

The mutation network for the hemagglutinin gene from the novel influenza A (H1N1) virus

Consortium for Influenza Study at Shanghai (HE YunGang¹, DING GuoHui¹, BIAN Chao¹, HUANG Zhong¹, LAN Ke¹, SUN Bing¹, WANG XueCai¹, LI YiXue^{1,9}, WANG HongYan², WANG XiaoNing^{2,5}, YANG Zhong^{2,9}, ZHONG Yang^{2,9}, JIN WeiRong^{3,14}, XIONG Hui³, DAI JianXin^{4,6}, GUO YaJun^{4,6}, WANG Hao^{4,6}, CHE XiaoYan⁷, WU Fan⁸, YUAN ZhenAn⁸, ZHANG Xi⁸, CAO ZhiWei^{9,10}, ZHOU XiaoNong¹¹, ZHOU JiaHai¹², MA ZhiYong¹³, TONG GuangZhi¹³, ZHAO GuoPing^{1,2,3,14†} & JIN Li^{1,2,3†})

¹ Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences (Key Laboratory of Computational Biology, CAS-MPG Partner Institute of Computational Biology; Center of Bioinformatics, Key Lab of System Biology; Lab of Synthetic Biology, Institute of Plant Physiology & Ecology; Institute Pasteur), Shanghai 200031, China;

² School of Life Sciences, Fudan University, Shanghai 200433, China;

³ Shanghai-MOST Key Laboratory for Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China;

⁴ Cancer Institute, Second Military Medical University, Shanghai 200433, China;

⁵ School of Life Sciences and Technology, South China University of Technology, Guangzhou 510641, China;

⁶ National Engineering Research Center for Antibody Drugs, Shanghai 201203, China;

⁷ Zhujiang Hospital, Southern Medical University, Guangzhou 510282, China;

⁸ Shanghai Municipal Center for Disease Control & Prevention, Shanghai 200336, China;

⁹ Shanghai Center for Bioinformation Technology, Shanghai 200235, China;

¹⁰ School of Life Sciences and Technology, Tongji University, Shanghai 200092, China;

¹¹ National Institute for Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai 200025, China;

¹² Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China;

¹³ Shanghai Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Shanghai 200241, China;

¹⁴ National Engineering Research Center for Biochip Technology, Shanghai Biochip Co. Ltd, Shanghai 201203, China

A mutation network for the hemagglutinin gene (*HA*) of the novel type A (H1N1) influenza virus was constructed. Sequence homology analysis indicated that one *HA* sequence type from the viruses mainly isolated from Mexico was likely the original type in this epidemic. Based on the 658A and 1408T mutations in *HA*, the viruses evolving into this epidemic were divided into three categories, the Mexico, the transitional and the New York type. The three groups of viruses presented distinctive clustering features in their geographic distributions.

influenza, novel type A (H1N1) virus, hemagglutinin, gene mutation

On April 15th and 17th 2009, the Centers for Disease Control and Prevention of the United States in Atlanta identified isolations from two children as a novel type A (H1N1) influenza virus^[1]. Sequence homology analysis indicated that the novel virus was generated from a reassortment of multiple influenza viruses (from swine, avian and humans). Further studies found that the novel virus appeared in Mexico as early as on March 18, 2009. Some patients infected by this novel virus developed

serious clinical syndromes and several patients died. As of Beijing time 14:00 May 20, 2009, 49 countries reported 9830 identified cases of the novel influenza virus A (H1N1) infection, including 79 deaths (http://www.who.int/csr/don/2009_05_19/en/index.html).

The genome of influenza virus A contains 8 separate

Received May 18, 2009; accepted May 22, 2009

doi: 10.1007/s11434-009-0428-4

†Corresponding authors (email: ljin007@gmail.com; gpzhao@sibs.ac.cn)

RNA segments, which code for different proteins which play certain roles in construction and duplication of the influenza virus. Among these proteins, hemagglutinin (HA) attracted much more attention. HA is located on the outside of the virus membrane. In hosts, the HA binds to sialic acids on the surface of target cells as one of the necessary steps for invasion. Therefore, HA is primarily responsible for the host range of influenza virus and immunity response of hosts to the infection^[2].

