数理解析研究所講究録
1066 巻 1998 年 145-163

ON EVOLUTION OF HIV

Numazu College of Technology Keiko Sato
Science University of Tokyo Masanori Ohya

Abstract

We analyze the variation of HIV for patient after his infection by means of an information measure called the entropy evolution rate and by writing phylogenetic trees of HIV. In our analysis of HIV, we use sequences of HIV, in particular a part of external glycoprotein gp120 including the V3 region, obtained from eight patients.

We conclude that the entropy evolution rate can be a measure grasping the course of progression to AIDS and the tree shows us the situation of patient as a whole.

1. INTRODUCTION

The main purpose of this study is to find a new criterion grasping the variation of HIV such that the immunity of patients from the gene level after their HIV infection.

In Section 2, we briefly explain HIV. In Section 3, the entropy evolution rate, a fundamental tool for our analysis, is reviewed. In Section 4, we summarize the data of HIV genes of eight patients reported in [2, 3, 4, 5]. Then a method of analysis how to use the entropy evolution rate is discussed, and the results are presented by the graph in this section. In Section 5, we show an axiomatic approach to write phylogenetic trees and construct the trees for evolution of HIV. We discuss our results and the usefulness of our method in Section 6.

2. ON HIV

The virus which causes AIDS (acquired immunodeficiency syndrome) is called HIV (human immunodeficiency virus) and has only RNA as its gene, the same as other retroviruses. After this virus (HIV) gets into a hostcell, the viral RNA is transcribed into DNA by reverse transcriptase. This DNA is integrated into chromosomal DNA of the host cell. The integrated proviral DNA is transcribed by a transcription apparatus of host, then the viral RNA transcribed produces viral protein and genomic RNA. These are assembled and are budded out of the cell. HIV infects other cell one by one in the above process.

The genome of HIV is principally composed of three proteins, called gag, pol and env. It is particularly known that the envelope protein has great many variations. The envelope gene consists of the gp120 (outer membrane) and the gp41 (transmembrane). The envelope glycoprotein gp120 contains the hypervariable regions (V1-V5) and the stable regions (C1-C5). The third variable region (V3) , composed of disulfide bounds of cysteine residues located in the amino acid 296 and 330 of the gp120, has particularly high mutation rate [6, 7]. Although it has been called the principal neutralisation domain (PND) which enables an antibody to block the HIV infection, the antibody for a specific virus is gradually losing its effect because of the mutation of the virus. Therefore, the V3 region are often chosen as a target to analyze the mutation of HIV.

The gp120 of HIV infects the cell with CD4 molecular, which is called CD4$^+$ cell and is a receptor of HIV. Therefore the increase of HIV is caused by the adsorption to CD4$^+$T-lymphocytes, macrophage, B-cell, Langerhans-cell and others, which carry CD4 molecular to the surface of cell. In particular, the immunodeficiency of patients infected HIV is due to the decrease of CD4$^+$T-lymphocytes. The CD4$^+$T-lymphocytes represents the number of immunocyte destroyed by HIV. The immunocyte for healthy people is around from 800 to 1000 per $\mu \ell$. When the CD4 count decreases and it becomes less than 200, the various infections are considered to appear. Therefore, according to the diagnosis standard of CDC(Center for Disease Control), when the CD4 become less than 200, the patient is recognized to have AIDS. The p24 antigen in blood reflects the amount of the virus, and it appears at early infection and reappears at time when the condition of a patient deteriorates, so that it is used as a value measuring the course of progression to disease.

3. ENTROPY EVOLUTION RATE

We consider two aligned amino acid (resp. base) sequences $\mathcal{A}$ and $\mathcal{B}$, which are composed of 20 (resp. 4) kinds of amino acids (resp. bases) and the gap *. The complete event system $(\mathcal{A},p)$ of $\mathcal{A}$ is determined by the occurrence probability p_i of each amino acid (resp. base) a_i and the gap * $(0 \le i \le 20)$ (resp. $0 \le i \le 4$) with $a_0 = *$;

$$\begin{pmatrix} \mathcal{A} \\ p \end{pmatrix} = \begin{pmatrix} *, & a_1, & \cdots, & a_{20} \\ p_0, & p_1, & \cdots, & p_{20} \end{pmatrix} \left(\text{resp.} \begin{pmatrix} *, & a_1, & \cdots, & a_4 \\ p_0, & p_1, & \cdots, & p_4 \end{pmatrix} \right)$$

In the same way, the complete event system $(\mathcal{B},q)$ of $\mathcal{B}$ is

$$\begin{pmatrix} \mathcal{B} \\ q \end{pmatrix} = \begin{pmatrix} *, & a_1, & \cdots, & a_{20} \\ q_0, & q_1, & \cdots, & q_{20} \end{pmatrix} \left(\text{resp.} \begin{pmatrix} *, & a_1, & \cdots, & a_4 \\ q_0, & q_1, & \cdots, & q_4 \end{pmatrix} \right)$$

We can construct the compound event system $(\mathcal{A} \times \mathcal{B}, r)$ for two sequences $\mathcal{A}$ and $\mathcal{B}$.

$$\begin{pmatrix} \mathcal{A} \times \mathcal{B} \\ r \end{pmatrix} = \begin{pmatrix} **, & *a_1, & \cdots, & a_{20}a_{20} \\ r_{00}, & r_{01}, & \cdots, & r_{2020} \end{pmatrix}$$

$$\left(\text{resp.} \begin{pmatrix} **, & *a_1, & \cdots, & a_4a_4 \\ r_{00}, & r_{01}, & \cdots, & r_{44} \end{pmatrix} \right)$$

where r_{ij} represents the joint probability of the event i of a and the event j of $\mathcal{B}$.

