Subtyping of type A influenza by sequencing the variable regions of *HA* gene specifically amplified with RT-PCR

Consortium for Influenza Study at Shanghai (YAN An[1*], DING GuoHui[2*], ZHOU ZhenFeng[1], DONG Hui[1], ZHANG YaKun[1], ZHU Lei[3], HE YunGang[3], ZHANG GuoQing[4], LI YiXue[2,4], SUN Bing[2], HUANG Zhong[2], LAN Ke[2], JIN Li[1,2,5], WANG HongYan[5], WANG XiaoNing[5,6], YANG Zhong[4,5], ZHONG Yang[4,5], DAI JianXin[7,8], GUO YaJun[7,8], WANG Hao[7,8], CHE XiaoYan[9], WU Fan[10], YUAN ZhenGan[10], ZHANG Xi[10], CAO ZhiWei[4,11], ZHOU XiaoNong[12], ZHOU JiaHai[13], MA ZhiYong[14], TONG GuangZhi[14], ZHAO GuoPing[1,2,3,5†], JIN WeiRong[3†] & XIONG Hui[1†])

[1] Shanghai-MOST Key Laboratory for Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China;
[2] Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences (Lab of Computational Genomics, 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;
[3] National Engineering Research Center for Biochip Technology, Shanghai Biochip Co. Ltd, Shanghai 201203, China;
[4] Shanghai Center for Bioinformation Technology, Shanghai 200235, China;
[5] School of Life Sciences, Fudan University, Shanghai 200433, China;
[6] School of Life Sciences and Technology, South China University of Technology, Guangzhou 510641, China;
[7] Cancer Institute, Second Military Medical University, Shanghai 200433, China;
[8] National Engineering Research Center for Antibody Drugs, Shanghai 201203, China;
[9] Zhujiang Hospical, Southern Medical University, Guangzhou 510282, China;
[10] Shanghai Municipal Center for Disease Control & Prevention, Shanghai 200336, China;
[11] School of Life Sciences and Technology, Tongji University, Shanghai 200092, China;
[12] National Institute for Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai 200025, China;
[13] Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China;
[14] Shanghai Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Shanghai 200241, China

The outbreak of a novel influenza A (H1N1) virus across the globe poses a threat to human health. It is of paramount importance to develop a rapid, reliable and inexpensive diagnostic procedure. Based on the bioinformatic information from public database, primers specific for influenza A virus surface protein haemagglutinin (HA) of several subtypes (including H1, H2, H3, H5, H7 and H9) were designed. Primer-specific PCR products were subjected to sequencing for accurately distinguishing H1 and H3 subtypes from others. This sequencing-based detection method will not only be applied to rapid detection and simultaneous subtype identification of new influenza A virus H1N1, but also provide the strategies to monitor other new types of influenza virus with explosive potential.

influenza A virus, *HA* gene, H1 and H3 subtypes, specific amplification by RT-PCT, DNA sequencing

A novel influenza A virus H1N1, originally identified in April 2009, is now spreading rapidly from human to human throughout the world (http://www.who.int/csr/don/2009_05_15/en/index.html). In China, more and more infected cases have been reported in several regions, such as Sichuan, Shandong, Peking and Guang-dong, which indicates that this new type of influenza A virus has embarked locally. This might lead to the rapid

spreading among given populations in certain areas. Therefore, prompt diagnosis of influenza infection would facilitate effective diagnosis of suspects, patient management as well as infection prevention and control. The recent availability of polymerase chain reaction (PCR) from trace samples makes it possible to rapidly detect viral nucleic acid sequences in specimens. However, due to the high degrees of variation in virus genes (mainly the viral envelopment glycoproteins hemagglutinin (*HA*) and neuraminidase (*NA*) genes)[1] and their uncertainty of variation, monitoring and diagnosis of H1N1 infection usually lag behind the ongoing prevention or outbreak control. Therefore, establishing a rapid, specific, hypersensitive detection procedure is one of the most urgent tasks to keep the infection of the novel H1N1 virus under control.