Sequences of the *HA* gene from the novel influenza virus A (H1N1) published before May 12, 2009 were downloaded from database of GISAID^[3]. Fifty-four full-length sequences with high quality passed a data clean procedure. The network of mutations for the sequences was constructed using NETWORK (version 4.5.1.0, <http://www.fluxus-technology.com/>) after the multiple alignment was conducted using ClustalW1.83^[4] (Figure 1). Every node in the network represented a sequence

type observed in the 54 full-length sequences. BLAST search in NCBI database of nucleotide sequences and homology analysis with an outgroup (viruses from swine) indicated that a node containing 4 sequences could be treated as an ancestral node of all the observed sequences^[5] (Figure 1, accession number in NCBI: GQ117067, FJ982430, GQ149692, FJ998208). Based on the mutation network, the ancestral sequence type first appeared in the late August of 2008 (with standard deviation 0.24 year, that is about 3 months), giving an estimation of 1.83 mutations (time unit in mutation) before May 2009^[6] and the mutation rate of the *HA* gene as 2.8 bases per year^[7].

Based on the 658A and 1408T mutations in *HA*, the viruses evolving into the epidemic could be divided into three categories, the Mexico, the transitional and the New York type. The Mexico type virus was found in Mexico, South of the United States (such as Texas,

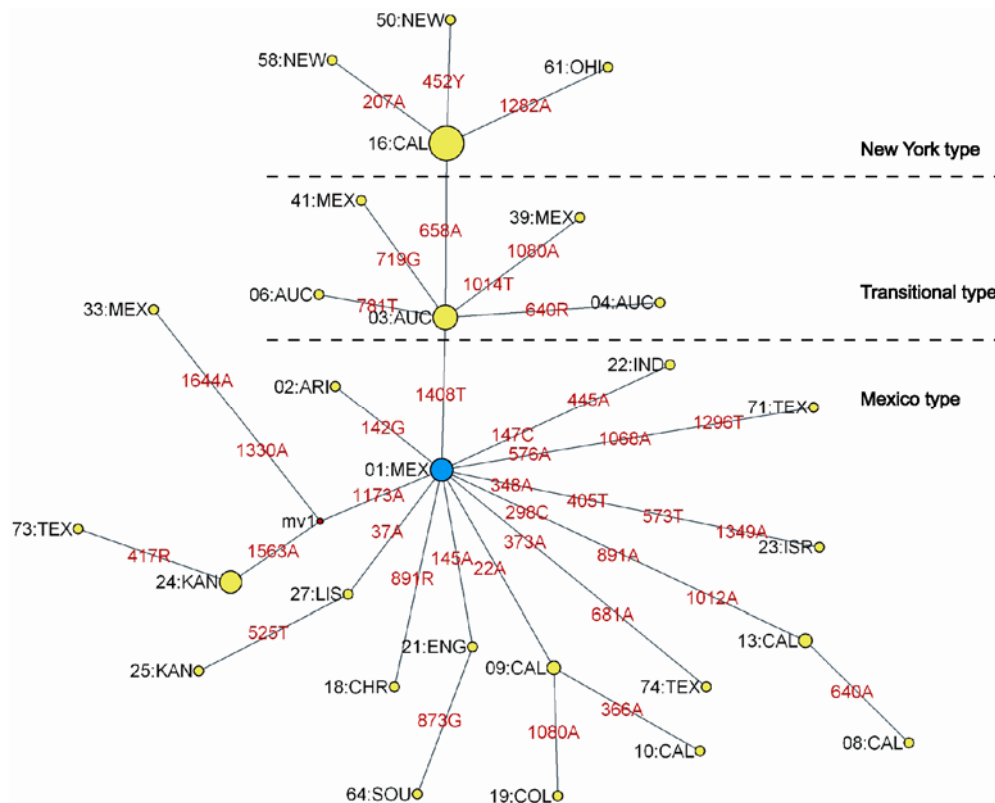


Figure 1 Mutation network for *HA* of the novel influenza virus A (H1N1). Each node is named in sequence ID and the first three letters of the name of the location where the sample was isolated. When multiple sequences are presented at one node, information of one sequence will be used randomly to name the node. The area of each node is in proportion to the number of sequences the node represents. The ancestral node representing the original sequence type is marked in blue. Location and status of mutations are marked on edges in red with the original sequence type as a reference. Based on information of the sequences in GISAID database, 54 samples were separately collected from Arizona (ARI), California (CAL), Colorado (COL), Indiana (IND), Kansas (KAN), Massachusetts (MAS), Michigan (MIC), New York (NEW), Ohio (OHI), South Carolina (SOU), Texas (TEX) of the United States; Christchurch (CHR) and Auckland (AUC) of New Zealand; Canada (CAN); Denmark (DEN); England (ENG); Israel (ISR), Lisbon (LIS) of Portugal; Mexico (MEX); Netherland (NET).

