

A new method for analyzing H5N1 avian influenza virus

Yanqiu Liu · Yusen Zhang

Received: 23 September 2009 / Accepted: 24 November 2009 / Published online: 11 December 2009
© Springer Science+Business Media, LLC 2009

Abstract In this paper a novel method for phylogenetic analysis of H5N1 avian influenza virus has been proposed. At first we provide a mapping of virus protein sequence. Based on this mapping, we propose a new distance measure and make use of the corresponding similarity matrix to construct phylogenetic tree without requiring multiple alignment. As an application, we construct phylogenetic tree for 123 species of H5N1 avian influenza virus. The phylogeny obtained is generally consistent with evolutionary trees constructed in previous studies.

Keywords Amino acid · H5N1 avian influenza virus · Similarity matrix · Phylogenetic tree

1 Introduction

Influenza A virus is a negative-stranded RNA virus with eight genomic segments encoding RNA polymerases (PB2, PB1, PA), hemagglutinin (HA), nucleoprotein (NP), neuraminidase (NA), matrix protein (MP) and non-structural protein (NS). They code protein PB1, PB2, PA, PB1-F2, HA, NP, NA, M1, M2, NS1 and NS2, respectively. A total of 16 HA (hemagglutinin) and 10 NA (neuraminidase) subtypes have been reported [2, 10]. All of these subtypes come from avian species [6]. New influenza viruses and genotypes continuously emerge due to the frequent evolutionary events including genetic reassortment, recombination and mutation. These emerging

Y. Liu · Y. Zhang (✉)

School of Mathematics and Statistics, Shandong University at Weihai, 264209 Weihai, China
e-mail: zhangys@sdu.edu.cn

influenza viruses have caused three pandemics, including H1N1 in 1918, H2N2 in 1957 and H3N2 in 1968 [14]. Influenza genotype analyses reflect influenza viral evolutionary footprints, and thus are critical for preparing a strategy to prevent and control influenza epidemics and pandemics.

Although some bench laboratory methods have been generated for genotyping [3], these types of approaches can be laborious and time consuming, and thus are not efficient for identifying genetic reassortment events, particularly when a pandemic occurs. A robust and efficient computational genotyping system is crucial; however, mathematical analyses are not trivial tasks. In this study, we present a novel approach for influenza virus protein analyses based on a four-letter model of the 20 amino acids and under the assumption that influenza A viruses form a niche in response to the environmental pressure. We apply this method in H5N1 influenza genotyping and the results are discussed.

2 Methods

2.1 The four-letter model of amino acids

A protein sequence consists of 20 different amino acids, so it can be expressed as a series of 20 kinds of letters. But it is not easy to find a visual representation for the primary sequence of a protein, because the sequence consists of 20 different letters, and the number of each kind of letters may vary considerably. Recently, much effort has been made by considering minimalist models with a few types of amino acid residues to simplify the natural set of residues of 20 types for better physical understanding and practical purposes. In these models the compositions are much simpler than the real ones. Following the introduction of John A. Bergera [1], the 20 amino acids can be classified into four groups:

$$\left\{ \begin{array}{l} \text{group - I} = (A, I, L, M, F, P, W, V), \\ \text{group - II} = (G, N, C, Q, S, T, Y), \\ \text{group - III} = (E, D), \\ \text{group - IV} = (K, R, H). \end{array} \right.$$

Each group contains some residues which interact with others in a similar way. Moreover, the four groups have representative residues, they are *A*, *G*, *E* and *K*, respectively. Thus a protein primary sequence can be reduced into a four-letter sequence by substituting (or replacing) each letter with its representative letter. For example, the sequence of *MERIKELRDLMQSRTREILTKTTVDHMAIIKKYTS*, the first 36 amino acid residues of the AIV protein sequence of (A/Goose/Guangdong/1/96(H5N1)), are *AEKAKEAKEAAGGGKKGKEAAGKGGAEKAAAKKGGG*. The four-letter sequence may be regarded as a coarse-grained description of the protein primary sequence. Via comparisons of the reduced sequences it will be easier to understand the biological function of various kinds of amino acid residues.

2.2 The mapping of virus protein sequences

For a given four-letter sequence, we give a curve mapping of the protein sequence. First we assign one letter as follows:

$$\begin{aligned}
 (\sqrt{2}, \sqrt{2}, \sqrt{2}) &\rightarrow A & (-1, 0, 0) &\rightarrow G \\
 (0, 1, 0) &\rightarrow K & (0, 0, -1) &\rightarrow E
 \end{aligned}$$

So that we can reduce a given sequence into a series of nodes $P_0, P_1, P_2 \dots P_N$, whose coordinates X_i, Y_i and Z_i ($i = 0, 1, 2, \dots, N$, where N is the length of the protein sequence being studied) satisfy:

$$\begin{cases}
 X_i = \sqrt{2}A_i - G_i \\
 Y_i = \sqrt{2}A_i + K_i \\
 Z_i = \sqrt{2}A_i - E_i
 \end{cases}$$

where A_i, K_i, G_i and E_i are the cumulative occurrence numbers of A, K, G and E , respectively, in the subsequence from the first base to the i -th base in the sequence. We define $A_0 = K_0 = G_0 = E_0 = 0$.