These event systems define various entropies, among which the following two are important:

(1) Shannon entropy

$$S(\mathcal{A}) = -\sum_i p_i \log p_i,$$

which expresses the amount of information carried by$(\mathcal{A},p)$.

(2) The mutual entropy

$$I(\mathcal{A},\mathcal{B}) = \sum_{i,j} r_{i,j} \log \frac{r_{i,j}}{p_i q_j}$$

which expresses the amount of information transmitted from $\mathcal{A}$ (resp. $\mathcal{B}$) to $\mathcal{B}$ (resp. $\mathcal{A}$).

Using the above information measures, a measure indicating the difference between two amino acid sequences was introduced in [1]. This measure is called the entropy evolution rate and defined as follows: Put

$$r(\mathcal{B}/\mathcal{A}) = \frac{I(\mathcal{A},\mathcal{B})}{S(\mathcal{A})}$$

which is the rate how much information is transmitted from $\mathcal{A}$ to $\mathcal{B}$, and it is symmetrized as

$$r(\mathcal{A},\mathcal{B}) = \frac{1}{2}\{r(\mathcal{A}/\mathcal{B}) + r(\mathcal{B}/\mathcal{A})\}$$

The entropy evolution rate $\rho(\mathcal{A},\mathcal{B})$ is defined by

$$\rho(\mathcal{A},\mathcal{B}) = 1 - r(\mathcal{A},\mathcal{B})$$

In this paper, we use this entropy evolution rate to examine the variation of HIV sequences of six patients. The entropy evolution rate takes the value in [0,1]; $\rho(\mathcal{A},\mathcal{B}) = 0$ if $\mathcal{A}$ and $\mathcal{B}$ are completely the same and $\rho(\mathcal{A},\mathcal{B}) = 1$ if they are completely different. Therefore the variation of HIV becomes larger, the entropy evolution rate is getting larger.

4. VARIATION OF HIV

4.1. Patient selection

The data used in our analysis are the base sequences of HIV for eight patients reported in [2, 3, 4, 5]. We obtained the data from International Nucleotide Sequence Database(DDBJ/EMBL/GenBank). Here, eight patients are designated as patient A to patient D, patient G, patient H, patient J and patient K. The facts reported for the eight patients are summarized in Table1.

Patients A and B were studied during a follow-up period of 5 years after primary HIV infection. It is reported [2] that sequences were derived by PCR from genomic RNA out of serum without cultivation. Here PCR is a method to detect a specific sequences by amplifying it in quantity. At the

Table1. Data used in our analysis

Designation in our analysis	Patient A	Patient B	Patient C	Patient D	Patient G	Patient H	Patient J	Patient K
Designation in ref.[2, 3, 4, 5]	Patient 1	Patient 495	Patient 82	S1	P1	P2	P4	P5
Presumed transmission mode	homosexual contact	homosexual contact	a single batch of factor VIII	no information	no information	no information	no information	no information
Clinical status	p24 antigenemia (1988)	p24 antigenemia AIDS (1989)	asymptomatic	no information	died within 36 months	died within 42 months	no information	no information
CD4 counts during the study	decreasing	decreasing	decreasing	fluctuating	decreasing	decreasing	fluctuating	stable
Antiviral therapy	none	AZT (1989)	none	none	none	AZT in 30 months	AZT in 64 months	none
Term	1985~ (about 5years period)	1985~ (about 5years period)	1984~1991 (7years period)	1985.11~1989.5 (42 months period)	1985~ (32 months period)	1985~ (32 months period)	1986~(about 3years period)	1984~ (47 months period)
Nucleotide sequences	183~276nt	183~276nt	234nt	332~335nt	658~665nt	653~662nt	659~668nt	650~665nt
Tissue	serum	serum	plasma	peripheral blood leucocyte	blood	blood	blood	blood
Molecular type	RNA	RNA	RNA	DNA	DNA	DNA	DNA	DNA

time patient B was diagnosed as having AIDS in 55 months after infection for both a decline in CD4 counts and a reappearance of p24 antigen, therapy with azidothymidine (AZT) was started.

Patient C was infected from a locally prepared batch of factor VIII was studied over the period of 7 years. It is reported [3] that viral RNA sequences were obtained directly from nested PCR amplified single molecular.

Patients D was studied over the period of 42 months. For patient D, his CD4 count changes as 470, 826, 273, and 515 from the early stage up to 4th stage [4].

Patients G, H, K and L were studied during a follow-up 32 months and 35 months, respectively. It is reported [5] that the proviral sequences amplified by the PCR from blood samples were used. Patients G and H died within 36 and 42 months of infection, respectively. Patient H received therapy with AZT at 30 months after infection.

We use the sequence, a part of the gp120 region including the V3 region whose mutation rate is particularly high in HIV.

The number of the sequence data observed from the eight patients are listed in Table 2.

Table 2. The number of sequences used in this paper.

