetic tree for these years, and a new method for reconstructing phylogenetic tree was developed based on the model-based complexity (MBC). In this study, by using MBC method we reconstruct the phylogenetic trees of three human immunodefiviency virus type 1 sequences which are isolated sequentially from one patient. The results indicate that not only neutral evolution but also the accelerated evolution caused by the positive selection may exist in virus evolution process. Therefore, a new model-inhomogeneous Markov evolution model for virus evolution is thought to be appropriate.

1 Introduction

It is well-known that the rate of virus evolution is very rapid compared with most DNA genomes, especially the rate of retrovirus evolution which has an error-prone reverse transcriptase is valued to be millionfold greater than for DNA genomes. This feature of virus is quite appropriate for analysis of molecular evolution. For this reason, the studies of virus evolution are widely studied for this years. In these studies, however, the problem of virus evolution is studied only within the framework of neutral evolution theory, that is, the substitution rate is assumed to equal to the constant mutation rate. We suspect the validity of this assumption. In this study, the evolutionary relation of three HIV-1 sequences which are sequentially isolated from a patient with AIDS are estimated by MBC method, and the results lead us to suggest a new model: inhomogeneous Markov evolution model for virus evolution.

1.1 Previous Work

In our previous studies, a model-based complexity method was proposed to be used for reconstruction of molecular phylogenetic tree. The results of computer simulation suggested that MBC method has a good asymptotic property compared with traditional maximum likelihood method or its modification by Akaike's AIC, especially in the case that the multifurcate tree is considered as a candidate topology of tree and/or long sequences could be used in reconstructing phylogenetic tree [1]. The program package of MBC method has been developed by us and will be made public in internet in next year.

2 Data and Method

The sequences (2571-bp) of *env* genes of HIV-1 which are denoted as WMJ1, WMJ2 and WMJ3 [2] were extracted from GenBank. The patient is a 2-year-old child with AIDS and infected perinatally by her HIV-positive mother. Three sequences were sequentially isolated from her in 10/1984, 1/1985 and 5/1985. We aligned these sequences and reorganized them into two groups: one group consists of the three sequences in which the nonsynonymously substituted sites are excluded from original

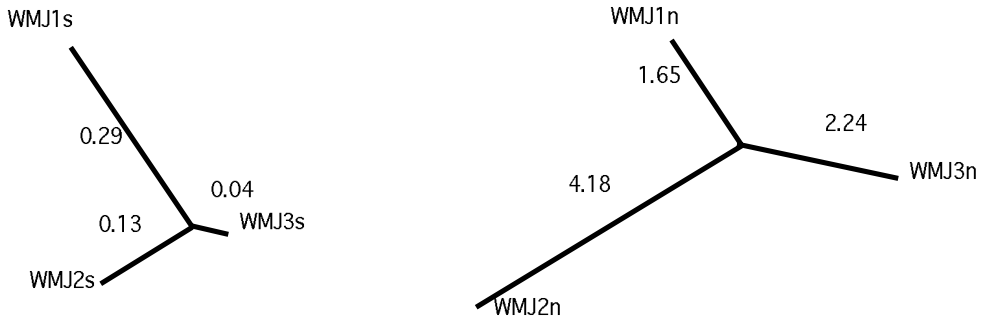

Figure 1: The phylogenetic trees of *env* genes (2571-bp) of HIV-1 type by using MBC method

env gene sequences, and the other group consists of the three sequences in which the synonymously substituted sites are excluded. Then we reconstructed the phylogenetic trees by using the data of two groups respectively. The programs of MBC method was employed to estimate the phylogenetic trees.

3 Results and Conclusions

The results are shown in Figure 1. The tree of left side is reconstructed by using the three sequences having synonymous substitutions. We can see that the distance which indicates the substitution numbers between WMJ1s and WMJ2s is 0.42, and that between WMJ2s and WMJ3s is 0.17. However, the distance between WMJ2s and WMJ3s is only 0.33. The tree of right side is reconstructed by using the three sequences having nonsynonymous substitutions. In this tree, the distance between WMJ1n and WMJ3n is still short than the sum of the distance between WMJ1n and WMJ2n plus the distance between WMJ2n and WMJ3n. Moreover, it should be noted that the WMJ2n is quite a long branch compared with the WMJ1n and WMJ3n.

These results indicate two important points. First point is that the virus had evolved in parallel, not directly, from a common progenitor virus. The second point, which is very important in virus evolution, is that the Markov model based on neutral evolution theory is not sufficient to describe virus evolution. The fact that the branch length of WMJ2n is very long compared with WMJ1n and WMJ3n makes it evident that the positive selection may operated on the genes and caused accelerated evolution in that period. Therefore we would like to suggest to use **Inhomogeneous Markov Evolution Model** for virus evolution. We think that two periods may exist in virus evolution: one is **exploring period** which is characterized by wide-ranging divergence, and the other is **positive selecting evolution period** in which the suitable direction of base substitution is found and positive selection operate on the genes, thus the evolution is accelerated. We would now like to go on to develop this theory by investigating more examples of virus sequences.

References

[1] Tanaka, H., Ren, F, Okayama, T. and Gojobori, T., "Inference of Molecular Phylogenetic Tree Based on Minimum Model-Based Complexity Method," *Proc. of ISMB97*, 319–328, 1997.

[2] Hahn, B.H. *et al*, "Genetic Variation in HTLV-III/LAV Over Time in Patients with AIDS or at Risk for AIDS," *Science*, 232:1548–1553, 1986.