2010

New Evidence Suggests Southern China as a Common Source of Multiple Clusters of Highly Pathogenic H5N1 Avian Influenza Virus

Bin Wu
Institute of Zoology

Chengmin Wang
Institute of Zoology

Guoying Dong
Institute of Zoology

Yunhai Guo
Institute of Zoology

Dale L. Nolte
USDA-APHIS-Wildlife Services, Dale.L.Nolte@aphis.usda.gov

See next page for additional authors

MAJOR ARTICLE

New Evidence Suggests Southern China as a Common Source of Multiple Clusters of Highly Pathogenic H5N1 Avian Influenza Virus

Bin Wu,[1,a] Chengmin Wang,[1,a] Guoying Dong,[1] Yunhai Guo,[1] Dale Louis Nolte,[4] Thomas Jude Deliberto,[4] Jianguo Xu,[3] Mingxing Duan,[2] and Hongxuan He[1]

[1]National Research Center for Wildlife Borne Diseases, Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, and [2]State Key Laboratory of Biomembrane and Membrane Biotechnology, School of Life Sciences, Tsinghua University, Beijing, and [3]Department of Respiratory, First Hospital, Jilin University, Changchun, China; [4]National Wildlife Research Center, Wildlife Services, Animal and Plant Health Inspection Service, United States Department of Agriculture, Fort Collins, Colorado

Highly pathogenic H5N1 avian influenza is considered an avian disease, although there is some evidence of limited human-to-human transmission of the virus. A global effort is underway to control or eradicate the highly pathogenic H5N1 avian influenza virus in poultry and prevent human exposure, both of which may also reduce the risk of pandemic emergence. Hemagglutinin gene sequences from 215 human H5N1 influenza viruses were used to trace the source and dispersal pattern of human H5N1 influenza viruses on a global scale. A mutation network and phylogenetic analyses of the hemagglutinin gene show that human H5N1 influenza viruses can be clearly divided among 4 clusters across geographic space. On the basis of analysis of the N-glycosylation sites at positions 100 and 170 in the hemagglutinin protein, human H5N1 influenza viruses were also divided into 3 types. When we combined these analyses with geographic information system data analyses, we found that Southern China is often a common source of multiple clusters of H5N1 influenza viruses and that each cluster has different dispersal patterns and individual evolutionary features. In summary, the genetic evidence presented here provides clear evidence for multiple clusters of human H5N1 influenza viruses that initially originated in Southern China.

Highly pathogenic avian influenza (HPAI) of the H5N1 subtype is an epidemic disease and has produced infections that have resulted in grand-scale economic losses [1]. The HPAI H5N1 virus emerged as a human pathogen in 1997 within the Hong Kong Special Administrative Region [2]. Since that outbreak, H5N1 influenza has spread to wild birds across Eurasia and as far west as England and Africa, thereby threatening to spread into the American and Australian continents. H5N1 influenza virus is primarily a pathogen of poultry, as more than 200 million poultry have died or been culled because of this virus [3]. The virus is also of great concern for public health because it has caused several hundred infections in humans. Although the predominant pathway of infection in humans is still avian-to-human transmission [4], the virus is always changing and mutations that make it more compatible with human transmission may occur at any time [5]. Approximately 60% of the 478 reported infections in humans (as of 17 February 2010) have been fatal [6], but very few have been transmitted from person to person [7]. Infections are usually acquired by contact with poultry or poultry products infected with H5N1

Received 28 January 2010; accepted 1 March 2010; electronically published 24 June 2010.

Potential conflicts of interest: none reported.

Financial support: National Key Basic Research and Development Program of China (grant 9732007BC109103); Chinese Academy of Sciences Innovation Program (grant KSCX2-YW-N-063); United States Department of Agriculture Animal and Plant Health Inspection Service Wildlife Services and Chinese Academy of Sciences Institute of Zoology (joint project 0760621234); National Science and Technology Ministry (ID 2009BAI83B01).

[a] B.W. and C.W. contributed equally to this work.

Reprints or correspondence: Prof Hongxuan He, National Research Center for Wildlife Borne Diseases, Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, No. 1-5, Beichen West Rd, Chaoyang District, Beijing 100101, China (hehx@ioz.ac.cn).

