Genetic diversity of hemagglutinin gene of A(H1N1)pdm09 influenza strains isolated in Taiwan and its potential impact on HA-neutralizing epitope interaction

Running title: HA mutation in Ab-binding interface

Krzysztof Łepek [1*], Beata Pająk [2,3 * ✉], Paweł Siedlecki[4], Marcin Niemcewicz[5], Janusz Kocik[5], Ho-Sheng Wu[6], Ji-Rong Yang[6], Krzysztof Kucharczyk[1], and Bogusław Szewczyk[3]

[*] K. Lepek and B. Pajak contributed equally to an article

[1]University of Gdansk and Medical University of Gdansk, Gdansk, Poland; [2] BioVectis Ltd, Warsaw, Poland; [3]Mossakowski Medical Research Centre, Warsaw, Poland; [4]Institute of Biochemistry and Biophysics, Warsaw, Poland; [5]Military Institute of Hygiene and Epidemiology, Pulawy, Warsaw, Poland; [6]Taiwan Centers for Disease Control, Taipei, Taiwan

Key words: A(H1N1) influenza, HA, neutralizing Ab, MSSCP, Taiwan isolates

✉ Correspondence to:

Beata Pajak, PhD, BioVectis Ltd,

Pawińskiego 5a/blok d, 02-106 Warsaw, Poland

Tel/fax: (+48) 22 6687147/ 22 6687164

e-mail: beata.pajak@biovectis.com

1**Abbreviations and acronyms**

2Ab – antibody

3CDC – Centers for Disease Control

4HA – hemagglutinin

5MSSCP – Multitemperature Single Strand Conformation Polymorphism

6PCR – Polymerase Chain Reaction

Abstract

Pandemic influenza A(H1N1)pdm09 virus is a global health threat and between 2009-2011 it became the predominant influenza virus subtype circulating in the world. We describe the MSSCP (Multitemperature Single Strand Conformation Polymorphism) analysis of the hemagglutinin (*HA*) region encompassing major neutralizing epitope in pandemic influenza isolates from Taiwan. In isolates obtained in 2010 and 2011, several genetically distinct changes have appeared. The majority of changes in HA protein, have not resulted in significant modifications, however three modifications were localized in epitope E of H1 and one was part of interface binding antibodies BH151 and HC45, which might make the current vaccine less effective.-Taking into account the possibility of the emergence of influenza A with antibody evading potential, the MSSCP method provides an alternative approach for detection of minor variants which escape detection by conventional Sanger sequencing.

Introduction

Influenza A is a serious disease causing recurring outbreaks with significant negative effects on the general health status globally as well as on the economy of affected populations. In April of 2009 a novel strain of (H1N1) swine-origin influenza A was reported in Mexico and the United States.[1-3] The rapid spread of this virus to over 70 countries of the world urged the WHO to raise its pandemic alert level to the highest phase 6 by June 11, 2009. This new A(H1N1)pdm09 virus was found to be a triple reassortant containing a combination of gene segments of swine, human and avian origin.[4-6] The 2009 pandemic influenza A turned out to be relatively mild in its symptoms when compared to other strains responsible for previous pandemics. Most individuals infected with the 2009 pandemic H1N1 strain usually developed uncomplicated illness with full recovery within a week. However, this may be a transient situation because influenza viruses are constantly undergoing changes, both by antigenic drift (mutations due to the error-prone genomic RNA amplification) as well as by antigenic shift (exchange of genomic segments during co-infection by two different strains). The best example of these phenomena was the 1918 Spanish flu pandemic virus. The initial outbreak was relatively mild in terms of clinical impact, but had acquired higher virulence when it returned in the next season. Changes are in some cases gradual and more predictable. Despite this fact the rate of genome mutation is high enough to cause the replacement of circulating strains every 3-5 years with variants that underwent antigenic changes sufficient for substantial or complete evasion of existing antibody response. Therefore, it is crucial to monitor, not only the mutations leading to resistance against antiviral drugs, but also those that modify the most important viral epitopes. The latter lead to gradual weakening of the antigen-antibody interactions and therefore reduce the efficacy of existing vaccine.