Patient A	month 0	month 13	month 22	month 33	month 46	month 59
Number of data collected	8	7	9	9	9	8
Number of data used	6	7	7	5	6	6

Patient B	month 0	month 11	month 23	month 35	year 44	year 56
Number of data collected	11	6	6	6	7	8
Number of data used	7	3	4	4	4	4

Patient C	year 0	year 3	year 4	year 5	year 6	year 7
Number of data collected	1	15	11	23	15	13
Number of data used	1	15	11	23	15	13

Patient D	month 0	month 20	month 36	month 42
Number of data collected	5	2	4	3
Number of data used	5	2	4	3

Patient G	month 9	month 13	month 16	month 22	month 26
Number of data collected	11	8	8	8	7
Number of data used	10	8	6	8	7

Patient H	month 3	month 9	month 15	month 18	month 21	month 27	month 31	month 35
Number of data collected	6	6	7	5	7	7	5	2
Number of data used	5	6	7	4	3	7	5	2

Patient J	month 8	month 21	month 25	month 37	month 48	month 54
Number of data collcted	12	8	9	10	13	6
Number of data used	11	7	7	7	6	6

Patient K	month 3	month 9	month 19	month 23	month 28	month 34	month 47
Number of data collcted	13	10	10	7	11	11	10
Number of data used	8	9	10	7	10	10	6

4.2. Method

We used the nucleotide sequences having the same length for each patient. For example, in the primary stage of patient A, we used 6 data out of 8 data because the length of six data is 276 and that of other two is 183. Moreover, in order to carry out our analysis, first we translate the nucleotide sequences of HIV collected from the patients into the amino acid sequences. Our analysis is done in the following two cases (I) and (II)..

(I) In order to compare the genome sequences of HIV in successive months (years), the entropy evolution rate is computed for the sequences obtained at one stage (month or year) with respect to those obtained at the next stage (month or year) (we call it the entropy evolution rate for each month), and we examine the variation by means of the entropy evolution rates for each stage (month or year) and the standard deviation for each stage (month or year).

(II) In order to check the variation of HIV from the primary stage, we compute the entropy evolution rate for the sequences of each stage (month or year) w.r.t. the primary stage (month or year) (we call it the entropy evolution rate for the primary stage). Similarly as the case (I), we examine the variation (mutation) rate with the mean of the entropy evolution rates for the primary stage (month or year) and their standard deviations.

As an example, we explain how to compute the entropy evolution rate and others mentioned above in (I) and (II) for patient A. From Table 2, the number of genome sequences for patient A are as follows: n=6 (month 0), n=7 (month 13), n=7 (month 22), n=5 (month 33), n=6 (month 46), n=6 (month 59). For the case (I), we compute every entropy evolution rate for the aligned sequences in successive stages (month or year), for instance, $\rho(\mathcal{A}_i^{33}, \mathcal{A}_j^{46})$ $(i = 1,\cdots,5.\ j = 1,\cdots,6)$ for the sequence $\mathcal{A}_i^{33}$ of the thirty-third month and the sequence $\mathcal{A}_j^{46}$ of the forty-sixth month. Then we compute their mean value given by

$$\bar{\rho}(\mathcal{A}^{33},\mathcal{A}^{46}) \equiv \frac{\sum_{i=1}^{5}\sum_{j=1}^{6}\rho(\mathcal{A}_i^{33},\mathcal{A}_j^{46})}{30}$$

which enables us to examine the variation of HIV. In the same way, we compute $\bar{\rho}(\mathcal{A}^{0},\mathcal{A}^{13})$, $\bar{\rho}(\mathcal{A}^{13},\mathcal{A}^{22})$, $\bar{\rho}(\mathcal{A}^{22},\mathcal{A}^{33})$, $\bar{\rho}(\mathcal{A}^{33},\mathcal{A}^{46})$, $\bar{\rho}(\mathcal{A}^{46},\mathcal{A}^{59})$. The standard deviation of the entropy evolution rate for month 33 and month 46 is defined as follows:

$$\sqrt{\frac{\sum_{i=1}^{5}\sum_{j=1}^{6}\left\{\rho(\mathcal{A}_i^{33},\mathcal{A}_j^{46})-\bar{\rho}(\mathcal{A}^{33},\mathcal{A}^{46})\right\}}{30}}$$

For the case (II), we compute the mean entropy evolution rates for every sequence of each stage with respect to that of the primary stage. For instance, the mean entropy evolution rate for the fifty-ninth month w.r.t. the primary stage is given by

$$\bar{\rho}(\mathcal{A}^{0},\mathcal{A}^{59}) \equiv \frac{\sum_{i=1}^{6}\sum_{j=1}^{6}\rho(\mathcal{A}_i^{0},\mathcal{A}_j^{59})}{36}$$

We similarly compute $\bar{\rho}(\mathcal{A}^{0},\mathcal{A}^{13})$, $\bar{\rho}(\mathcal{A}^{0},\mathcal{A}^{22})$, $\bar{\rho}(\mathcal{A}^{0},\mathcal{A}^{33})$, $\bar{\rho}(\mathcal{A}^{0},\mathcal{A}^{46})$, $\bar{\rho}(\mathcal{A}^{0},\mathcal{A}^{59})$ and their standard deviations.

