The panorama of the diversity of H5 subtype influenza viruses

**Hong-Chao Ma · Ji-Ming Chen · Ji-Wang Chen ·
Ying-Xue Sun · Jin-Ming Li · Zhi-Liang Wang**

Abstract To elucidate the global diversity of H5 influenza viruses from a dynamic view, haemagglutinin (HA) sequences of 170 isolates were selected and analyzed in this study. Our results showed that H5 influenza isolates could be divided into two distinct lineages that circulated in the Eastern Hemisphere and the Western Hemisphere, respectively. This may be due to the separate migration routes and habitats of birds in the two hemispheres. The two distinct lineages, having existed at least for decades, possibly began divergence in 1850s. Each of the two distinct HA lineages could be further divided into some sublineages, but there was little correlation between the minor lineages and their isolation places, isolation time, neuraminidase subtypes, host species or virulence. The panorama of the diversity of H5 influenza viruses presented here integrated all known H5 epidemics including the current severe H5N1 avian epidemics in the Eastern Hemisphere and suggested that H5 virulent viruses could originate from multiple sublineages and associate with multiple NA subtypes. Our study provided a framework for the studies on the evolution and epidemiology of H5 influenza viruses.

Keywords Haemagglutinin · Lineage · H5 · Influenza virus · Evolution · Diversity

H.-C. Ma · J.-M. Chen (✉) · Y.-X. Sun · J.-M. Li ·
Z.-L. Wang
Chinese Center for Animal Health and Epidemiology,
Nanjing Road 369#, Qingdao 266032, China
e-mail: chenjiming2004@yahoo.com.cn

J.-W. Chen
Department of Medicine, Northwestern University,
Chicago, IL 60611, USA

Currently, H5 subtype high pathogenic avian influenza (HPAI) is posing heavy burdens to the global poultry and severe threats to the world public health. Although the arduous and expensive culling with or without immunization strategy has been implemented in most involved countries to control the HPAI epidemics since the year 2003, the epidemics continue expanding from several spots in Asia in 2003 to most regions in Eurasia in 2005 and even to Africa in 2006. Meanwhile, the toll of highly fatal human infections with the HPAI virus from more and more countries also continues rising [1, 2] (See http://www.who.int and http://www.oie.int).

The diversity of H5 subtype influenza viruses has been explored from regional views with limited sequences [3–17], but not from the global and dynamic view. In this study, we try to present such a panorama and explore its significance through analyses of haemagglutinin (HA) sequences of H5 influenza viruses available online.

Until April 1, 2006, approximately 716 H5 influenza isolates whose complete sequences of HA1 domain in HA gene were reported to the influenza sequence database (http://www.flu.lanl.gov). These 716 isolates covered 7 neuraminidase (NA) subtypes, namely, N1, N2, N3, N6, N7, N8 and N9, respectively, with the number of 538, 123, 25, 3, 3, 3 and 12, and the NA subtype of the rest 9 isolates were unknown. More H5N1 and H5N2 sequences other than the NA subtypes' were reported, which might be because only H5N1 and H5N2 subtypes have widely circulated in the domestic fowls in the past decades.

Through sequence analyses with the online alignment and BLAST software in the influenza sequence database (http://www.flu.lanl.gov), 533 H5N1 strains

isolated after 1996 were found to be more similar to A/Goose/Guangdong/1/1996 than any other strains isolated before 1997 in HA1 sequences, and therefore only 19 (A/Chicken/Hong Kong/1203/97, A/Hong Kong/156/97, A/duck/Zhejiang/15/2000, A/duck/Fujian/17/2001, A/duck/Shanghai/37/2002, A/chicken/Korea/ES/03, A/wild duck/Guangdong/314/2004, A/chicken/Jilin/9/2004, A/swan/Guangxi/307/2004, A/cat/Thailand/KU-02/04, A/whooper swan/Mongolia/4/05, A/grebe/Novosibirsk/29/2005, A/quail/Guangxi/575/2005, A/turkey/Turkey/1/2005, A/Chicken/Vietnam/NCVD12/2005, A/quail/Thailand/Nakhon Pathom/QA-161/2005, A/chicken/Nigeria/641/2006, A/swan/Iran/754/2006, A/chicken/Egypt/960N3-004/2006) among these 534 H5N1 isolates were randomly selected. Similarly, some H5N2 strains isolated in the same place and in the same year and with HA1 sequence similar to that of some selected strains were also excluded for further analyses. After the preliminary selection, HA1 sequences of 170 isolates[1] including 24 (H5N1), 97 (H5N2), 25 (H5N3), 3 (H5N6), 3 (H5N7), 3 (H5N8), 12 (H5N9), and 3 NA-unknown strains were selected in this study. Among these 170 isolates, 62 were from chickens, 36 from domestic ducks, 11 from turkeys, 2 from geese, 2 from mammals (A/HongKong/156/97 and A/cat/Thailand/KU-02/04) and the rest from wild birds. All H5 influenza viruses previously isolated from mammals were assumed to be transmitted from infected fowls and could not be transmitted among mammals [1]. In addition, 53 among these 170 isolates were isolated in