The viral surface protein hemagglutinin (HA) is the primary target for neutralizing antibodies in natural infections.[7] The antigenic variation of HA is the main mechanism

employed by the influenza virus to escape the response of the host immunological system. The HA of the pandemic A(H1N1)pdm09 strain has changed relatively little in its genetic and antigenic characteristics since it emerged in 2009. In 2010 a genetically distinct variant containing several amino acid changes in hemagglutinin and neuraminidase has emerged in Singapore, Australia and New Zealand. It was a dominant strain in the second and third quarters of 2010 in these countries, although it did not represent a significant antigenic change of the virus.[8]

In this study we present an analysis of the HA region encompassing the main neutralizing epitope of HA from Taiwan isolates of flu seasons 2009-2011. Isolates of 2010 and 2011 from these samples show significant changes in HA amino acid composition. Three of the detected mutations are likely to affect the virus-antibody interaction.

Results

Nineteen clinical specimens (throat or nasal swabs) from outpatients with influenza-like illnesses, who developed severe complications, were collected in Taiwan during the 2009-2011 flu season (Table 1). To determine viral genotype, partial nucleotide sequences (positions 393-1182) of the *HA* gene were first analyzed by RT-PCR using primers published by WHO. All the isolates in this study were characterized as A/California/07/2009-like viruses. To verify whether the observed clinical diversity could be correlated with genetic alterations or the presence of minor genetic variants, the isolates were further analyzed. Bearing in mind the crucial role of hemagglutinin mutations for influenza virus virulence, representative *HA* gene fragments encompassing nucleotides 125 to 302 were amplified from cDNAs as described in Materials and Methods. This region corresponds to a fragment of influenza virus HA1 polypeptide starting 25 amino acids (H1 numbering is used throughout this paper) after the N-terminal signal peptide of hemagglutinin. A crucial part of the influenza A/H1N1 epitope reacting with neutralizing antibodies is located within this region. To check for the presence of minor genetic variants of the A(H1N1)pdm09 pandemic strains within the obtained amplicones, we performed MSSCP (Multitemperature Single Strand Conformation Polymorphism) analysis.

MSSCP is a native electrophoretic separation performed under sequentially changed gel temperature. This improves the sensitivity of mutation detection and reduces time of analysis. The temperature changes increase the probability for the PCR products to adopt different ssDNA conformations during the electrophoretic run if they contain nucleotide substitutions All the amplified *HA* fragments, including corresponding fragments of the reference seasonal (s) (A/Brisbane/59/2007) and pandemic (p) (A/Mexico/4486/09) strains, were denatured and the resulting ssDNA fragments were subjected to the native electrophoresis in optimal conditions for the MSSCP analysis (15-10-5 °C, 450 Vxh/per

phase, 10% PA). Results of this experiment (after visualization with silver stain) are shown in Fig. 1. According to the electrophoretic profiles (Fig. 1), none of the samples contains fragments corresponding to the predominant influenza A seasonal strain (s) which excludes the possibility of co-infection with seasonal and pandemic strains. Samples designated as 2009-02626, 2009-00940, 2009-08542, 2010-00842, 2010-06031, 2011-02054, 2011-00623, 2009-06078, 2009-04909, 2009-00937, 2009-08575, 2011-04512, 2011-02068 and 2011-704611 exhibited MSSCP profiles identical to the reference pandemic strain, while the electrophoretic profiles of five samples: 2010-03994, 2011-01219, 2010-01164, 2010-05270 and 2010-05347 were different from that of the pandemic reference strain. For further analysis, if profiles reflected distinct DNA sequences, ssDNA bands from the samples indicated by arrows in Fig. 1 were extracted from the gel, re-amplified as described in Materials and Methods, and the PCR products were Sanger sequenced. Additionally, the reference pandemic ssDNA bands were analyzed in the same manner.

Sanger sequencing of the ssDNA bands confirmed that fourteen out of the nineteen analyzed samples, were identical with the A(H1N1)pdm09 pandemic strain reference sequence (Table 2). For the five samples with electrophoretic profiles different from the reference strain, Sanger sequencing revealed the presence of many point mutations. Schematic representation of all detected mutations and their localization within analyzed HA amplicone are presented in Fig. 2. Sample 2010-03994 contained two point mutations, 2011-01219 – eight, 2010-01164 – three, 2010-05270 – seven and 2010-05347 – five. Six mutations were present in more than one sample (Fig. 2), and nine were unique to single isolates. It seems unlikely that mutations arose during the short passages of the original virus from swabs in MDCK cells.