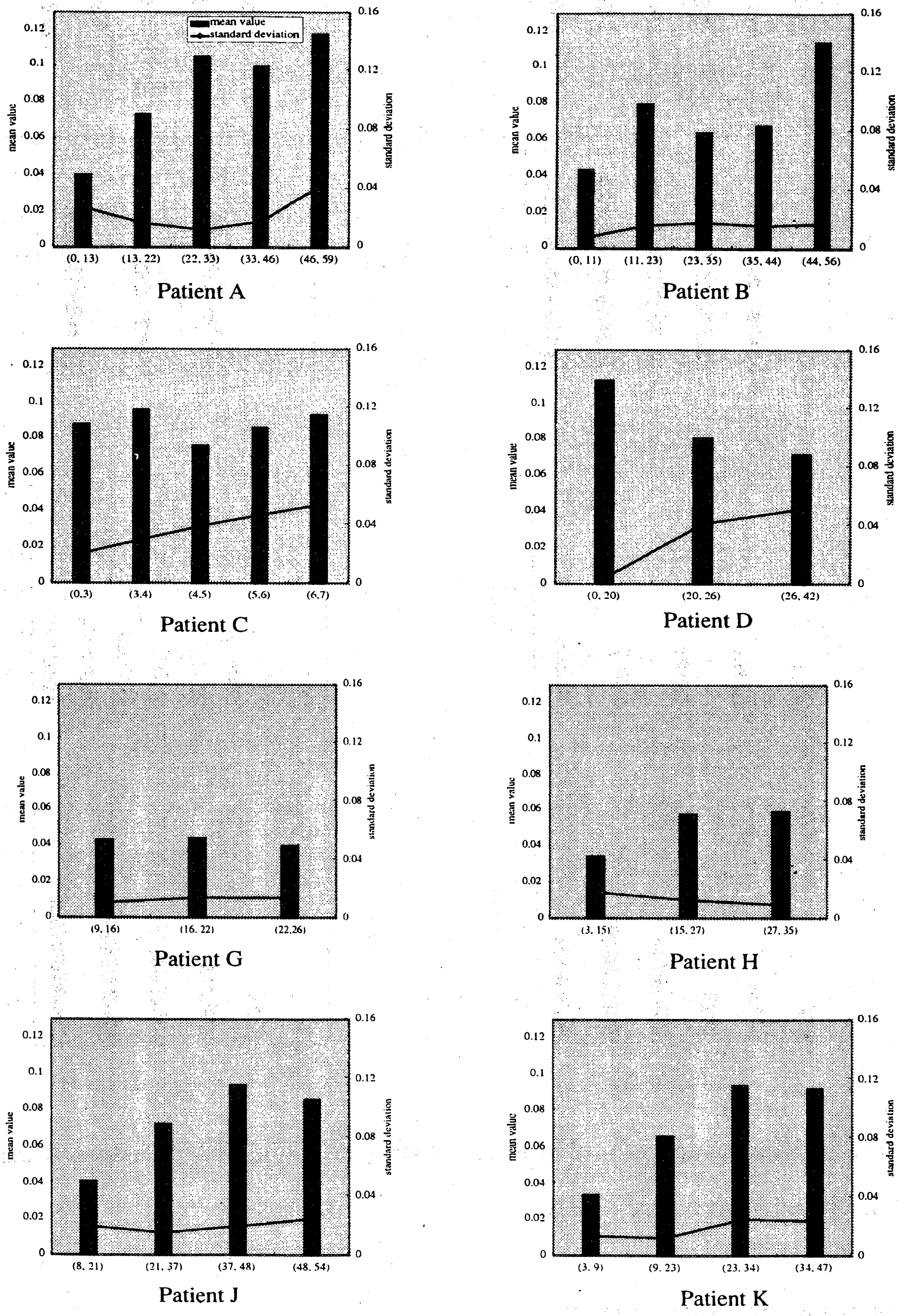
All eight patients are examined with these quantities, and our results are shown in the next subsection.

Here we note that we should align the sequences to compute the entropy evolution rate, and the alignment is done by a method developed in [8, 9].

4.3. Results

The following figures (Fig.1) is the results of the mean entropy evolution rates and the standard deviations for each stage (month or year). Here (i, i+1) denotes the (i+1)-th stage w.r.t. i-th stage and the mean value is the mean entropy evolution rate. The "i" of the seven patients except patient C indicates the months after infection. For patient C, it does how the years after infection. We take the data of the four patients G, H, J, K with almost one year interval so that we can analyze them in the same standing position as patient A, B, C.

Fig.2 shows the results of the mean entropy evolution rate for the primary stage and their standard deviations, so that (0, i) denotes the i-th stage w.r.t. the primary stage.