1990s, and other 76 were isolated in 2000s, and the rest 41 were isolated in 1959–1989.

Sequence alignment, which was done with the online software in the influenza sequence database (http://www.flu.lanl.gov), demonstrated that there were no insertions or deletions among the HA1 sequences except the HA0 cleavage site that is located at the end of HA1 sequence. The phylogenetic tree (Fig. 1) of the sequences was constructed using software PHYLIP 3.62 (Fig. 1) with maximum likelihood method [18, 19]. Although many H5N1 sequences similar to that of A/Goose/Guangdong/1/1996 and some H5N2 strains were excluded from the tree as stated above, the main topology of the phylogenetic tree (Fig. 1) changed little when these sequences were included through tests (data not shown). Therefore, the phylogenetic tree (Fig. 1) was assumed to be the panorama of the diversity of H5 influenza viruses.

The phylogenetic tree (Fig. 1) suggested that all the H5 strains isolated after 1961 could be aligned into one of the two lineages except A/Turkey/Ramon/73. The two lineages were designated as the Eastern lineage and the Western lineage, respectively, because strains in the Eastern lineage were all isolated from the Eastern Hemisphere (Asia, Europe or Africa) and strains in the Western lineage were all isolated from the Western Hemisphere (America) except A/chicken/Taiwan/1209/03 and A/duck/Hokkaido/84/02, which were also assumed to be oddly introduced from the Western Hemisphere (See: http://www.recombinomics.com/News/09040501/H5N2_WBF_Japan.html). This result was consistent with some previous regional studies, which suggested that avian influenza viruses of some HA subtypes could be assorted into Eurasian and North American lineages [2, 5, 10]. The geographical distribution of the two lineages was possibly because of the separate migration routes and habitats of birds in the two hemispheres [20]. Migration birds, which may carry the viruses to a faraway place, were considered as the main reservoir of the influenza viruses [3]. According to this hypothesis, H5 influenza viruses in South America isolated there in the future will probably belong to the Western lineage because many birds migrate between the North and South America [20].

Each of the two distinct lineages harbored several sublineages (Fig. 1). For example, the isolates causing HAPI in USA in 1980s and those causing HAPI in Mexico in 1990s were located in different sublineages. Figure 1 suggested that isolates of the same NA subtype, or of similar virulence, which was deduced from the amino acid residues at the HA0 cleavage site, or isolated from same region, or isolated from the same host species, or isolated in the

---

[1] The accession numbers of the 170 sequences in the influenza sequence database were (1) H5N1: AB233321, AF036356, AF082040, AF082043, AF098539, AF144305, AY585369, AY585371, AY585372, AY653200, AY676035, DQ230521, DQ236077, DQ320897, DQ334768, DQ406728, DQ407519, DQ440535, DQ447199, ISDN124036, S68489, U79453, X07869; (2) H5N2: AB241614, AF082042, AF098537, AF100180, AF164655~AF164665, AF194169, AF194990, AF194991, AJ305306, AJ621811, AJ632269, AY296064~AY296083, AY497065~AY497088, AY497093~AY497095, AY573917, AY684894, AY849793, AY995883~AY995885, AY995889, AY995896, CY005575, CY005918, DQ092869, DQ104701, DQ122147, DQ256383, DQ387854, J04325, L46587, M10243, U05330~U05332, U28920, U37172, U37174, U37179, U37181, U37182, U79449, U79450, U79454; (3) H5N3: AB241615, AB241616, AB241617, AB241619, AB241621~AB241626, AF082038, AF082039, AF290443, AF303057, AJ621807, AJ632268, AY296084, AY296085, AY995891, AY995895, CY005926, CY006028, ISDN49024, U20460, U79452; (4) H5N6: AF082041, AY995892, U79451; (5) H5N7: AY531029, CY005927, CY005969; (6) H5N8: AY296086, M18450, M18451; (7) H5N9: AF194992, AY995886~AY995888, AY995890, AY995893, AY995894, AY995897, AY995898, M30122, U67783, U79456; (8) NA unknown: AY950235, AY950236, U69277.