With the progress in molecular biology and modern biotechnology, molecular techniques have been widely used for the rapid detection of influenza virus. For instance, most of the molecular detection criteria recommended recently by World Health Organization (WHO) (http://www.who.int/csr/resources/publications/swineflu/ realtimeptpcr/en/index.html) are based on real-time reverse transcription RT-PCR. With the advantage of easy preparation, high sensitivity and less time-consuming, PCR-based detection methodology still has several defects, such as false-positive results due to the nonspecific primers and/or probes; false-negative results due to unavailability of specific primers and/or probes to novel or mutant virus, etc.[2]. These disadvantages might largely hamper the definitive diagnosis of influenza A virus H1N1. Instead, the sequencing-based technique is more specific and accurate, and thus recommended by WHO as the diagnosis methods. In the present study, according to the bioinformatic analysis, we have designed a panel of primers specific for H1 and H3 subtypes located in the comparatively conserved domains. After PCR amplification, conventional sequencing was performed which helped to determine the subtypes of influenza A virus with the alignment of the known subtype sequence database and the similarity clustering analysis. H1N1 subtype can thus be rapidly and accurately identified.

Based on the strategy described above, the primers were designed according to the following principles: (i) Primers were targeted to the comparatively conserved domain of H1 and H3 subtypes; (ii) the amplified sequences can be used for identifying the six subtypes H1,

H2, H3, H5, H7 and H9, especially for distinguishing the normal seasonal H1N1 virus strains from 2009 North American H1N1 virus strains. The H1, H2, H3, H5, H7 and H9 subtypes were chosen for testing according to the previous reports that they were able to infect humans as well as to transmit between humans in a given time, some of which might even have the potential to cause explosive outbreak.

The latest sequence data of influenza virus and corresponding descriptive information were collected from NCBI Influenza Virus Resource. Up to May 2009, there had been about 187327 pieces of genome sequences encompassing 23094 samples in that database. The local integrated information system of influenza virus was established (http://lifecenter.sgst.cn/flu). Since the number of influenza sequences was huge and the number of each subtype varied in the databases, we selected 50 complete HA sequences newly reported for each subtype from our database for further study (http://lifecenter.sgst. cn/flu/doc/flu_1_2_3_5_7_9-1.htm). Among the selected 300 reference HA sequences 197 are from avian, 69 from human, 15 from swine, 1 from Plateau pika, and 18 from environment-hosted virus. As the subtypes of influenza A viruses have been proved to be capable of cross-infecting among humans, swine and avian, the avian or environment-hosted viruses were also included in our study. Three hundred pieces of sequences were analyzed by multiple sequence alignment using Clustalw software[3], and the comparatively conserved HA regions ranging from 750 nt to 1100 nt in the comparatively conserved domain was selected for primer designing (Figure 1).

In the present study, only the primers for detecting H1 and H3 subtypes are described and tested since no RNA samples of other subtypes are available currently (Table 1).

Based on the similarity of the comparatively variable sequences in *HA* genes (750 nt to 1100 nt), a cluster tree was pictured to test the efficiency of our subtyping method (Figure 2). Instead of relying on the evolution model, the average distance clustering was used for subtyping that was more likely based on the original sequence similarity. As shown in Figure 2, the primers targeted at sequences within "the comparatively variable regions" located in "the comparatively conserved domain" could accurately subtype H1, H2, H3, H5, H7 and H9 individually. More importantly, our clustering results indicated that the 2009 novel H1N1 belonged to a dis-
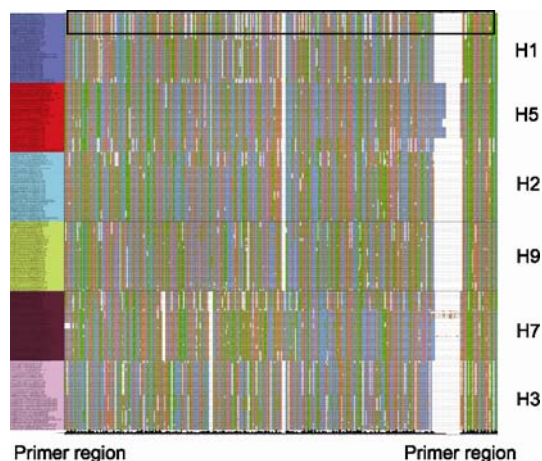
Figure 1 The comparatively variable sequences ranging from 750 nt to 1100 nt and the comparatively conserved regions flanking these sequences. The red box represents the 2009 novel H1N1 virus.

Figure 2 Subtyping of influenza A viruses by average distance clustering analysis based on the similarity of variable regions. Bootstrapping between main types was 100%. The arrow indicates the 2009 novel H1N1 virus.

tinct group.