We called the corresponding plot set be characteristic plot set. The curve connecting all plots of the characteristic plot set in turn is called 3DD-Curve (3D curve) [16–19], this representation has been proved to be reasonable and nondegeneracy [9, 17].

2.3 Constructing phylogenetic tree

For any virus protein sequence, we have a set of points (X_i, Y_i, Z_i) , ($i = 1, 2, 3, \dots, N$, where N is the length of the protein sequence). The coordinates of the geometrical center of the points, denoted by X^0, Y^0 and Z^0 , may be calculated as follows:

$$\begin{cases}
 X^0 = \frac{1}{N} \sum_{i=1}^N X_i \\
 Y^0 = \frac{1}{N} \sum_{i=1}^N Y_i \\
 Z^0 = \frac{1}{N} \sum_{i=1}^N Z_i
 \end{cases}$$

We construct a covariance matrix CM , where,

$$CM = \begin{bmatrix}
 CM_{xx} & CM_{xy} & CM_{xz} \\
 CM_{yx} & CM_{yy} & CM_{yz} \\
 CM_{zx} & CM_{zy} & CM_{zz}
 \end{bmatrix}$$

And the entries of covariance matrix are defined:

$$\begin{cases} CM_{xx} = \frac{1}{N} \sum_{i=1}^N (X_i - X^0)(X_i - X^0) \\ CM_{xy} = \frac{1}{N} \sum_{i=1}^N (X_i - X^0)(Y_i - Y^0) = CM_{yx} \\ CM_{xz} = \frac{1}{N} \sum_{i=1}^N (X_i - X^0)(Z_i - Z^0) = CM_{zx} \\ CM_{yy} = \frac{1}{N} \sum_{i=1}^N (Y_i - Y^0)(Y_i - Y^0) \\ CM_{yz} = \frac{1}{N} \sum_{i=1}^N (Y_i - Y^0)(Z_i - Z^0) = CM_{zy} \\ CM_{zz} = \frac{1}{N} \sum_{i=1}^N (Z_i - Z^0)(Z_i - Z^0) \end{cases}$$

The above six numbers give a quantitative description of a set of point (X_i, Y_i, Z_i) , $(i = 1, 2, \dots, N)$ scattering in a three-dimensional space. Obviously, the matrix is a real symmetric 3×3 one. There are three real eigenvalues for a matrix CM [8]. So that there are one geometrical center and three eigenvalues corresponding a virus sequence. For an Influenza A virus, there are eight segments: PB2, PB1, PA, HA, NP, NA, MP and NS. They can code protein PB1, PB2, PA, PB1-F2, HA, NP, NA, M1, M2, NS1 and NS2, respectively. Based on the previous studies [10, 12, 15] and the feature of the segments protein sequences of H5N1 AIV, eight segment protein sequences are integrated to interpret the genotypes for each virus is reasonable. Here we only consider 123 H5N1 AIVs with eight full-length segments PB2, PB1, PA, HA, NP, NA, M1 and NS1. So, there are eight geometrical centers and twenty four eigenvalues corresponding a virus protein sequence.

Let $(X_i^{0,t}, Y_i^{0,t}, Z_i^{0,t})$, $i = 1, 2, \dots, M$, $t = 1, 2, \dots, 8$ denote the geometrical center of the curve belonging to segment t of strain i , and $\lambda_1^{i,t}, \lambda_2^{i,t}, \lambda_3^{i,t}$ denote the three eigenvalues of matrix $CM_{i,t}$, $i = 1, 2, \dots, M$ ($t = 1, 2, \dots, 8$), corresponding to segment t of strain i , then we define E_i^t as:

$$E_i^t = \sum_{k=1}^3 \left| \lambda_k^{i,t} \right|$$

We construct a 4-component vector

$$\eta = (X_i^{0,t}, Y_i^{0,t}, Z_i^{0,t}, E_i^t)$$

Then we get a correspondence between the protein sequences and 4-component vectors $(X_i^{0,t}, Y_i^{0,t}, Z_i^{0,t}, E_i^t)$, So $(X_i^{0,t}, Y_i^{0,t}, Z_i^{0,t}, E_i^t)$ can characterize the corresponding protein sequences. Comparison between protein sequences becomes comparison between these 4-component vectors. And the distance d_{ij}^t between the segment t belonging to strains i and j is defined by:

$$d_{ij}^t = \sqrt{(X_i^{0,t} - X_j^{0,t})^2 + (Y_i^{0,t} - Y_j^{0,t})^2 + (Z_i^{0,t} - Z_j^{0,t})^2 + (E_i^t - E_j^t)^2}$$

where $i, j = 1, 2, \dots, M$, and M is the total number of all strains ($M = 123$, here). Then we obtain a real $M \times M$ symmetric matrix D^t whose elements are d_{ij}^t .