Fig. 1 The phylogenetic tree of 170 H5 influenza isolates based on their HA1 nucleic acid sequences estimated with software PHYLIP 3.62 using maximum likelihood method [18]. The transitions/transversions ratio was set as 8.0, and the substitution rates were set in Gamma distribution. The earliest isolate, A/chicken/Scotland/59, was selected as the outgroup. Some bootstrap values out of 100 replicates were given near to the corresponding nodes, and the isolation time was given at the right side. Most isolate designations were omitted due to space limitation. "*" indicated that the isolates were highly pathogenic with >3 basic amino acid residues at the HA0 cleavage site, and "Δ" indicated that there were 2 or 3 basic amino acid residues at the HA0 cleavage site and the isolates remained low pathogenic. Words in rectangles expressed the place, serotype and outbreak time of related severe avian influenza epidemics. The vertical lines were just for spacing while the transversal lines represented genetic distances whose scale was given at the right below corner
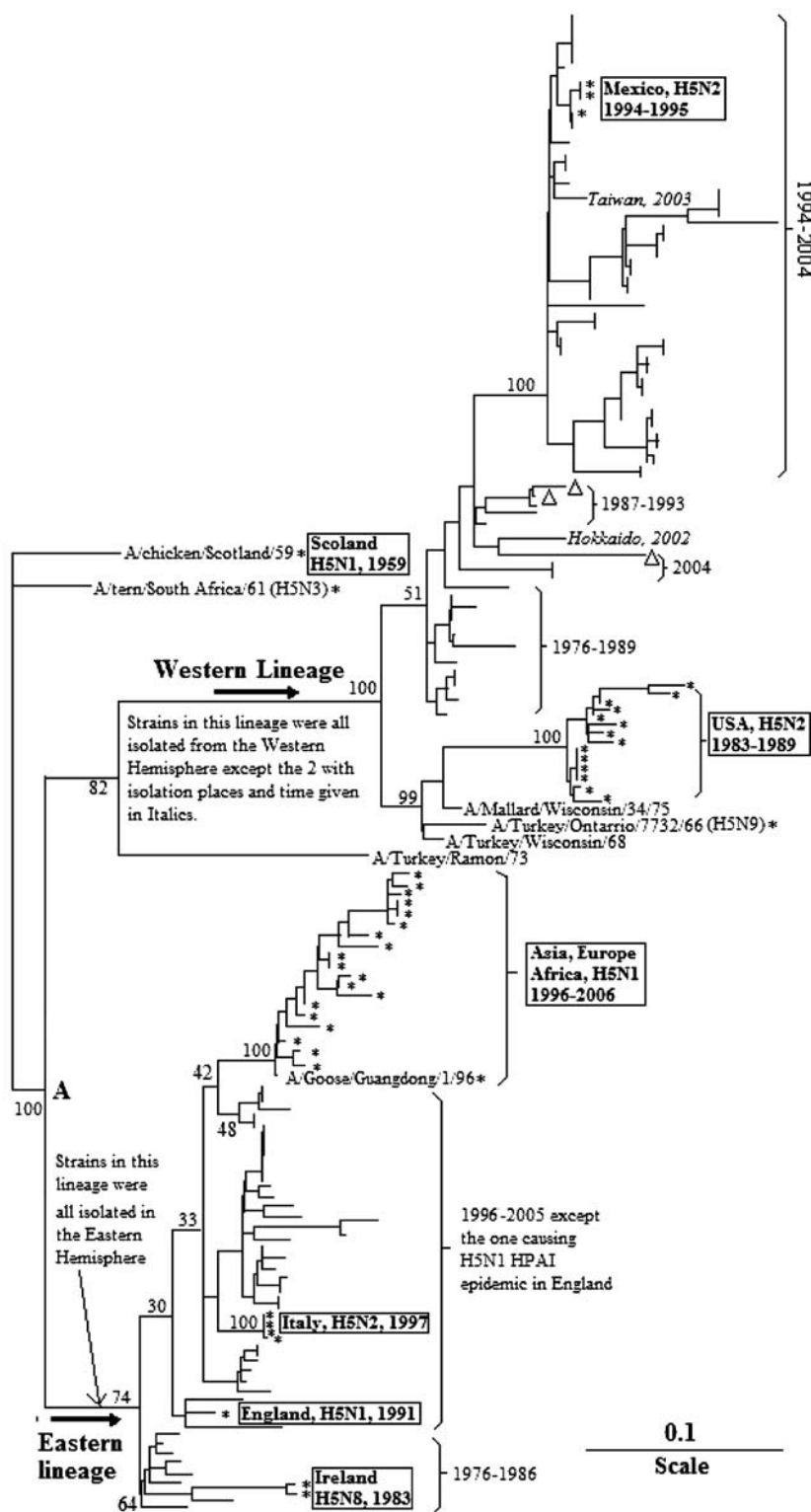
same year were possibly situated in different HA minor lineages, and a HA minor lineage possibly contained isolates of different NA subtypes, or of different virulence, or isolated from different regions or different host species, or isolated in different years. In addition, both of the two distinct lineages covered all the aforementioned 7 NA subtypes (Table 1).