DNA codons containing detected point mutations were translated to amino acids and compared with the pandemic reference sequence. Furthermore, their physico-chemical

properties and localization within HA protein structure were also determined. All detected point mutations as well as corresponding changes of amino acids, in comparison to reference sequence, are summarized in Table 2. Two mutations (TAC>AAC, GTA>ATA) close to the vicinity of 3' end of the amplicones, were not analyzed as they were localized within the primer binding region. Five mutations were synonymous substitutions (Table 2), that did not affect the HA protein sequence. In three cases the mutations led to substitutions preserving the physico-chemical properties of the encoded amino acids. However, some of the non-synonymous substitutions affected the physico-chemical properties of amino acids (Table 2). Three point mutations (GTA>GGA, GCC>GAC, GAG>AAG), leading to amino acid changes (valine (V)›glycine (G), alanine (A)›aspartic acid (D), glutamic acid (E)›lysine (K) respectively), were localized in a region presumed to be part of epitope E of H1. To check if these changes could affect HA-Ab binding and modulate the potential immune response against A(H1N1)v, we analyzed this region by molecular modeling and estimated the possible impact on vaccine effectiveness.

Using sequence based structural alignment method we have matched the HA1 structure of H1 hemagglutinin (3ZTN) with the neutralizing antibody directed to H3 hemagglutinin from 1QFU, the 3-D structures were processed and aligned to obtain maximum resemblance of their HA-Ab binding interfaces. According to Figure 3, one of the detected mutations - E66K was localized in an HA-Ab binding interface.

Substitutions in the amino acid sequence, that distinguish a circulating strain from a vaccine strain, can be characterized by antigenic distance measured by values like P-value or $p_{epitope}$ [9] and then used for determination of how these substitutions may influence the vaccine effectiveness level predicted by a previously established mathematical model. We provide such quantification based on five substitutions. All calculations are described in detail in Materials and Methods. The calculated $p_{epitope}$ value based on substituted amino acids located

in epitope E, shows that these substitutions can result in 17.5% decrease of the predicted vaccine efficacy.

The mutated fragments listed in Table 2 were compared with available *HA* sequences of Taiwan pandemic influenza isolates deposited in the NCBI and EpiFlu data banks. The phylogenetic tree constructed for the 935 nt *HA* gene fragment, showed that the isolates may be divided into several clusters (Fig. 4). The isolates described in this work form a distinguishable cluster except isolate 2010-01164 which shares the V47G mutation with a group represented by isolate EPI382342. It should be pointed out that the A48D and E66K mutations identified here were not found in any other group of isolates that were subjected to comparative analysis. They may therefore represent novel changes in HA which appeared in the 2010/2011 flu season and suggest a short-term or geographically limited emergence of influenza A virus strains with a high antibody evading potential

Discussion

Over four years have passed since the emergence of a novel H1N1 influenza A virus (A(H1N1)v) which caused the first global pandemic in the 21st century. The virus is still circulating in many countries, though at a much lower level than in its peak season of 2009-2010.[10] The impact of A(H1N1)v during the upcoming seasons will be directly correlated with the immunity of the population provided the virus does not undergo significant antigenic changes.[11] So, as long as the virus is genetically stable and the percentage of population resistant to infection is high (immunity being attained either by vaccinations or prior exposures to cross-reacting strains), it should not pose a major risk to public health. However, the influenza virus is a master of immune evasion and the possibility of an explosive emergence of new mutants of A(H1N1)v cannot be excluded.

The evasion of the host immune response by an influenza virus can be accomplished by at least three ways. Firstly, by sequential changes in *HA* (and other influenza virus genes) owing to the error-prone copying of the RNA genome by the viral RNA polymerase; this phenomenon is called as antigenic drift. Second, by much more abrupt changes due to antigenic shift during co-infection with different HA subtypes; it can be described as importation by reassortment of a gene encoding a different *HA* subtype concurrently with or without other viral gene segments. The third mode is a special case of the antigenic shift involving intrasubtypic reassortment which may occur during co-infection with two different strains of the same serotype. The HA serotype in the progeny virus which underwent reaasortment remains the same, though the *HA* sequence may differ significantly from the *HA* sequences of the parental viruses.

The intrasubtypic reassorment may be particularly dangerous when two major strains of the same serotype co-circulate in an area, as was the case with the seasonal and pandemic A(H1N1) viruses. We showed previously that among patients in Polish hospitals co-infection

with these two strains was observed at a fairly high frequency.[12] In contrast, in the selected samples diagnosed as pandemic A(H1N1)pdm09 in Taiwan, mixed infections with seasonal and pandemic strains were not detected, but instead some samples exhibited unique, specific electrophoretic profiles, different from the A/Mexico/4486/09 reference pandemic strain. Five out of nineteen samples contained point mutations. Notably, all samples isolated in 2009 were identical to the reference sequence, even though originating from Mexico and the mutations were observed only in samples isolated in 2010 and 2011, suggesting geographically limited virus evolution.