The Journal of Infectious Diseases 2010; 202(3):452–458

0022-1899/2010/20203-0016$15.00

DOI: 10.1086/653709

This figure is available in its entirety in the online version of the *Journal of Infectious Diseases.*

Figure 1. Global geographical distribution of highly pathogenic H5N1 avian influenza.

influenza virus, and the H5N1 influenza viruses isolated from cases of infection in humans are often virtually identical to isolates from poultry [8]. A global effort is underway to control or eradicate H5N1 influenza virus in poultry and prevent exposure to humans, both of which may also reduce the risk of pandemic emergence. The scientific support for these programs is provided by ecological, virological, epidemiological, and immunological studies. In particular, molecular and functional characterization of H5N1 influenza viruses from poultry will assist in the development and implementation of public health control measures involving diagnosis, immunization, and antiviral drug therapy.

In this study, we used hemagglutinin gene sequences of viruses isolated from 215 cases of H5N1 influenza in humans to trace the dispersal pattern of human H5N1 influenza viruses on a global scale. The results of these analyses were combined with geographic information system data to determine the geographic origin of the human H5N1 influenza viruses that are currently in circulation.

MATERIALS AND METHODS

HPAI H5N1 virus infection data. Information on avian influenza is often rare and incomplete. The most reliable information is from the World Organization for Animal Health (OIE). We obtained the data on avian influenza from the OIE report on disease (http://www.oie.int/eng/info/hebdo/A_INFO.HTM), which is the official report by the member countries of the OIE. The information in the report includes the date of occurrence, the site of occurrence, the type of avian influenza virus, and the kind and number of dead animals (domestic or wild). Information about sites and virus type is often incomplete. However, in comparison with the information on cases in animals, the information on cases in humans from the World Health Organization (http://www.who.int/csr/disease/avian_influenza/updates/en/index.html) is relatively complete. The information includes date, outbreak situation, sex, age, and location (county, province, and town). The information on the latitude and longitude of each site was derived using relevant maps or mapping software such as Google Earth, 3D World Map, and MapQuest. Our database has 5885 records of avian H5N1 virus infection during the period from December 2003 through December 2009 and 215 records of human H5N1 virus infection during the period from 2004 through 2009. The database is managed with the use of ArcGIS geographic information system software (version 9.2; ESRI).

Phylogenetic analyses. We extracted hemagglutinin (HA) gene sequences of 215 human H5N1 influenza virus strains that were isolated from 2004 through June 2009 and some avian H5N1 influenza virus reference strains from the National Center for Biotechnology Information (NCBI) Web site (http://www.ncbi.nlm.nih.gov). Phylogenetic analyses were performed by use of the neighbor-joining method with Mega software (version 4.0) [9] and median-joining networks [10]. The locations of the N-glycosylation sites in the HA protein were initially predicted with data from the NetNGlyc server (http://www.cbs.dtu.dk/services/NetNGlyc/) by use of artificial neural networks that examine the context of Asn-Xaa-Ser/Thr sequences [11]. To further identify the evolutionary features between the 4 clusters, the 3-dimensional structure of the HA protein was constructed using Swiss-Model software [12] (available through a Web interface at http://swissmodel.expasy.org/SWISS-MODEL.html) and visualized using Visual Molecular Dynamics software (version 1.8.6; available at http://www.ks.uiuc.edu/Research/vmd/) [13].

RESULTS AND DISCUSSION

The putative ancestor of the currently circulating H5N1 influenza virus is A/goose/Guangdong/1/96 (Gs/GD/1/96), named after the province in Southern China that is also thought to be the epicenter of H5N1 influenza [14, 15]. However, contemporary H5N1 influenza viruses carry only the HA gene derived from the Gs/GD/1/96 lineage [16]. The other 7 genes of contemporary H5N1 influenza viruses were acquired from other avian influenza viruses by means of genetic reassortment [17]. The HA gene from Gs/GD/1/96 has evolved extensively in the past decade, diverging into multiple clades [18]. In 2005, unprecedented outbreaks of H5N1 influenza in wild migratory birds at Qinghai Lake in Western China apparently facilitated its spread to >30 countries in Europe, the Middle East, and North Africa [19, 20]. The body of evidence is growing that some migratory birds have the capacity to spread the highly pathogenic H5N1 influenza virus over extended distances [21].