To further verify the ability of primers (H1F722 and H1F721 mixture) and H1R1062 to separate the previously and recently reported H1N1 virus, we selected 100 pieces of HA sequences from 2009 North American H1N1 virus and 100 pieces from the previous H1N1 virus. Results from multiple sequence alignment and average distance clustering analysis showed that previous and novel H1N1 viruses could be clearly distin-
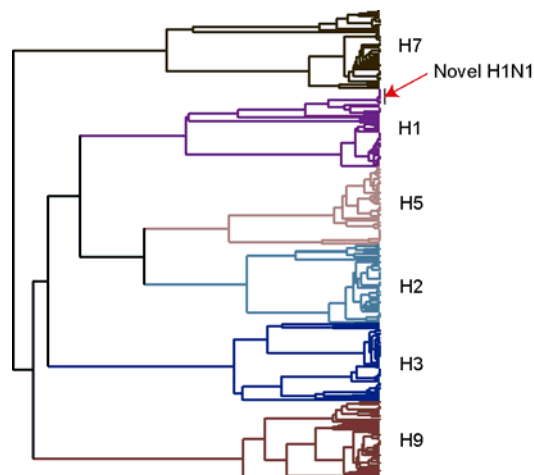
guished from each other by our designed primers (date not shown).

Six virus samples isolated from either humans or swine were selected to test the specificity and sensitivity of primers and optimize PCR conditions. As shown in Figure 3, the pair of primers for H1 subtyping could specifically amplify the samples from human and swine H1 subtype virus but not the H3 subtype without nonspecific products. So is the case for the pair of primers

Table 1 Specific primers used for the amplification of H1 and H3 subtypes in influenza A strains[a]

Subtype	Primers	Location	Sequences (5′ to 3′)	Size of product (bp)
	H1F722	722－744	AAGGGAGAATGAACTATTACTGG	
H1	H1F721	721－744	GAAGGAAGAATCAACTACTACTGG	341－342
	H1R1062	1046－1062	AATGAAACCGGCAATGG	
	H3F731	731－752	GCAGAATAAGCATCTATTGGAC	
H3	H3R1064	1048－1064	ATGAAACCCGCGATTGC	334－337
	H3R1067	1048－1067	TCTATAAACCCGGCTATTGC	

a) The forward primers H1F722 and H1F721 were mixed at 1:1 ratio for subtyping of H1, while the reverse primers H3R1064 and H3R1067 were mixed at 1:1 ratio for subtyping of H3.
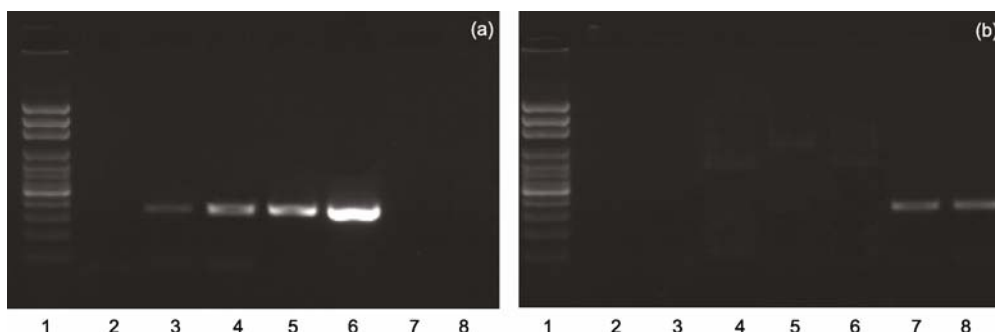