Patient A

Patient B

Patient C

Patient D

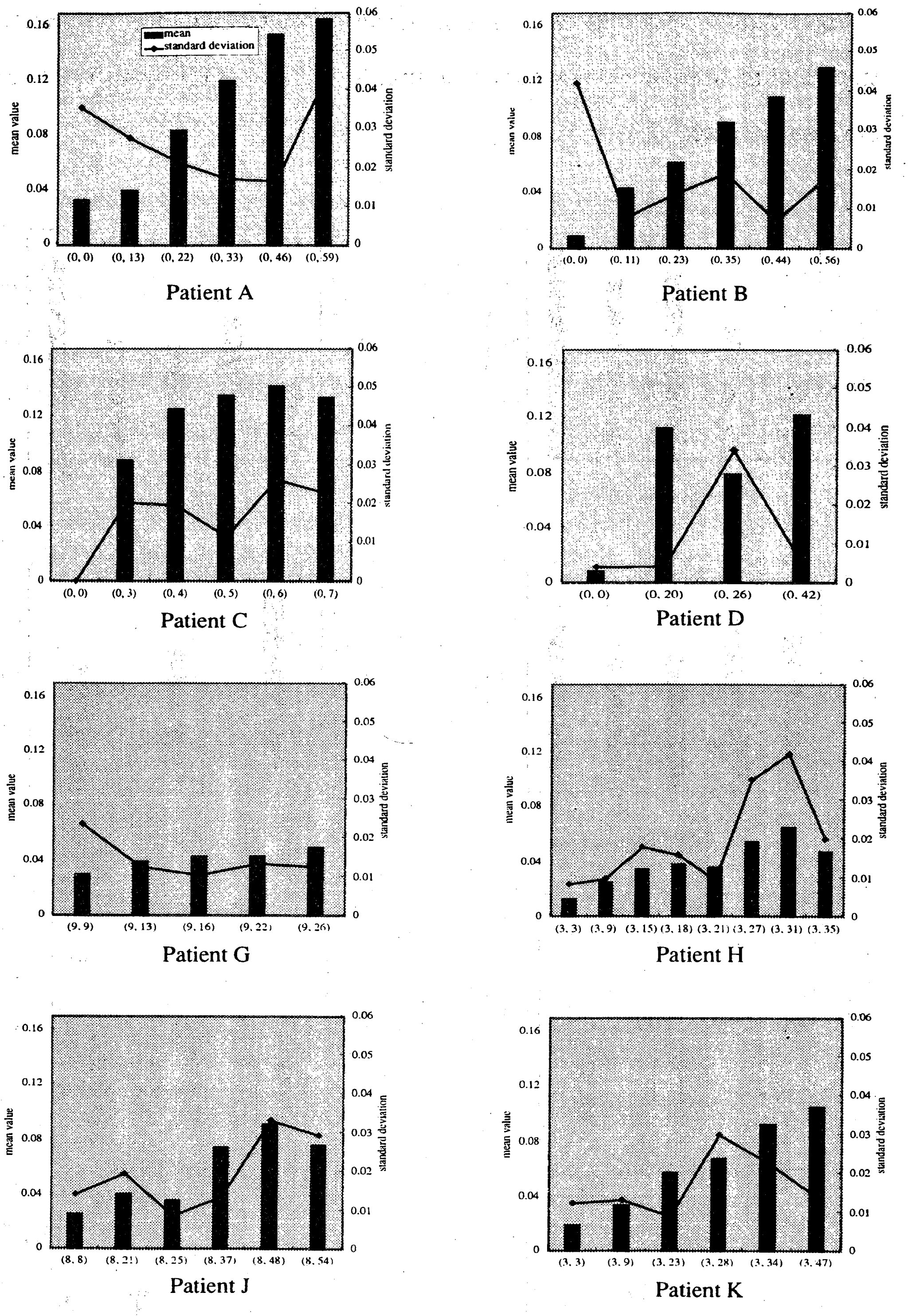
Patient G

Patient H

Patient J

Patient K

Fig.1. Mean entropy evolution rate (bars) and standard deviation (lines) for each stage (month or year).



Patient A

Patient B

Patient C

Patient D

Patient G

Patient H

Patient J

Patient K

Fig.2. Mean entropy evolution rate (bars) and standard deviation (lines) measured from primary stage.

5. PHYLOGENETIC TREES

5.1. Writing algorithm

Phylogenetic tree represents the process of evolution of organisms, so that it can be used it to presume how to branch and classfy the organisms.

The genetic difference between organism a and b is expressed by the entropy evolution rate $\rho(a, b)$ defined in the section 3. In order to write phylogenetic trees, we need genetic matirx. Let us consider n organisms $(A_1, A_2, \cdots, A_n)$. The genetic difference between i-th and j-th of these n organisms is given by $D_{ij} = \rho(A_i, A_j)$ $(i, j = 1, \cdots, n)$, which makes a $n \times n$ matrix $D = (D_{ij})$, called "genetic matrix" for n organisms $(A_1, A_2, \cdots, A_n)$.

Based on the genetic matrix, there exist several methods to write phylogenetic tree such as UPG, MF, NJ. Here we first discuss an axiomatic approach , namely mathematical structure writing the trees. Then we show that UPG and NJ methods are special expressions of our axiomatic setting.

Take a set G of n species to observe phylogenetic relation as

$$G = \{s_1, s_2, \cdots, s_n\}$$

and let $G_s (\subset 2^G)$ be the set of all groups construting trees (a subset of the power set of G) satisfying the following conditions:

$$X \cap Y = \varnothing \text{ for any } X, Y \in G_s, \text{ and } \bigcup_{X \in G_s} X = G$$

The difference $d_{X \circ Y}$ between two groups X and Y, including two organisms, should satisfy the following conditions:

(1) $d_{X \circ Y} \geq 0$ (2) for any $X \in G_s$ there exists $Y \in G_s$ attaining the minimum value of $d_{X \circ Y}$ to make a new group with X.

That is, we have to find a proper operation $\circ$ providing the difference between X and Y, in other words, we have to set how to compute the difference between two elements in G_s.

Unweighted Pair Group Clustering Method

UPG method is a method introduced by Sokai and Michener in 1958. In the UPG method, the pair having the smallest difference makes the first group and compute the difference between two groups according to the following simple average: The differenece $d_{X \circ Y}$ between two groups $X = \{x_1, x_2, \cdots, x_k\}$ and $Y = \{y_1, y_2, \cdots, y_l\}$ is given as

$$d_{X \circ Y} = \frac{1}{k \cdot l} \sum_{i=1}^{k} \sum_{j=1}^{l} D_{x_i y_j}$$

where $D_{x_i y_j}$ is the (i,j)-element of the genetic matrix.

Compute all differences for all two groups of G_s then two groups giving a minimum difference make a new group. Repeat this procedure untill all organisms forms one group.