California and Arizona), several European countries (England, Denmark, Portugal, Netherland and Germany), and New Zealand. Most of the European patients found in the early stage of the pandemic had been to Mexico. One patient found in Sichuan Province of China also carried the novel influenza virus of the original type. The transitional type characterized by 1408T mutation was mainly found in New Zealand and Mexico. Sequences published after May 12, 2009 showed the type that appeared in Thailand, Texas, Michigan and Wisconsin (data not shown). The virus of the New York type with both 1408T and 658A mutations mostly appeared in North and East of America, such as New York, Massachusetts, Ohio and Canada. The virus of this type also appeared in California but the cases were rare in GISAID sequence database. The New York type virus has been found in New Jersey and Nebraska based on the recently released *HA* sequences (data not shown). To

summarize, the three groups of novel influenza A(H1N1) virus presented distinctive clustering features in their geographic distributions (based on the sequencing data released before May 20, 2009). The New York type was found in North and East of America; the Mexico and transitional types mostly appeared in other countries and areas.

The 658A mutation carried by the New York type mutated from Serine to Threonine at residue 220 of HA. This residue is located in the sialic acid bonding domain of HA and close to the important 200-loop region (221—228 amino acid)^[8]. It has been reported that the host of 1918 influenza virus A (H1N1) switched between humans and avian when the residues at 190 and 225 sites were mutated^[9]. However, there is no solid evidence in this study to support the effect of 658A mutation on binding capability of HA and sialic acids. Further studies are necessary to address this issue in future.

- 1 Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team. Emergence of a Novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med*, 2009, doi: 10.1056/NEJMoa0903810
- 2 Kobasa D, Takada A, Shinya K, et al. Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. *Nature*, 2004, 431: 703—707
- 3 Bogner P, Capua I, Lipman D J, et al. A global initiative on sharing avian flu data. *Nature*, 2006, 442(7106): 981
- 4 Bandelt H J, Forster P, Sykes B C, et al. Mitochondrial portraits of human populations using median networks. *Genetics*, 1995, 141: 743—753
- 5 Bao Y, Bolotov P, Dernovoy D, et al. The influenza virus resource at the National Center for Biotechnology Information. *J Virol*, 2008, 82: 596—601
- 6 Saillard J, Forster P, Lynnerup N, et al. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet*, 2000, 67: 718—726
- 7 Ferguson N M, Galvani A P, Bush R M. Ecological and immunological determinants of influenza evolution. *Nature*, 2003, 422: 428—433
- 8 Gamblin S J, Haire L F, Russell R J, et al. The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science*, 2004, 303: 1838—1842
- 9 Tumpey T M, Maines T R, Van Hoeven N, et al. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science*, 2007, 315: 655—659