The evolutionary linkages between the strains of H5N1 influenza virus that emerged in East Asia in the 1990s and those that were formerly isolated in Europe have yet to be elucidated, although clear differences in the pathogenicity of the Asian and European viruses have been documented [22]. In summarizing

Table 1. Hemagglutinin Gene Sequences of the Human H5N1 Influenza Viruses and Reference Strains Studied

This table is available in its entirety in the online version of the *Journal of Infectious Diseases*

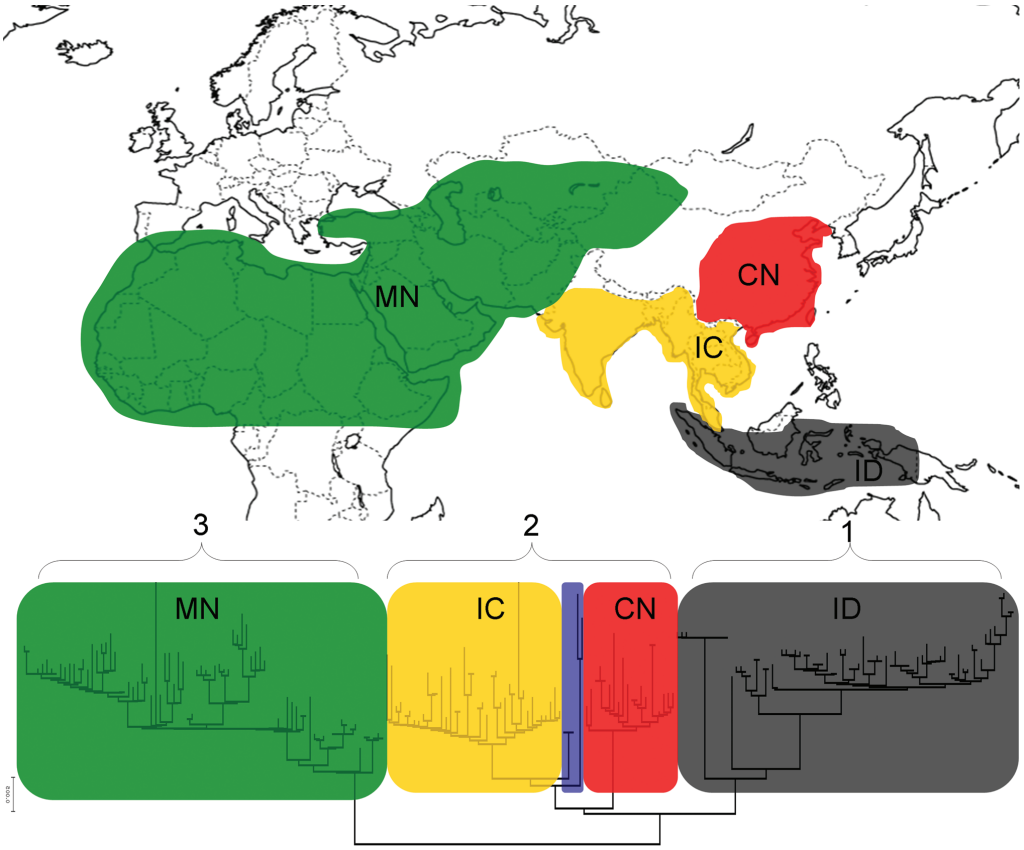

Figure 2. Consensus tree (determined with the neighbor-joining method) of the hemagglutinin genes of 215 human H5N1 avian influenza viruses rooted in a common source (Mv77/avian/China/2001 or Mv77/avian/China/2002). A total of 4 clusters (each represented by a specific color and a corresponding region on the global map) are contained within 3 major clades on the tree (1, 2, and 3). ID, Indonesia cluster; CN, Southern China cluster; IC, Indo-China cluster; MN, Middle Asia–North Africa cluster.

the evidence for the geographical origins of the current HPAI H5N1 virus panzootic, Chen et al [23] suggested Southern China as the most likely (immediate) source on account of (1) the original detection of the precursor HPAI H5N1 virus in 1996 and (2) the high degree of genetic diversity of HPAI H5N1 viruses in the region. Their findings are consistent with the influenza epicenter hypothesis—a hypothesis that identifies Southern China as the common source of avian and human strains of influenza virus [5]. In addition, our present study presents new data that show global human HPAI H5N1 virus strains are divided into 4 clusters contained within 3 major clades.