Summing over t , we can obtain the overall distance d_{ij} between the strains i and j :

$$d_{ij} = \sum_{t=1}^8 d_{ij}^t, \quad i, j = 1, 2, \dots, M$$

Accordingly, a real symmetric $M \times M$ matrix D is obtained whose elements are d_{ij} and used to reflect the evolutionary distance between the strains i and j . And we can present the dendrogram tree base on linkage cluster analysis using Euclid distances of these vectors which consist of a real symmetric $M \times M$ matrix D . In Fig. 1, we present the phylogenetic tree belonging to 123 strains' HA, NA and NP according to D^t , and in Fig. 2 we present the phylogenetic tree belonging to 123 H5N1 AIVs protein sequences according to D .

2.4 Datasets

The influenza datasets were downloaded from influenza virus resource database at GenBank. The final analyses were focused on 123 H5N1 AIVs with eight full-length protein segments (Table 1).

2.5 Results

Based on the feature of the segmented RNA genome, eight segment genotypes are integrated to interpret the genotypes for each virus. In this study, our analyses only focus on the H5N1 AIV protein sequences using eight segment of it. The phylogenetic tree in Fig. 2 demonstrates the evolution pathways of 123 H5N1 AIVs. All the viruses are chosen from different areas, different periods, and different species virus hosts, so they have universality. Now according to the phylogenetic tree, we analyze in detail.

Firstly, the continuous isolation of AIVs indicates that China has been the epicenter of H5N1 AIVs. In China, many genotypes were identified in the past 10 years, and many of these genotypes have become local strains interwoven with the local virus pool [4, 7], several genotypes are identified with Chinese provinces including Guangdong, Fujian, Jilin, and Guangxi and so on as shown in Fig. 2.

Secondly, from Fig. 2 we can obtain that phylogenetic relationship among H5N1 avian influenza virus closely relates to distributions of time and geography. In Fig. 2, we can obtain that these clades (77, 70, 69) (75, 73) (66, 65, 62, 61, 21) (53, 18) (79, 45, 44, 7) (49, 20, 17) (37, 71, 2) (63, 64) (97, 100, 96) (48, 39) (35, 32, 31, 8) (16, 14) (99, 98) (92, 91, 6) (86, 50) (114, 93) (103, 116, 115, 102, 101, 104, 106, 105) formed according to time, so we could deduce that time factor plays an important role in the AIVs evolution process. In generally, from Fig. 2 we find that virus strains separately existing in China, Indonesia, Vietnam, and Hong Kong are different branches in this phylogenetic tree, these results are identified with the previous studies [10, 15].