Figure 1 demonstrated that the two distinct lineages have existed at least for decades. The high bootstrap

Table 1 The neuraminidase subtype distributions of the two lineages in Fig. 1

	N1	N2	N3	N6	N7	N8	N9
Eastern lineage	21	19	19	2	1	2	9
Western lineage	3	77	5	1	2	1	3

Table 2 The amino acid residues different between the two lineages in 2000s

Positions	30	51	69	120	172	179	185	190	210	211	226	250	256	260	291	298	310	326
Eastern lineage	E	K	R	L	A	S	Q	V	P	T	V	K	N	N	N	M	I	R
Western lineage	K	E	K	M	V	T	V/I	I	S/L	N	I	R	S	T	D	V	V	K

The positions were numbered according to the HA0 sequences of A/swan/Iran/754/2006, and the amino acid residues given in this table were applicable to most rather than all isolates in the either lineage in 2000s

values at nodes A, B, C and D in Fig. 1 suggested that A/chicken/Scotland/59, A/tern/SouthAfrica/61 and A/Turkey/Ramon/73 could be considered as the intermediate strains of the two lineages with confidence [19, 21]. However, because inadequate sequences of strains isolated before 1970, the exact divergence time remains unclear. Though there was no obvious correlation between minor HA lineages and their isolation time, which is different from human influenza virus [22], the distance between the two distinct lineages has increased in the past decades.

To estimate when the two distinct lineages began divergence, the divergence rate of the two lineages was calculated. Firstly, the distances between the two lineages were estimated using software PHYLIP 3.62 according to Kimura-2-Parameter model with substitution rates differing in Gamma distribution and transistions/transversions ratio was set as 8.0 [19]. Because of inadequate sequences before 1990, only the means of the distances between the two lineages in 1990s (= 0.2996 substitution/site) and in 2000s (= 0.3129 substitution/site) were calculated with the nucleotide sequences of the aforementioned 53 selected isolates isolated in 1990s (Mean of the isolation time = 1996.09) and 76 selected isolates isolated in 2000s (Mean of the isolation time = 2002.45). Therefore, the rate of the divergence was approximately 0.00209 substitution/site/year). According to this rate, the two lineages possibly began divergence in 1850s. However, this was just a preliminary estimate based on the assumption that the two lineages diverged at a constant rate under the Kimura-2-Parameter model.

In 2000s, there have been 17 amino acid residues that were found to be different between the two lineages (Table 2), but the significance of these residues need to be investigated experimentally. Despite these differences in amino acid sequences between the two lineages, inactivated vaccine viruses from the Western lineage could provide complete protection from clinical signs and deaths caused by the current H5N1 HPAI viruses from the Eastern lineage and reduce the number of chickens infected and shedding virus [2]. This was possible because these different residues were not situated on the HA antigenic epitopes [23].

Figure 1 integrated all important H5 epidemics in the past decades. For example, it indicated that the current rampant H5N1 HAPI viruses in the Eastern Hemisphere were phylogenetically closer to the H5N1 HPAI viruses circulating in England in 1959 than to the H5N2 HPAI viruses circulating in Mexico in 1990s. Figure 1 also provided a framework for the studies on the evolution and epidemiology of H5 influenza viruses. For example, it indicated that H5 HPAI viruses possibly originated from multiple minor lineages in each of the distinct lineages and their NA subtypes could be N1, N2, N3, N8 or N9.