To investigate the relationship between human HPAI H5N1 viruses and avian HPAI H5N1 viruses on a global scale, we obtained data on avian influenza from the OIE report on disease and information on cases in humans from the World health Organization and managed our database with ArcGIS software (see Materials and Methods). The results indicate that cases in humans are mainly distributed in China, Azerbaijan, Bangladesh, Cambodia, Djibouti, Egypt, Indonesia, Iraq, Laos, Myanmar, Nigeria, Pakistan, Thailand, Turkey, and Vietnam, but avian cases are more widely distributed in Asia, Europe, and Africa (Figure 1). Interestingly, the geographic pattern of cases of HPAI H5N1 virus infection in humans follows that of the avian cases.

We also extracted HA gene sequences (Table 1) from 215 human HPAI H5N1 virus strains isolated from 2004 through June 2009 and some avian HPAI H5N1 virus reference strains from the NCBI Web site (see Materials and Methods). Our phylogenetic analyses shows that the basal lineages of human HPAI H5N1 viruses show some global structuring. Globally, human HPAI H5N1 virus strains are clearly divided into 4 clusters across 3 clades: the Southern China cluster, the Indonesia cluster, the Indo-China cluster, the and Middle Asia–North Africa cluster, contained within clades 1, 2, and 3 (Figure 2). Among these clusters, viruses in Japan, Korea, and European countries are not included because no cases were detected in humans despite the occurrence of outbreaks of highly pathogenic H5N1 avian influenza (Figure 2). In Southeast Asia, which includes Southern China, Vietnam, Myanmar, Laos, Cambodia,