Neighbor Joining Method

NJ method was introduced by Saitou and Nei, in which the evolution rate is supposed to reflect the length of branch. When a group $X = \{x_1, x_2\}$ connects to a group $Y = \{y_1, y_2, \cdots, y_n\}$ as a neighbor, the difference between group X and group Y (the length of the branch connecting X and Y) is given as

$$\begin{aligned} d_{X \circ Y} &= d_{(x_1, x_2)(y_1, y_2, \cdots, y_n)} \\ &= \frac{1}{2 \cdot n} \sum_{i=1}^{2} \sum_{j=1}^{n} D_{x_i y_j} + \frac{1}{2} D_{x_1 x_2} + \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=1}^{n} D_{y_i y_j} . \end{aligned}$$

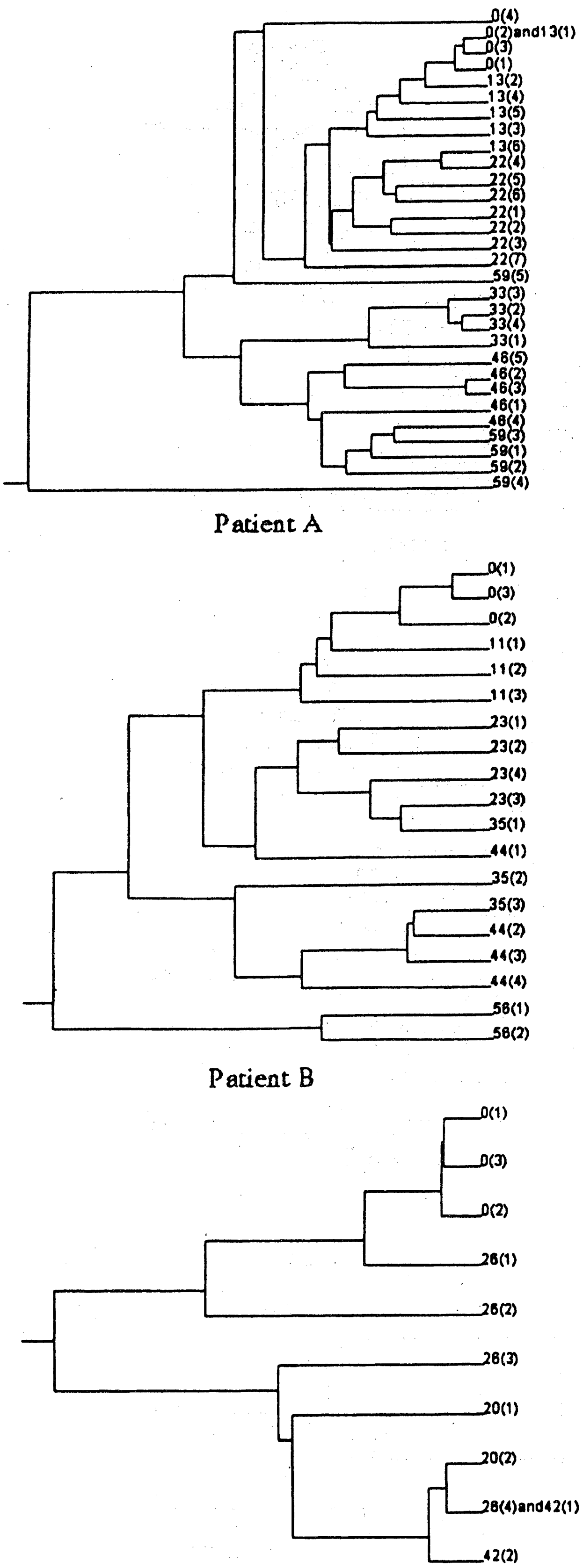
Moreover the length (difference) of a branch between an element x_1 of X and the node C_X of X (point bundling all elements of X) is calculated as

$$L_{x_1, C_X} = \frac{1}{2} D_{x_1 x_2} + \frac{1}{2 \cdot n} \sum_{j=1}^{n} \left(D_{x_1 y_j} - D_{x_2 y_j} \right).$$

We first compute the difference of each pair of organisms by using the genetic matrix, and we determine a pair (first neighbor) giving the minimum value of such differences. Then we compute the difference indicating the total length of a branch in phylogenetic tree as above to find other neighbors. We continue this procedure until all organisma forms one neighbor, and we determine the branching point of the ancestor of all organisms as the center of the longest branch of the whole tree.

5.2. Results

The following figures (Fig.3, Fig.4 and Fig.5) are the phylogenetic trees written by each method. We here show the philogenetic trees of patients A, B, D.