In conclusion, this study described a whole map of the diversity of H5 influenza viruses, which provided us some information about the distributions of NA subtypes, isolation places, isolation years, host species and virulence of the viruses. It also provided a framework for the studies on the evolution and epidemiology of H5 influenza viruses including the investigations on the further development of the two distinct lineages. For example, whether will one disappear? whether will they circulate in overlapping areas and further evolve into two separate subtypes?

References

1. D. Kobasa, Y. Kawaoka, Curr. Mol. Med. **5**, 791–803 (2005)
2. D.E. Swayne, C.V. Lee, E. Spackman, Avian Pathol. **35**, 141–146 (2006)
3. C. Abolnik, E. Cornelius, S.P. Bisschop, M. Romito, D. Verwoerd, Dev. Biol. (Basel) **124**, 189–199 (2006)
4. C. Bender, H. Hall, J. Huang, A. Klimov, N. Cox, A. Hay, V. Gregory, K. Cameron, W. Lim, K. Subbarao, Virology **254**, 115–123 (1999)
5. I. Donatelli, L. Campitelli, L. Di Trani, S. Puzelli, L. Selli, A. Fioretti, D.J. Alexander, M. Tollis, S. Krauss, R.G. Webster, J. Gen. Virol. **82**, 623–630 (2001)
6. M. Garcia, J.M. Crawford, J.W. Latimer, E. Rivera-Cruz, J. Gen. Virol. **77**, 1493–1504 (1996)
7. M. Garcia, D.L. Suarez, J.M. Crawford, J.W. Latimer, R.D. Slemons, Virus Res. **51**, 115–124 (1997)
8. Y. Guan, M. Peiris, K.F. Kong, K.C. Dyrting, T.M. Ellis, T. Sit, L.J. Zhang, K.F. Shortridge, Virology **292**, 16–23 (2002)
9. T. Horimoto, E. Rivera, J. Pearson, D. Senne, S. Krauss, Y. Kawaoka, R.G. Webster, Virology **213**, 223–230 (1995)
10. Y. Kawaoka, A. Nestorowicz, D.J. Alexander, R.G. Webster, Virology **158**, 218–227 (1987)
11. C.W. Lee, D.A. Senne, J.A. Linares, P.R. Woolcock, D.E. Stallknecht, E. Spackman, D.E. Swayne, D.L. Suarez, Avian Pathol. **33**, 288–297 (2004)
12. C.W. Lee, D.A. Senne, D.L. Suarez, J. Virol. **78**, 8372–8381 (2004)
13. M. Matrosovich, N. Zhou, Y. Kawaoka, R. Webster, J. Virol. **73**, 1146–1155 (1999)
14. V.J. Munster, A. Wallensten, C. Baas, G.F. Rimmelzwaan, M. Schutten, B. Olsen, A.D. Osterhaus, R.A. Fouchier, Emerg Infect Dis. **11**, 1545–1551 (2005)
15. C. Rohm, T. Horimoto, Y. Kawaoka, J. Suss, R.G. Webster, Virology **209**, 664–670 (1995)
16. D.L. Suarez, D.A. Senne, Avian Dis. **44**, 356–364 (2000)
17. T. Saito, T. Horimoto, Y. Kawaoka, D.A. Senne, R.G. Webster, Virology **201**, 277–284 (1994)
18. S. Ota, W.H. Li, Mol. Biol. Evol. **17**, 1401–1408 (2000)
19. W.H. Li, *Molecular Evolution*. (Sinauer Associates, Sunderland Associates, Massachusetts, 1997), p. 75
20. P. Berthold, *Bird Migration: A General Survey*. (Oxford University Press, New York, 2001), p. 56
21. B. Efron, E. Halloran, S. Holmes, Proc. Natl. Acad. Sci. USA **93**, 7085–7090 (1996)
22. R.M. Bush, C.A. Bender, K. Subbarao, N.J. Cox, W.M. Fitch, Science **286**, 1921–1925 (1999)
23. Y. Ha, D.J. Stevens, J.J. Skehel, D.C. Wiley, EMBO J. **21**, 865–875 (2002)