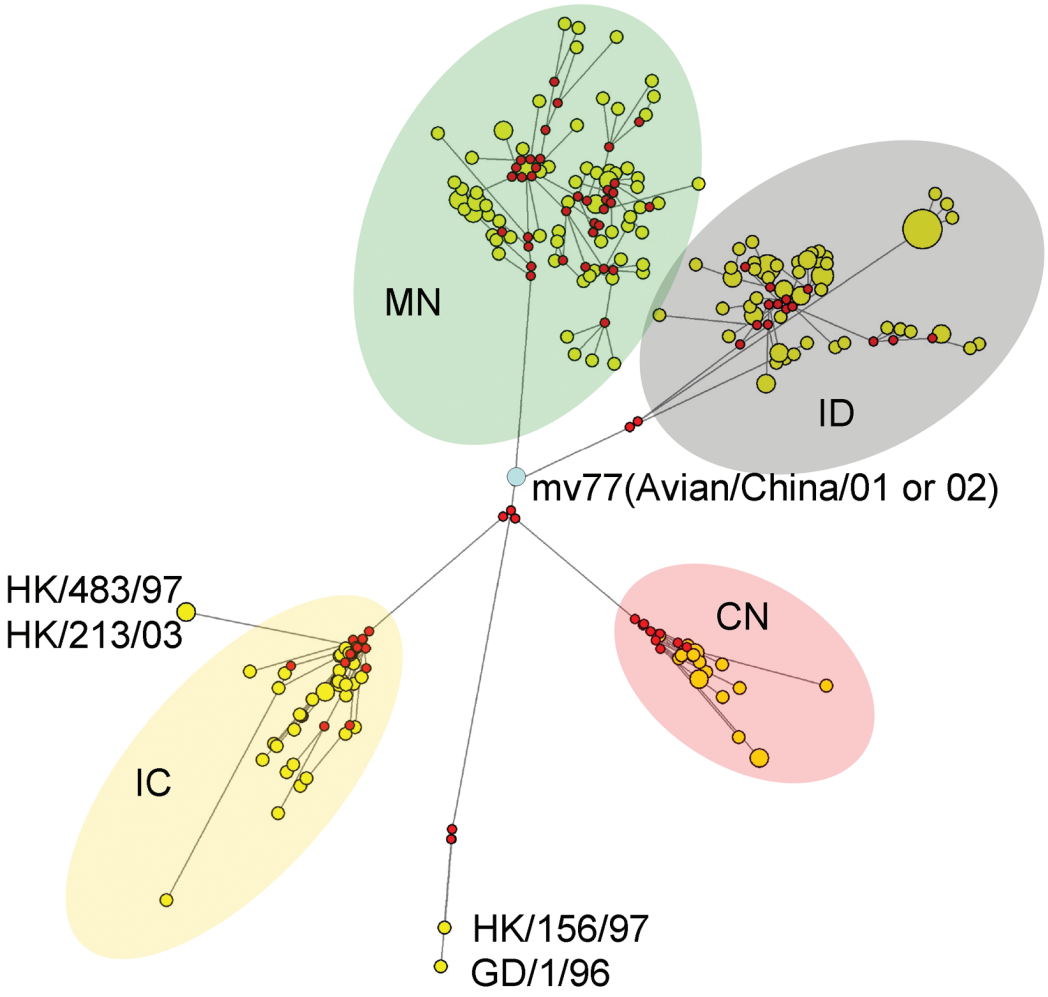

Figure 3. Median-joining network depicting the relationship between human H5N1 avian influenza viruses from other regions in the world and those from China. Colors are identical to those in Figure 2. *Gray,* Indonesia cluster (ID); *red,* Southern China cluster (CN); *yellow,* Indo-China cluster (IC); *green,* Middle Asia–North Africa cluster (MN). The key node is named as the virus strain. The area of each node is proportional to the number of sequences that the node represents. The ancestral node representing the original sequence type is marked by a blue circle (Mv77/avian/China/2001 or Mv77/avian/China/2002). The analysis is based on 215 hemagglutinin gene sequences in the National Center for Biotechnology Information Influenza Virus Resource of human virus strains that were separately collected from China, Azerbaijan, Bangladesh, Cambodia, Djibouti, Egypt, Indonesia, Iraq, Laos, Myanmar, Nigeria, Pakistan, Thailand, Turkey, and Vietnam. Human H5N1 influenza virus strains from the ID, CN, IC, and MN clusters are inferred to have the same ancestor from Southern China. The results show that strains A/Hong Kong/483/97 and A/Hong Kong/213/2003 in Hong Kong are haplotypes and genetically very close to the IC cluster viruses.

and Thailand, substandard methods of processing poultry products has led to more cases of infection in humans. A similar situation also appears in some North African countries such as Egypt and Nigeria.

To infer the common ancestor of human HPAI H5N1 viruses, the network of mutations for the HA gene sequences was constructed using Network software (version 4.5.1.0; Fluxus; available at http://www.fluxus-technology.com) after multiple alignment was conducted using ClustalW software (version 1.83). A Basic Local Alignment Search Tool (BLAST) search in the NCBI database of nucleotide sequences and homology analyses with an out-group (viruses from avian cases) indicated that a node (Mv77/avian/China/2001 or Mv77/avian/China/2002) containing 6 sequences could be treated as an ancestral node of all the 4 clusters (A/chicken/Hunan/23/2002 [accession no. CY028962], A/chicken/Zhengzhou/1/02 [accession no. DQ211923], A/Ck/HK/61.9/02 [accession no. AY575876], A/silkychicken/Shantou/4071/2002 [accession no. CY028951], A/duck/Guangxi/50/2001 [accession no. AY585375], and A/duck/Yunnan/862/2002 [accession no. CY028971]; data not shown). On the basis of the mutation network, we concluded that the ancestral sequence type first appeared in duck cases in Guangxi province in 2001 (A/duck/Guangxi/50/2001[H5N1]) (Figure 3). This evidence suggests that a virus from avian cases in 2001 or