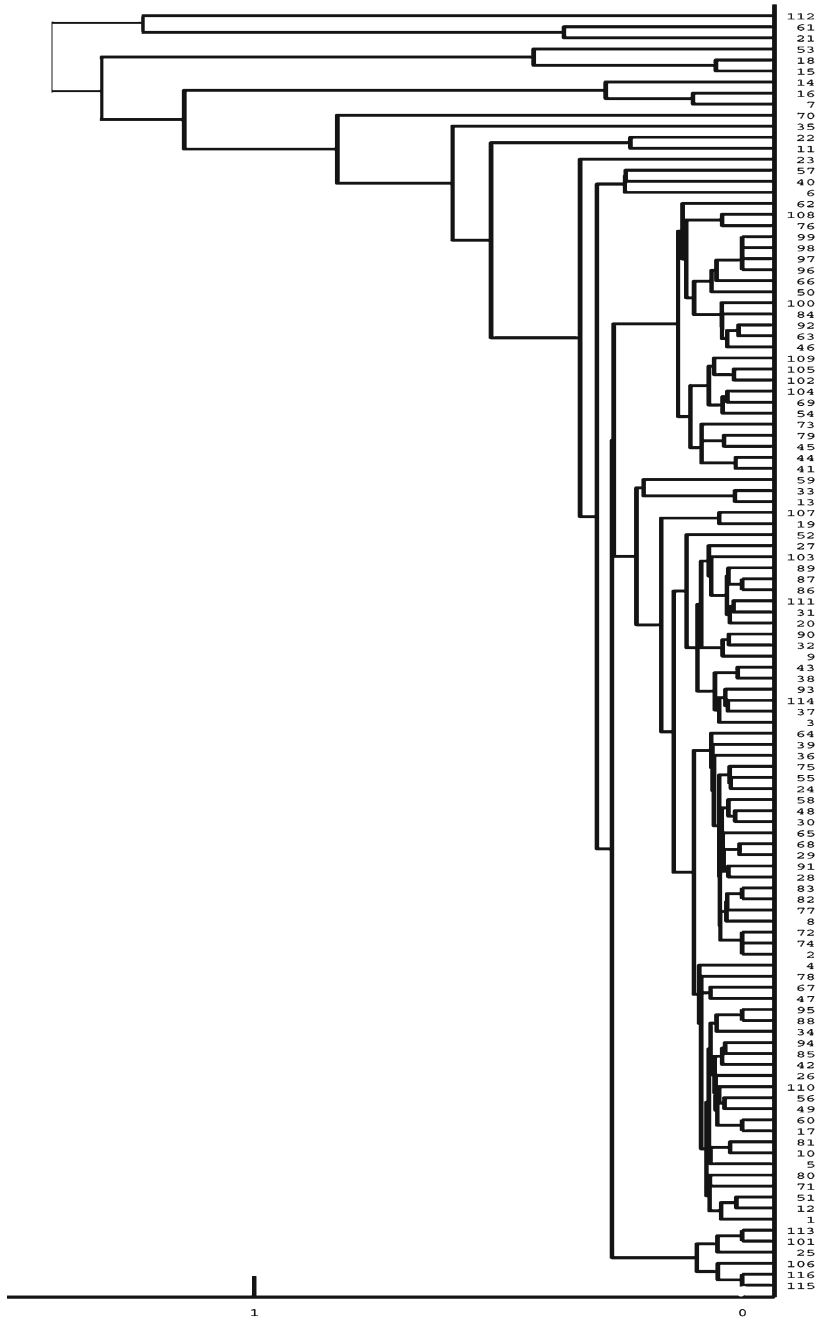


Fig. 1 Phylogenetic tree of 123 H5N1 AIVs HA, NA, NP sequences

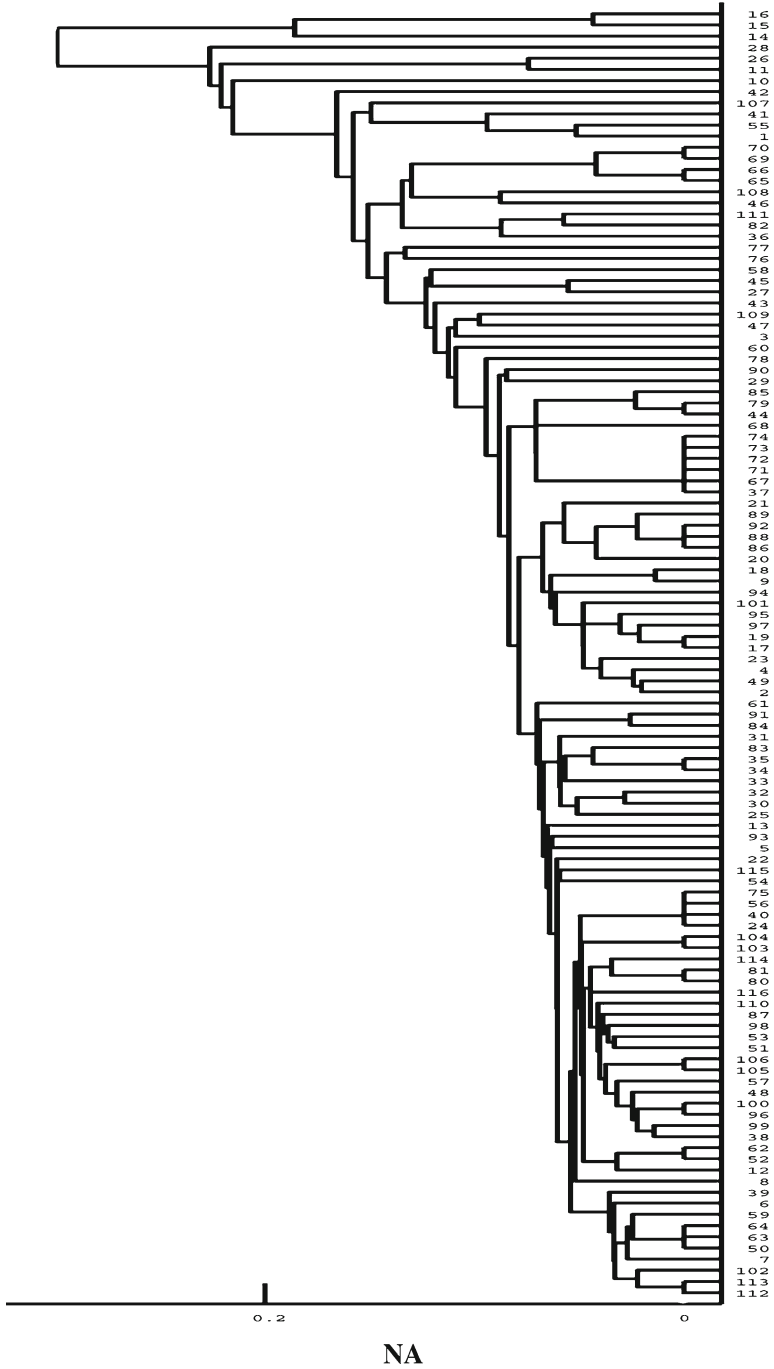


Fig. 1 continued

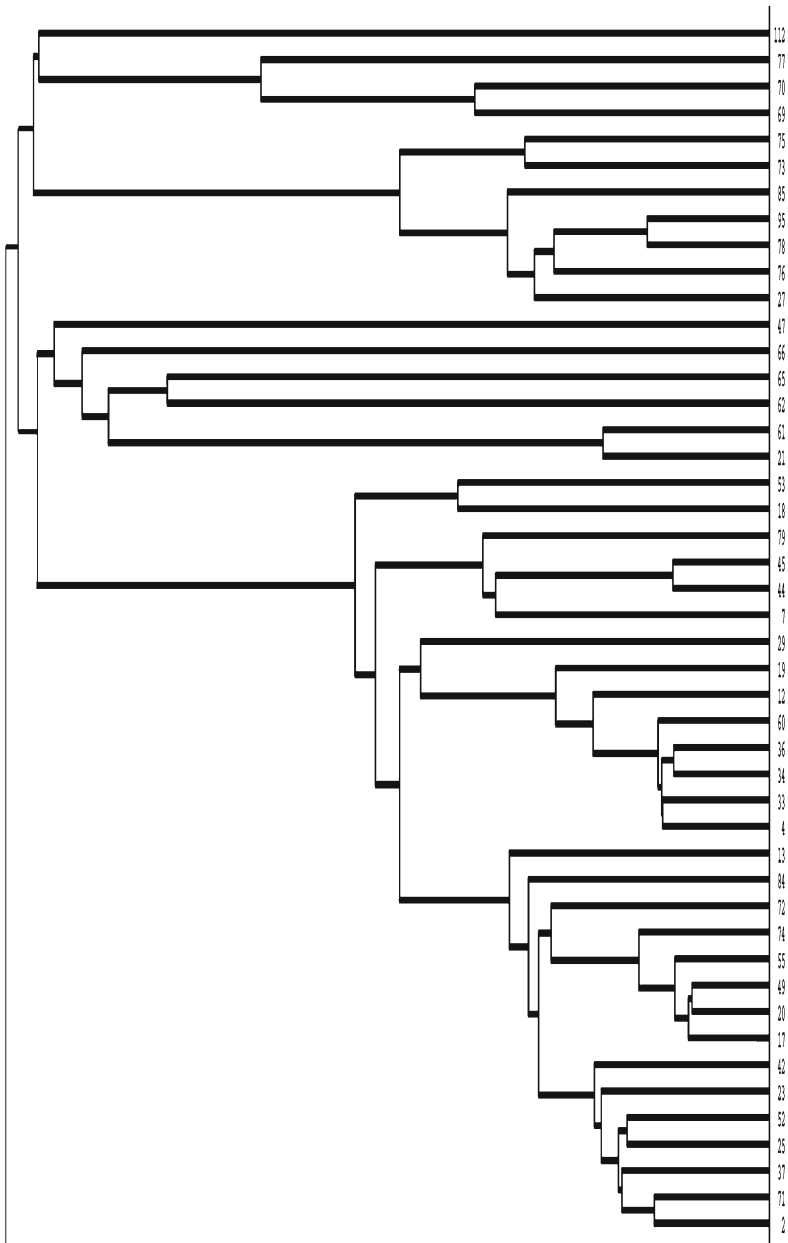


Fig. 2 Phylogenetic tree of 123 H5N1 AIVs protein sequences



Fig. 2 continued

Table 1 Database source

No	Species
1	A/Goose/Guangdong/1/96(H5N1)
2	A/duck/Guangdong/01/2001(H5N1)
3	A/duck/Guangdong/07/2000(H5N1)
4	A/duck/Guangdong/22/2002(H5N1)
5	A/chicken/Guangdong/174/04(H5N1)
6	A/duck/Guangdong/173/04(H5N1)
7	A/goose/Guangdong/xb/2001(H5N1)
8	A/chicken/Guangdong/1/2005(H5N1)
9	A/Duck/Anyang/AVL-1/2001(H5N1)
10	A/duck/China/E319-2/03(H5N1)
11	A/teal/China/2978.1/2002(H5N1)
12	A/duck/Fujian/01/2002(H5N1)
13	A/duck/Fujian/19/2000(H5N1)
14	A/babbler/Fujian/320/04(H5N1)
15	A/golden mountain thrush/Fujian/376/04(H5N1)
16	A/black bulbul/Fujian/439/04(H5N1)
17	A/duck/Fujian/9651/2005(H5N1)
18	A/duck/Fujian/10160/2005(H5N1)
19	A/chicken/Fujian/10313/2005(H5N1)
20	A/chicken/Fujian/584/2006(H5N1)
21	A/duck/Fujian/671/2006(H5N1)
22	A/duck/Guangxi/07/1999(H5N1)