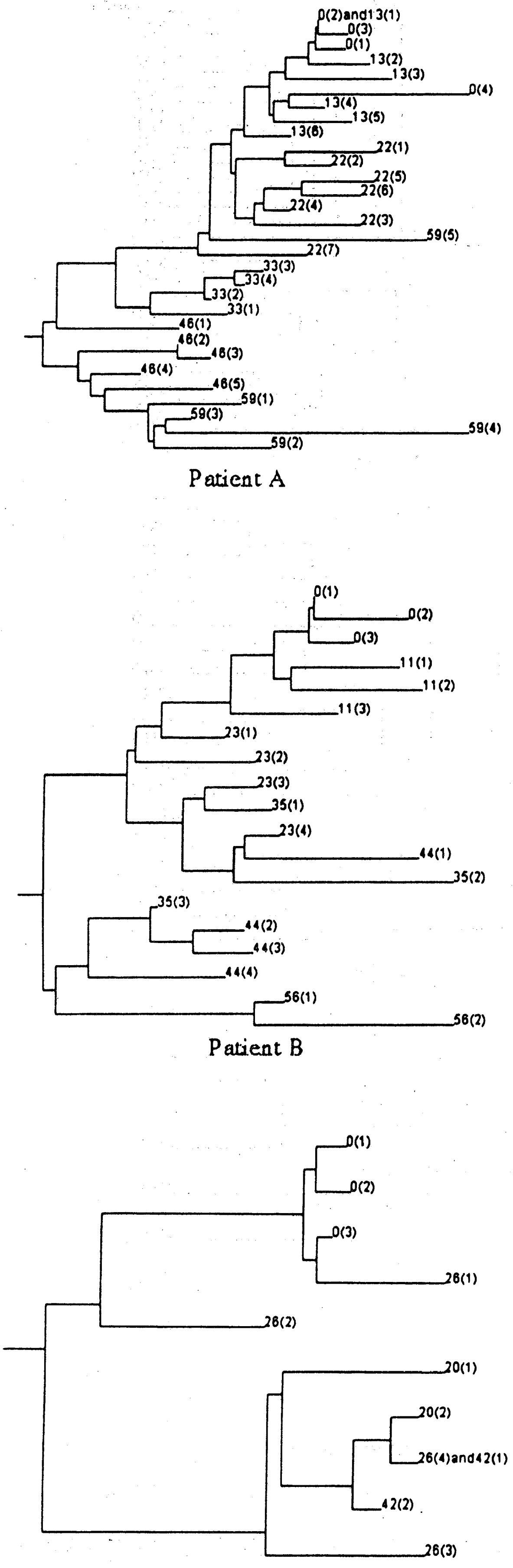


Patient A

Patient B

Patient D

Fig.3. Phylogenetic trees by UPG method. Here i(j) denotes that "i" indicates months (stage) after the time of primary infection and "j" indicates the number of sequence.

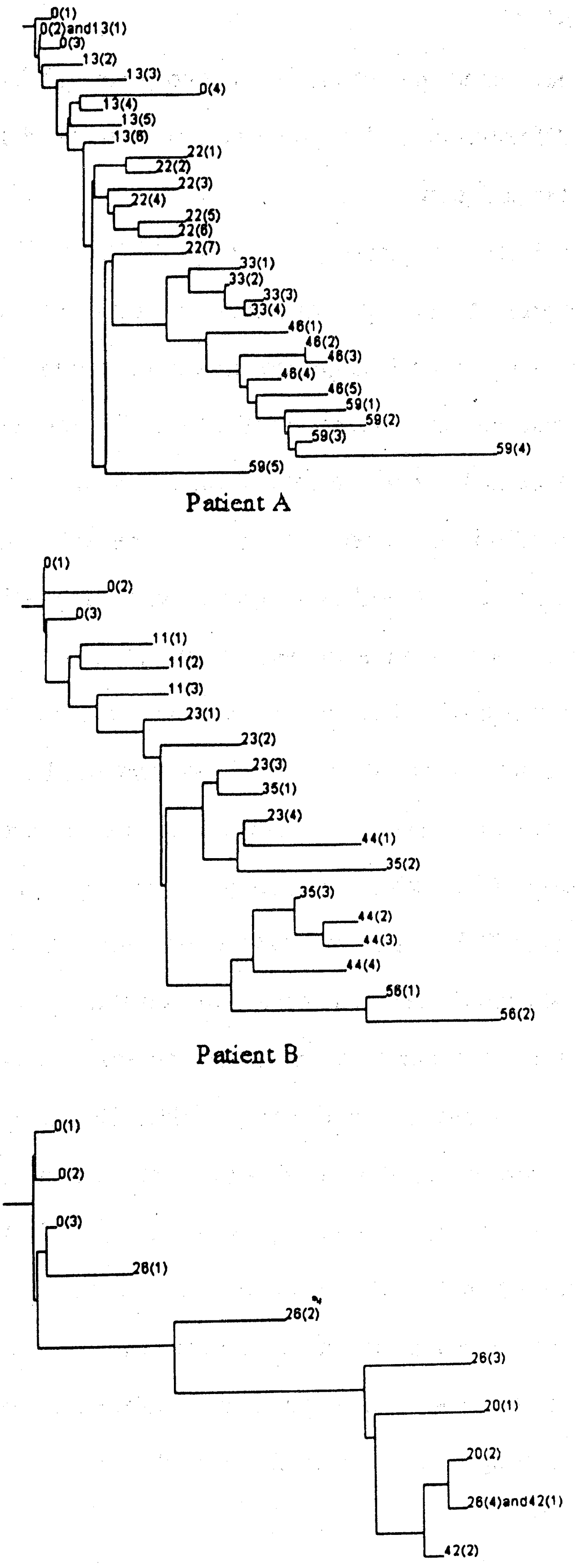


Patient A

Patient B

Patient D

Fig.4. Phylogenetic trees by NJ method.



Patient A

Patient B

Patient D

Fig. 5. Phylogenetic trees measured from primary stage (month or year) by NJ method.

6. DISCUSSION

Our study have been mainly carried in two parts; (1)Variation of HIV by EER and (2)Evolution of HIV by trees, so that we discuss main results concerning the above two.