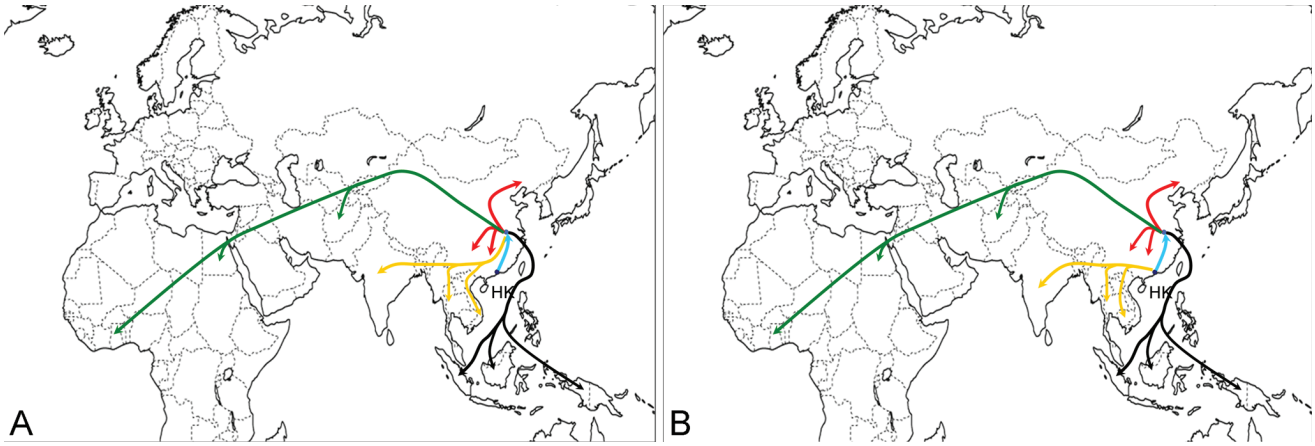

Figure 4. Maps of dispersal events of human H5N1 influenza viruses as inferred by parsimony through maximum likelihood phylogeny for 215 hemagglutinin gene sequences of viruses isolated from cases of infection in humans across Eurasia and Africa from 2004 through 2009. On the basis of the analysis depicted in Figure 2, we inferred that Southern China may be a common source of 4 major clusters of human H5N1 influenza virus. Different colors indicate the dispersal patterns of the different clusters. Interestingly, viruses in the Indo-China cluster (*yellow arrows*) could have originated in 1 of 2 different regions, Southern China (*A*) or Hong Kong (HK) (*B*). The Indo-China cluster viruses exhibit a dispersal pattern that indicates the early virus strains in Hong Kong may have evolved independently of the common ancestor (Mv77/avian/China/2001 or Mv77/avian/China/2002) in Southern China and spread directly into the Indian subcontinent.

2002 are common ancestors of the HPAI H5N1 viruses from cases in humans since 2004 (4 clusters), but GD/1/96 and A/Hong Kong/156/97 (HK/156/97) are the common ancestor of both human and avian H5N1 influenza viruses. In summary, current HPAI H5N1 viruses may have evolved from GD/1/96 and HK/156/97 by 2-step evolution through the China avian influenza viruses in 2001 or 2002. Through the mutation network, we also discovered that A/Hong Kong/483/97 and A/Hong Kong/213/2003 in Hong Kong are haplotypes (Figure 3) and genetically very close to the Indo-China cluster virus. Therefore, we think that Indo-China cluster viruses may have another dispersal pattern (Figure 4) whereby early virus strains in Hong Kong may have evolved independently or directly spread into Indo-China rather than from the common ancestor (Mv77/avian/China/2001 or Mv77/avian/China/2002) in Southern China.

On the basis of the N-glycosylation sites 100 and 170 in the HA protein (Figure 5), we divided viruses of the current epidemic among 3 categories: the GD196 type, the transitional type, and the Indonesia type. The GD196 type, without N-glycosylation sites 100 and 170, was found in several provinces of China (such as Guangdong and Xinjiang Uygur Autonomous Region) and in the Middle East and North Africa (Djibouti, Egypt, Azerbaijan, Iraq, Nigeria, Pakistan, and Turkey). BLAST search results of nucleotide sequences indicated that A/Bar-headed Goose/Qinghai/China/05 could be treated as an ancestral sequence of Middle Asia–North Africa cluster viruses (Table 2). The transitional type, characterized by the presence of an N-glycosylation site at either the 100 or the 170 position, was mainly found in China and the Indian subcontinent (Bangladesh, Cambodia, Laos, Myanmar, Thailand, and Vietnam). This virus type had been traced earlier to A/chicken/Zhejiang/24/2005 and A/chicken/Vietnam/5/2003 (Table 2). The virus of the Indonesia type, with N-glycosylation sites at both positions 100 and 170, mostly appeared in Indonesia. In this virus type, one early strain was traced to A/Ck/Indonesia/BL/2003 (Table 2). In conclusion, the 3 types of human influenza A(H5N1) virus presented distinctive clustering features in their geographic distributions. The Indonesia type was found only on Indonesian islands, and the GD196 and transitional types mostly appeared in other countries and areas. Our results show that the 4 different clusters have individual evolutionary features and provide further information concerning the classification of human H5N1 influenza virus strains.