23	A/duck/Guangxi/22/2001(H5N1)
24	A/swan/Guangxi/307/2004(H5N1)
25	A/duck/Guangxi/351/2004(H5N1)
26	A/goose/Guangxi/914/2004(H5N1)
27	A/chicken/Guangxi/2439/2004(H5N1)
28	A/chicken/Guangxi/2461/2004(H5N1)
29	A/goose/Guangxi/345/2005(H5N1)
30	A/quail/Guangxi/575/2005(H5N1)
31	A/chicken/Guangxi/604/2005(H5N1)
32	A/duck/Guangxi/793/2005(H5N1)
33	A/chicken/Guangxi/1212/2006(H5N1)
34	A/duck/Guangxi/1258/2006(H5N1)
35	A/duck/Guangxi/1436/2006(H5N1)
36	A/goose/Guangxi/1458/2006(H5N1)
37	A/duck/Guangxi/xa/2001(H5N1)
38	A/duck/Shanghai/08/2001(H5N1)
39	A/migratory-duck/Jiangxi/1653/2005(H5N1)
40	A/duck/Zhejiang/11/2000(H5N1)

Table 1 continued

41	A/duck/Zhejiang/52/2000(H5N1)
42	A/duck/Shantou/1930/2001(H5N1)
43	A/chicken/Shantou/2535/2001(H5N1)
44	A/goose/Shantou/4325/2001(H5N1)
45	A/duck/Shantou/4407/2001(H5N1)
46	A/goose/Shantou/157/2002(H5N1)
47	A/partridge/Shantou/478/2002(H5N1)
48	A/Quail/Shantou/911/05(H5N1)
49	A/chicken/Shantou/3840/2006(H5N1)
50	A/chicken/Jilin/9/2004(H5N1)
51	A/goose/Jilin/hb/2003(H5N1)
52	A/duck/Yunnan/5236/2005(H5N1)
53	A/goose/Yunnan/5299/2005(H5N1)