Figure 3 PCR products amplified by H1 and H3 subtyping primers using different samples as templates. (a) PCR product amplified by H1 subtyping primer; (b) PCR product amplified by H3 subtyping primer. 1, 100 bp DNA Ladder; 2—8, PCR products from different templates. 2, Negative control; 3, cDNA of human H1 reference strain; 4, cDNA of human H1 virus strain; 5, swine H1 standard strain cDNA; 6, plasmid containing H1N1 *HA* genes; 7, human H3 standard strain cDNA; 8, plasmid containing swine H2N2 *HA* gene.
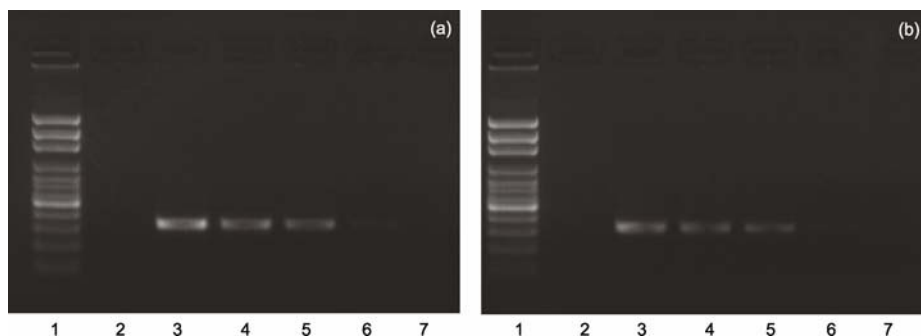
Figure 4 Primer sensitivity test. (a) Amplification of different concentration of human H1 standard strain cDNA by H1 subtyping primers; (b) amplification of different concentration of human H3 standard strain cDNA by H3 subtyping primers. 1, 100 bp DNA Ladder; 2, negative control; 3—7, PCR products amplified using cDNA with dilutions 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5} and 10^{-6}, respectively.
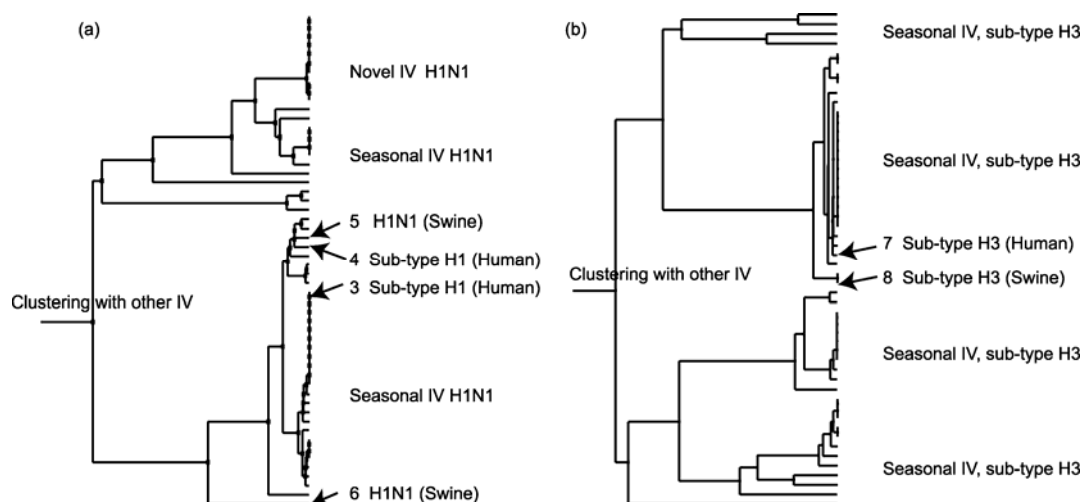

Figure 5 Subtyping based on the sequencing of comparatively variable regions from human and swine H1 and H3 subtypes. (a) and (b) show the H1 and H3 branches from the clustering tree (data shown in online supplement material 1) of comparatively variable regions from the tested and standard virus, respectively. The tested virus can be categorized into its expected branch accurately by clustering analysis of sequence similarity. Note: the names of virus samples of 3—8 are consistent with those in Figure 3.

for subtyping of H3.

To evaluate the sensitivity of the specific primers, the rough RNA extract from virus strains was further tested. 2 μg of H1 and H3 RNA each was used in a total volume of 40 μL for reverse transcription. 2 μL of the reverse transcribed product was used as template in the subsequent PCR system. As illustrated in Figure 4, PCR products from templates with 10^5 dilutions could still be apparently observed, and be subjected to sequencing. These results indicated the high sensitivity and productivity of our designed primers during PCR reaction.

Results from sequencing (data not shown) of PCR products from human and swine influenza A virus and

comparative average distance clustering analysis (Figure 5) with 300 standard HA sequences showed that the virus strains 3—8 could be categorized into the expected position. Our data presented here indicated that these sets of primers could not only accurately identify the novel H1N1 virus, but also identify other human susceptibility HA subtype (such as H3). In addition, the method introduced here was able to identify the newly mutated H1N1 subtype according to the specific HA sequence, which might be potential for predicting virus variation and epidemics. With the long-term data collection, our work might be helpful for the government to enact national preventive measures.

1   Ghedin E, Sengamalay N A, Shumway M, et al. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature, 2005, 437: 1162－1166

2   Leutenegger C M, Higgins J, Matthews T B, et al. Real-time TaqMan PCR as a specific and more sensitive alternative to the branched-

chain DNA assay for quantitation of simian immunodeficiency virus RNA. AIDS Res Hum Retroviruses, 2001, 17: 243－251

3   Li K B. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. Bioinformatics, 2003, 19: 1585－1586