We propose that Southern China was the center of the outbreak of highly pathogenic H5N1 avian influenza from which the virus radiated into Southeast Asia and Western and Northern China and made a final, progressive spread across Eurasia into the Middle East and North Africa. An initial dispersal from Southern China into the Indian subcontinent was followed by subsequent transmission to Indonesian islands. This strong phylogeographic structure provides an opportunity to trace the geographic origins of the human HPAI H5N1 virus

Table 2. Different Types of Human H5N1 Influenza Virus, Based on N-Glycosylation Site Analysis and Related Early Emerging Virus Strains

This table is available in its entirety in the online version of the *Journal of Infectious Diseases*

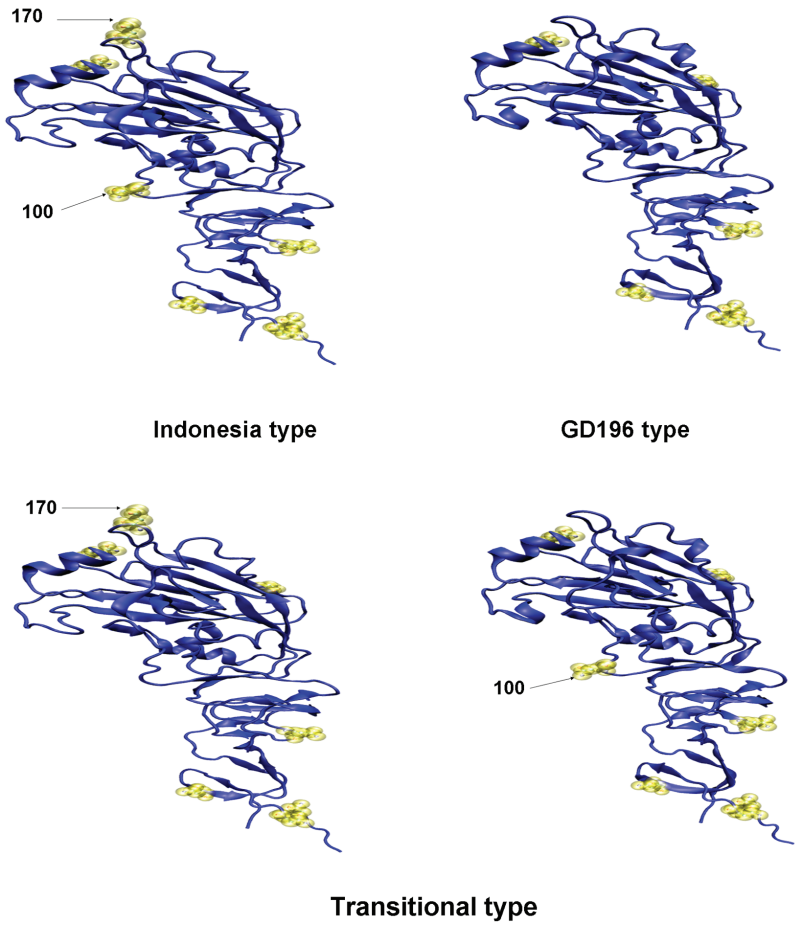

Figure 5. Models of the structure of the hemagglutinin proteins of representative strains of 4 different clusters of human H5N1 influenza virus, based on analysis of the N-glycosylation site. The hemagglutinin protein amino acid sequences were obtained from the NetNGlyc server; the predicted locations of N-glycosylation sites 100 and 170 are marked. The results of analysis of the N-glycosylation site show that the human influenza A(H5N1) virus presented distinctive clustering features in their geographic distributions. Viruses of the Middle Asia–North Africa cluster belong to the GD196 type or the original type, found in Guangdong and Xinjiang Uygur Autonomous Region of China and in the Middle East and North Africa, including in Djibouti, Egypt, Azerbaijan, Iraq, Nigeria, Pakistan, and Turkey. The Southern China and Indo-China cluster viruses belong to the transitional type and were found in most provinces of China and in the Indian subcontinent, including in Bangladesh, Cambodia, Laos, Myanmar, Thailand, and Vietnam. But viruses of the Indonesia cluster belong to Indonesia type, which is characterized by the presence of N-glycosylation sites at both positions 100 and 170; these viruses appeared only in Indonesia.

strains. In summary, the genetic evidence presented here provides a clear picture of multiple clusters of human HPAI H5N1 virus and indicates that the source was Southern China.