To assess the potential impact of the detected genetic changes on the virus fitness, we analyzed them more closely. We based our conclusions on the epitope mapping done by Deem and Pan[13] in which epitopes A-E of H3 hemagglutinin were aligned with the corresponding amino acids in H1 hemagglutinin. After sequence alignment, similarity of the epitopes was verified by 3-D structure alignment of H1 and H3 hemagglutinins. Therefore we refer to this bioinformatically mapped regions of H1 as epitopes A-E of H1 hemagglutinin. Three amino acid substitutions affecting their physico-chemical properties: (GTA>GGA, GCC>GAC and GAG>AAG), detected in three Taiwanese isolates localized to the HA region presumed to be a part of the epitope E of H1.[13] It is also worthy to point out that residue 66 (E) of H1 hemagglutinin is also a part of nine amino acid sequence (ILGNPECEL) listed among predicted epitopes for A2 supertype of MHC Class I molecule[14] by Influenza Research Database[15]

Hemagglutinin is a homo-trimer with each of the subunits comprising two polypeptides: HA1 and HA2. They fold into an α-helical stem and a globular domain, both of which can interact with the host cell membrane. Also neutralizing antibodies can bind to the both structures, as seen in crystallographic studies e.g, 3ZTN or 1EO8, but primarily they target the globular domain, where the HA receptor binding site is also located.[16,17] Based on

the approach proposed by Deem and Pan[13] we used data for H3 hemagglutinin from the 1QFU complex and matched its structure with that of H1 hemagglutinin (3ZTN) to predict their interaction interface. The HA-Ab interface with the BH151 antibody is formed by 14 residues (50-54, 65-70 and 81-85 of HA1). The E66K mutation (Fig. 3) is within this interface and could affect it as it changes both the charge and size of the amino acid. Single mutations within HA-AB interface, or even those located in its close vicinity, have already been shown to allow viruses to escape neutralization.[18] Mutations of the interface residue D54 to N or to Y make the influenza virus immune to the HC45 and BH151 antibodies.[19,20] This effect may be caused by introducing a new glycosylation site.[21] It is noteworthy that the mutation E66K may create a new glycosylation site and residue 66 is in close contact with D54, forming a hydrogen bond with its main chain in the crystal structure 1EO8.

On the other hand, although mutations at a protein-protein interface may be a strong indication of structural hindrances hampering complex formation, there are also studies showing the ability of antibodies to bind protein variants with one or a few amino acids changed.[20, 22] The close proximity to a known escape mutation residue, exposure and interface location suggest the major role of residue 66 in modulating HA-Ab affinity. Nevertheless, this hypothesis needs further studies.

It is also worth mentioning that the observed substitutions present in epitope E of H1 may affect vaccine effectiveness when they become widespread. The vaccine effectiveness undergoes annual changes, partially due to antigenic distances between the recommended vaccine strain and the circulating strain.[13] One of factors to quantify such antigenic distance is p_{eptope} value which describes the highest fraction of substituted amino acids in all epitopes.[23, 24] Recent studies showed that vaccine effectiveness correlates with the p_{epitope} value, thus one can quantify the impact of substituted amino acids on the predicted level of vaccine effectiveness for influenza-like illnesses, which is 52.7%.[24] According to our p_{epitope} and vaccine

effectiveness calculations , the five substitutions found in epitope E of HA can decrease the predicted vaccine effectiveness from 52.7% to 35.2%. Obviously, there are many other factors that can alter the vaccine effectiveness. Although the presented calculations are only based on available epidemiological data, conclusions derived from such mathematical models may be useful in future understanding of influenza virus diversity.[24]

Materials and Methods

Sample collection and virus propagation

Clinical specimens (throat or nasal swabs) from outpatients (southern, eastern, northern and central Taiwan communities) with influenza-like illnesses and from hospitalized patients who developed severe complications, were collected and transported to the laboratories of the influenza surveillance network in Taiwan, which is coordinated by the Centers for Disease Control (Taiwan CDC), for influenza diagnosis using virus culture or/and real-time RT-PCR. All influenza isolates from positive cases were transported to the Taiwan CDC where they were shortly passaged in MDCK cells at 34°C in serum free DMEM medium (Life Technologies, cat no. 21855-025) with TPCK-trypsin (Thermo Scientific, cat no. 20-233) and further characterized by analyzing the viral antigenicity and sequences of HA genes. Sample information with clinical features of A(H1N1)pdm09 infection is summarized in Table 1.