54	A/Chicken/Yunnan/447/05(H5N1)
55	A/goose/Yunnan/6384/2006(H5N1)
56	A/chicken/Yunnan/6885/2003(H5N1)
57	A/chicken/Yunnan/1215/2002(H5N1)
58	A/duck/Hunan/1386/2003(H5N1)
59	A/chicken/Hunan/999/2005(H5N1)
60	A/duck/Hunan/5106/2005(H5N1)
61	A/duck/Hunan/988/2006(H5N1)
62	A/domestic green-winged teal/Hunan/79/2005(H5N1)
63	A/Bar-headed Goose/Qinghai/68/05(H5N1)
64	A/Great Black-headed Gull/Qinghai/2/05(H5N1)
65	A/bar-headed goose/Qinghai/F/2006(H5N1)
66	A/black-headed gull/Qinghai/3/2006(H5N1)
67	A/Goose/Hong Kong/ww26/2000(H5N1)
68	A/Goose/Hong Kong/ww28/2000(H5N1)
69	A/Duck/Hong Kong/ww381/2000(H5N1)
70	A/Duck/Hong Kong/ww461/2000(H5N1)
71	A/Chicken/HongKong/FY150/01(H5N1)
72	A/Pheasant/HongKong/FY155/01-MB(H5N1)
73	A/Chicken/Hong Kong/FY150/01 (H5N1)
74	A/Pheasant/HongKong/FY155/01(H5N1)
75	A/Silky Chicken/Hong Kong/SF189/01 (H5N1)
76	A/Quail/Hong Kong/SF203/01 (H5N1)
77	A/Pigeon/Hong Kong/SF215/01 (H5N1)
78	A/Ck/HK/409.1/02 (H5N1)
79	A/Gf/HK/38/2002(H5N1)
80	A/chicken/Hong Kong/915/97(H5N1)
81	A/Duck/Hong Kong/p46/97 (H5N1)
82	A/common magpie/Hong Kong/2125/2006(H5N1)