(1)Patient B was diagnosed as having AIDS at 55 months after the primary infection. According to the result of the mean entropy evolution rate (m-EER for short) for each stage, the variation of the m-EER for patient B is met the second extreme increase at that time. The variation of the m-EER (Fig.1) for patient B is considered as a fundamental pattern of the outbreak of AIDS. Based on this pattern, we may say the following conclusions for other patients. Patient A will be diagnosed as having AIDS in a few years because the second extreme increase seems starting. Similarly, patient C will have an attack of AIDS soon, because the second moderate increase is occurred. The patterns of patient J and K are very similar to that of patient A before the second increase start. Therefore, we can expect that patient J and K will develope AIDS when the second increase of m-EER is met, namely, the sequence of HIV has many variants as patient B. Patient D has few number of data and the sequences are collected at irregular interval, so that I merely say a few comments. Patient D may possibly increase here after. The CD4 count of patient G and H are less than 200 in 16 and 21 months, respectively. Moreover, they died within 36 and 42 months from the estimated time of primary infection, respectively [5]. They might have been infected with HIV before the estimated time of primary infection, because the change of their m-EER is almost same and the values of their m-EER are small in comparison with other patients. It means that the variation of the sequence is getting smaller when the patient is in near death after having AIDS.

According to the results of the mean entropy evolution rate measured from the primary stage, we can classify the eight patients into three categories; the first one is that of patiens A, B, C, K, the second is G, H and the third is D, J. For the patients in the first category, their values of m-EER clearly increases as shown in Fig.2. For the patiens in the second category, the values of m-EER are small and their increase is moderate. For the patiens in the last category, the values of m-EER fluctuate. This consequence agrees with the reports [2, 3, 4, 5] concerning the change of the CD4 count for the patients. That is, the gradual decrease of the CD4 count for the patients except D and J is strongly related to the increase of the m-EER for the primary year, and the CD4 counts of the patient D and J fluctuate. This result means that there exists a positive correlation between the m-EER for the primary year and the CD4.

We merely note that the standard deviation shows how many different HIV exist in each stage (year).

From our analysis, we may conclude that the mean entropy evolution rate can be a measure of the variation of HIV and the outbreak of AIDS as the CD4 count. This part of study for HIV is based on a paper [10].

(2)From the trees written by UPG, each patient has the following distinct clusters;

A=[{0(4)}, {0(1)~0(3), 13(1)~13(5)}, {13(6), 22(1)~22(7)}, {59(5)}, {33(1)~33(4)}, {46(2), 46(3), 46(5)}, {46(1), 46(4), 59(1)~59(3)¥}, {59(4)}]

B=[{0(1)~0(3), 11(1)~11(3)}, {23(1)~22(4), 35(1)}, {44(1)}, {35(2)}, {35(3), 44(2)~44(4)}, {56(1), 55(2)}]

D=[{0(1)~0(3), 26(1), 26(2)}, {20(1), 20(2), 26(3), 26(4), 42(1), 42(2)}]

so that D can be considered in the earlier stage compared with A and B. Further B has peculiar sequences at 56 months after the appearance of AIDS, which is definitely shown in Fig.3.

The tree from NJ methods are of two types, one of which is the tree due to starting from a common ancestor and another is due to the primary stage (sequence). We have the similar clusters as UPG, but we can understand when a HIV in a ceratin stage appears since the length of branch is proportional to the evolution time. Seeing from the ancestor, some HIV may go back to the older HIV like at 33 month of B. In any tree, the sequences of B at 56th month, particularly 56(2), are differently branched from others.

Phylogenetic tree can not be a measure indicating the symptoms of patient, however the branching complexity of the tree might help us reading the situation of patient.

References

[1] M.Ohya, Information theoretical treatment of genes, The Trans. of The IEICE, Vol. E 725, pp.556-560 (1989)

[2] T.W.Wolfs, G.Zwart, M.Bakker, M.Valk,C.Kuiken, and J.Goudsmit, Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution, Virology 185, pp.195-205 (1991)

[3] Holmes, L.Q.Zhang, P.Simmonds, C.A. Ludlam, and A.J. L.Brown, Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient, Evolution, Vol. 89, pp.4835-4839 (1992)

[4] T.McNearney, Z.Hornickova, R.Markham, A.Bird-well, M.Arens, A.Saah, and L.Ratner, Relationship of human immunodeficiency virus type 1 sequence hetero-geneity to stage of disease, Medical Sciences, Vol. 89, pp.10247-10251 (1992)

[5] Steven M. Wolinsky, Bette T. M.Korber, Avidan U. Neumann, Michael Daniels, Kevin J. Kunstman, Amy J. Whetsell, Manohar R. Furtado, Yunzhen Cao, David D. Ho, Jeffrey T. Safrit, Richard A. Koup, Adaptive Evolution of Human Immnodeficency Virus-Type 1 During the Natural Course of Infection, SCIENCE, Vol.272, pp.537-542 (1996)

[6] J.D.Watson, M.Gilman, J.Witkowski, and M.Zoller, Recombinant DNA 2nd Edition, Freeman and Company(1993)

[7] J.J.de Jong, J.Goudsmit, W.Keulen, B.Klaver, W.Krone, M.Tersmette, and A.de Ronde, Human immunodeficiency virus type 1 clones chimeric for

the envelope V3 domain differ in syncytium formation and replication capacity, Journal of Virology, 66, pp.757-765 (1992)

[8] M.Ohya and Y.Uesaka, Amino acid sequences and DP matching: A new method for alignment, Information Sciences 63, pp.139-151 (1992)

[9] S.B.Needleman and C.D.Wunsch, A general method applicable to search for similarities in the amino acid sequence of two proteins, J.Mol.Biol., pp.443-453 (1970)

[10] K.Sato, S.Miyazaki and M.Ohya, Analysis of HIV by entropy evolution rate, Amino Acids 14: pp.343-352 (1998)