References

1. Zhou JY, Shen HG, Chen HX, et al. Characterization of a highly pathogenic H5N1 influenza virus derived from bar-headed geese in China. J Gen Virol **2006**; 87:1823–1833.
2. Saw T, Tam J, Liu KK, et al. Isolation of avian influenza A(H5N1) viruses from humans—Hong Kong. MMWR Morb Mortal Wkly Rep **1997**; 46:1204–1207.
3. Food and Agricultural Organization (FAO). Bird flu outbreaks continue in Africa, Asia, Europe and Near East. FAO Web site. http://www.fao.org/newsroom/en/news/2006/1000266/index.html. Published 6 April 2006. Accessed 1 May 2008.
4. Xu CL, Dong LB, Xin L, et al. Human avian influenza A (H5N1) virus infection in China. Sci China C Life Sci **2009**; 52:407–411. doi:10.1007/s11427-009-0067-7. Published 27 May 2009. Accessed 10 June 2009.
5. Webster RG, Govorkova EA. H5N1 influenza—continuing evolution and spread. N Engl J Med **2006**; 355:2174–2177.
6. World Health Organization (WHO). Cumulative number of confirmed human cases of avian influenza A/(H5N1) reported to WHO. WHO Web site. http://www.who.int/csr/disease/avian_influenza/country/en/index.html. Accessed 24 January 2010.
7. Kandun IN, Wibisono H, Sedyaningsih ER, et al. Three Indonesian clusters of H5N1 virus infection in 2005. N Engl J Med **2006**; 355:2186–2194.
8. Webster RG, Hulse-Post DJ, Sturm-Ramirez KM, et al. Changing epidemiology and ecology of highly pathogenic avian H5N1 influenza viruses. Avian Dis **2007**; 51:269–272.
9. Kumar S, Tamura K, Jakobsen IB, et al. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **2001**; 17:1244–1245.
10. Hans-Jürgen B, Peter F, Arne R. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol **1999**; 16:37–48.

11. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics **2004**; 4:1633–1649.
12. Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res **2003**; 31:3381–3385.
13. William H, Andrew D, Klaus S. VMD: Visual molecular dynamics. J Mol Graph **1996**; 14:33–38.
14. Wallace RG, Hodac H, Lathrop RH, Fitch WM. A statistical phylogeography of influenza A H5N1. Proc Natl Acad Sci U S A **2007**; 104: 4473–4478.
15. Wan XF, Nguyen T, Davis CT, et al. Evolution of highly pathogenic H5N1 avian influenza viruses in Vietnam between 2001 and 2007. PLoS One **2008**; 3:e3462.
16. Duan L, Campitelli L, Fan XH, et al. Characterization of low-pathogenic H5 subtype influenza viruses from Eurasia: implications for the origin of highly pathogenic H5N1 viruses. J Virol **2007**; 81:7529–7539.
17. Zhao Z, Shortridge KF, Garcia M, Wan XF. Genotypic diversity of H5N1 highly pathogenic avian influenza viruses. J Gen Virol **2008**; 89:2182–2193.
18. World Health Organization/World Organization for Animal Health/Food and Agriculture Organization H5N1 Evolution Working Group. Toward a unified nomenclature system for highly pathogenic avian influenza virus (H5N1). Centers for Disease Control and Prevention Web site. http://www.cdc.gov/EID/content/14/7/e1.htm. Published July 2008. Accessed 6 September 2008.
19. Liu J, Xiao H, Lei F, et al. Highly pathogenic H5N1 influenza virus infection in migratory birds. Science **2005**; 309:1206. doi:10.1126/science.1115273. Published 6 July 2005. Accessed 25 May 2009.
20. Chen H, Smith GJ, Zhang SY, et al. Avian flu: H5N1 virus outbreak in migratory waterfowl. Nature **2005**; 436:191–192.
21. World Health Organization. Avian influenza: frequently asked questions. Wkly Epidemiol Rec **2005**; 80:377–384.
22. Dybing JK, Cherry SS, Swayne DE, Suarez DL, Perdue ML. Distinct pathogenesis of Hong Kong-origin H5N1 viruses in mice compared to that of other highly pathogenic H5 avian influenza viruses. J Virol **2000**; 74:1443–1450.
23. Chen H, Li Y, Li Z, et al. Properties and dissemination of H5N1 viruses isolated during an influenza outbreak in migratory waterfowl in western China. J Virol **2006**; 80:5976–5983.