Table 1 continued

83	A/house crow/Hong Kong/2858/2006(H5N1)
84	A/chicken/Hong Kong/282/2006(H5N1)
85	A/Ck/HK/SSP141/2003(H5N1)
86	A/Ck/Indonesia/PA/2003(H5N1)
87	A/Dk/Indonesia/MS/2004(H5N1)
88	A/Ck/Indonesia/BL/2003(H5N1)
89	A/chicken/Indonesia/CDC25/2005(H5N1)
90	A/chicken/Purworejo/BBVW/2005(H5N1)
91	A/quail/Yogyakarta/BBVet-IX/2004(H5N1)
92	A/chicken/Kupang-2-NTT/BPPV6/2004(H5N1)
93	A/duck/Parepare/BBVM/2005(H5N1)
94	A/chicken/Tebing Tinggi/BPPVI/2005(H5N1)
95	A/quail/Tasikmalaya/BPPV4/2004(H5N1)
96	A/chicken/Korea/IS2/2006(H5N1)
97	A/chicken/Korea/CA7/2006(H5N1)
98	A/duck/Korea/Asan5/2006(H5N1)
99	A/duck/Korea/Asan6/2006(H5N1)
100	A/quail/Korea/KJ4/2006(H5N1)
101	A/open-bill stork/Thailand/VSMU-20-AYA/2004(H5N1)
102	A/pigeon/Thailand/VSMU-11-KRI/2005(H5N1)
103	A/tree sparrow/Thailand/VSMU-14-KRI/2005(H5N1)
104	A/open-bill stork/Thailand/VSMU-15-ATG/2005(H5N1)
105	A/Ck/Thailand/9.1/2004(H5N1)
106	A/Qa/Thailand/57/2004(H5N1)
107	A/duck/Nong-Khai/Thailand/KU-56/2007(H5N1)
108	A/chicken/Thailand/NS-342/2008(H5N1)
109	A/duck/Thailand/CU-329/07(H5N1)
110	A/quail/Thailand/CU-330/06(H5N1)
111	A/chicken/Vietnam/C58/04(H5N1)
112	A/chicken/Viet Nam/VL-008/2004(H5N1)
113	A/quail/Vietnam/36/04(H5N1)
114	A/duck/Viet Nam/1/2005(H5N1)
115	A/chicken/Viet Nam/2/2005(H5N1)
116	A/chicken/Viet Nam/6/2005(H5N1)
117	A/duck/Vietnam/1771/2005(H5N1)
118	A/duck/Vietnam/1/2007(H5N1)
119	A/chicken/Vietnam/29/2007(H5N1)
120	A/Muscovy duck/Vietnam/33/2007(H5N1)
121	A/duck/Vietnam/34/2007(H5N1)
122	A/chicken/Vietnam/20/2003(H5N1)
123	A/mallard/Vietnam/21/2003(H5N1)

Thirdly, in the phylogenetic tree of Fig. 2, (70, 69) (49, 20) (61, 21) (42, 23, 52, 25, 37, 71, 2) (35, 32, 31, 8) (99, 98) (108, 89) (86, 50) (38, 3) (111, 56) (87, 88, 94, 5) (114, 93) (116, 115) formed a clade respectively according to the species virus host. So we can obtain that the phylogenetic relationship also relates to species virus host, various virus strain sharing the same kind of hosts have close phylogenetic relationship though the geographic area and time of the virus strain differ [10, 13].

Meanwhile, in Fig. 2 we find (95, 78, 76, 27) (72, 74, 55) (107, 81, 80, 26, 82, 22) (110, 46) (30, 10) (58, 24) formed a cluster separately, but these virus strains are from different geographic areas, different time and different host, by analyzing the particularity of these virus strains we can infer that these genotypes may be first introduced by migrating birds and then complicated by domestic poultry movements [6, 11, 13]. Although the factors associated with the evolution pathways are not clear, the different environmental factors in different geographic areas and species may be linked to the formation of these emerging influenza genotypes. One common factor may be the differences between bird species range, population size, farming system and the structure of the poultry industry.

Finally, in Fig. 2 we can obtain that (19, 12, 60, 36, 34, 33, 4) (49, 20, 17) (35, 32, 31, 8) (37, 71, 2) (87, 88, 94, 5) (103, 116, 115, 102, 101) formed a small clade respectively, and the phylogenetic analyses revealed that the genotype isolated from the duck was closely clustered with the chicken.

The surface protein HA and NA, most importantly HA, were considered as the major antigens of influenza viruses. HA may influence the pathogenicity and lethality of the virus and the trait of the host. In Fig. 1 we give the phylogenetic tree of HA, NA. Here we found some rather unexpected sequences, they are 112(A/chicken/VietNam/VL-008/2004(H5N1)), 61(A/duck/Hunan/988/2006(H5N1)), 21(A/duck/Fujian/671/2006(H5N1)) in the phylogenetic tree of HA, 16(A/black bulbul/Fujian/439/04(H5N1)), 15(A/golden mountain thrush/Fujian/376/04(H5N1)), 14(A/babbler/Fujian/320/04(H5N1)) in the phylogenetic tree of NA, and 77(A/Pigeon/Hong Kong/SF215/01(H5N1)) in the phylogenetic tree of NP. They are different from other sequences, which have great distance between them and other sequences. We infer that (112, 61, 21) may have a relationship with the complexity of the HA mutation, and (16, 15, 14, 77) are due to the special species virus host. Overall, from the phylogenetic tree of HA, we obtain that this phylogenetic relationship mainly relates to species virus host and geographical distribution, while the phylogenetic relationship of NA mainly relates to time and geographical distributions which we can see from the phylogenetic tree of NA. So (112, 61, 21) form a small clad in the phylogenetic tree of HA, while (16, 15, 14) form a small clad in the phylogenetic tree of NA. The result is identified with the previous study [5].

We analyze the phylogenetic tree of HA, NA from every small clade, and we can see that the similarity to HA, NA is identified with phylogenetic tree (Fig. 2) basically. In addition, in Fig. 1 we also give the phylogenetic tree of NP, and we can find *it's* phylogenetic relationship mainly relates to species virus host and geographical distributions obviously, so we can also affirm HA, NA and NP have performed significant functions in genetic reassortment, recombination and mutation, the similar results can be found in references [10].

3 Conclusion

This novel method based on a mapping which can be recaptured mathematically without loss of textual information and gave a new method to compute the distance between strains i and j which is very simple for calculation.

The examination of phylogenetic tree belongs to 123 H5N1 avian influenza virus protein sequences illustrated the utility of our approach. We obtained that phylogenetic relationship among H5N1 avian influenza virus closely relates to distributions of time, geography and species virus host, and we inferred that the genotype isolated from the duck was closely clustered with the chicken, moreover, phylogenetic analyses showed that the segment HA, NA and NP have performed significant functions in the evolution of H5N1 avian influenza virus.

Acknowledgments This work was supported in part by the Shandong Natural Science Foundation (Y2006A14).

References

1. J.A. Bergera, S.K. Mitra, M. Carli, A. Neri, Visualization and analysis of DNA sequences using DNA walks. *J. Franklin Inst.* **341**, 37–53 (2004)
2. R.A. Fouchier et al., Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J. Virol.* **79**, 2814–2822 (2005)
3. A.L. Ghindilis et al., CombiMatrix oligonucleotide arrays: genotyping and gene expression assays employing electrochemical detection. *Biosens. Bioelectron* **22**, 1853–1860 (2007)
4. Y. Guan et al., H5N1 influenza: a protean pandemic threat. *Proc. Natl. Acad. Sci. USA* **101**, 8156–8161 (2004)
5. Y. Hiromoto, Y. Yamazaki, T. Fukushima et al., Evolutionary characterization of the six internal genes of H5N1 human influenza A virus. *J. Gen. Virol.* **81**, 1293–1303 (2000)
6. E.D. Kilbourne, Perspectives on pandemics: a research agenda. *J. Infect. Dis.* **176**(Suppl. 1), S29–S31 (2007)
7. Y. Li et al., Detection of Hong Kong 97-like H5N1 influenza viruses from eggs of Vietnamese waterfowl. *Arch. Virol.* **151**, 1615–1624 (2006)
8. B. Liao, J. Luo, R. Li, W. Zhu, Novel method for analyzing proteome. *Int. J. Quantum Chem.* **107**, 1295–1300 (2006)
9. M. Randic, J. Zupan, A.T. Balaban, Unique graphical representation of protein sequences based on nucleotide triplet codons. *Chem. Phys. Lett.* **397**, 247–252 (2004)
10. C. Rohm et al., Characterization of a novel influenza hemagglutinin, H15: criteria for determination of influenza A subtypes. *Virology* **217**, 508–516 (1996)
11. G.J. Smith et al., Evolution and adaptation of H5N1 influenza virus in avian and human hosts in Indonesia and Vietnam. *Virology* **350**, 258–268 (2006)
12. The World Health Organization Global Influenza Program Surveillance Network, Evolution of H5N1 avian influenza viruses in Asia. *Emerging infectious diseases.* **11**, 1515–1521 (2005)
13. X. Wan, G. Chen, F. Luo, M. Emch, R. Donis, A quantitative genotype algorithm reflecting H5N1 avian influenza niches. *Bioinformatics* **18**, 2368–2375 (2007)
14. R.J. Webby, R.G. Webster, Are we ready for pandemic influenza?. *Science* **302**, 1519–1522 (2003)
15. Y. Weng, Y. Yan, Y. Zhang, S. Yang, X. Shen, Genome sequencing and analysis of a high pathogenic avian influenza virus strain A/Fujian/1/2007(H5N1) isolated from a patient in Fujian, China. *Chin. J. Zoonoses* **7**, 635–638 (2007)
16. Y. Zhang, B. Liao, K. Ding, On 2D graphical representation of DNA sequences of nondegeneracy. *Chem. Phys. Lett.* **411**, 28–32 (2005)
17. Y. Zhang, B. Liao, On 3DD-curves of DNA sequences. *Mol. Simul.* **32**, 29–34 (2006)

18. Y. Zhang, On 3D graphical representation of RNA secondary structure. *MATCH Commun. Math. Comput. Chem.* **57**, 157–168 (2007)
19. Y. Zhang, W. Chen, A new approach to molecular phylogeny of Primate mitochondrial DNA. *MATCH Commun. Math. Comput. Chem.* **59**, 625–634 (2008)