Total RNA extraction and cDNA synthesis

Viral RNA was extracted from cultured viral isolates using QIAamp Viral RNA Mini Kits (Qiagen, cat. no. 52904). In brief, 140 µl of a clinical specimen was mixed with AVL buffer, followed by 560 µl of ethanol. The resultant suspensions were applied on a QIAamp Mini column for RNA binding. After washing with wash buffers and eluting with TE buffer, the eluted viral RNA was used in the reverse transcription assay. Alternatively, automated extraction was conducted using a MagNa Pure LC extraction system (Roche). For cDNA preparation, BluePrint(TM) 1st Strand cDNA Synthesis Kit (Takara BIO Inc, cat. no. 6110) was used. Template viral RNA and random primers were mixed in a total volume of 10 µl. After incubating at 65°C for 5 minutes followed by cooling on ice, reaction buffer, reverse transcriptase and water were added. The resultant mixture was incubated at 30°C for 10

minutes, then at 42°C for 30 minutes. The obtained cDNA was used in PCR assays. To determine the viral genotype, partial nucleotide sequences (position 393-1182) of the HA genes were first analyzed by RT-PCR using primers published by WHO. Genotypes of the isolated viruses were then determined by blasting the sequences through the NCBI database. All of the A(H1N1)pdm09 isolates in this study were identified as A/California/07/2009-like viruses.

Hemagglutinin gene fragment amplification by PCR

The primers specific for a fragment of HA gene of both pandemic A(H1N1)v and seasonal A(H1N1)v strain were as follows: H1msscp1 (5'-AGTAACACACTCTGT-3') and H1msscp2 (5'-ACAATGTAGGACCATGA-3'). The primers were synthesized by IBB PAN (Warsaw, Poland). PCR was performed in a 25 µl reaction volume with standard reagents, 0.4 µM each primer and 1 µl of cDNA solution. The assay was performed in a GeneAmp PCR System 2700 (Applied Biosystems Inc., USA) according to the following procedure: 7 min. at 94°C for initial denaturation, then 45 cycles of denaturation at 94°C for 10 s, annealing at 46°C for 30 s and extension at 68°C for 35 s. After the last cycle the reaction was completed by a final extension at 68°C for 7 min.

MSSCP-based minor genetic variants enrichment procedure

The PCR products were analyzed by the MSSCP method at strictly controlled (± 0.2ºC) gel temperature in a dedicated apparatus: DNA*Pointer*® System/BioVectis (Warsaw, Poland) as described by Kaczanowski et al.[25] Briefly, the PCR products were heat denatured and ssDNA conformers were resolved on a 9% polyacrylamide gel in native conditions (TBE buffer), the gel temperature decreasing during the run from the initial 15°C to 10°C to 5°C. Electrophoretic profiles of analyzed amplicones were compared to reference samples

representing seasonal: A/Brisbane/59/2007 and pandemic: A/Mexico/4486/09 A(H1N1) strains. Subsequently, DNA bands were visualized by silver nitrate staining (SilverStain DNA Kit, BioVectis, cat. no. 200-101). Fragments of the MSSCP gel containing bands of interest (for samples with electrophoretic profile in accordance with pandemic reference sample: the first ssDNA band from the bottom of the gel; for samples with electrophoretic profile distinct from pandemic reference sample: the first ssDNA band from the top of the gel) were cut out, ssDNA was eluted and re-amplified using primers and PCR conditions as described above. For subsequent DNA Sanger sequencing (3730xl DNA Analyzer, Applied Biosystems, Carlsbad, CA, USA) 1/10 vol. of obtained PCR products, purified with exonuclease I (Exo) and shrimp alkaline phosphatase (Sap) enzymes (Fermentas, cat. no. EN0581 and EF0511) was used.[26]

Bioinformatic analysis of HA structures

HA structures were downloaded from the RCSB website http://www.rcsb.org (PDB accession codes: 3ZTN and 1QFU) and subjected to a short 200 step minimization procedure with the backbone atoms fixed. Minimization was done with Tripos SybylX and AMBER force field to relax crystalographic constrains imposed especially on the surface exposed amino acid side chains. Next, structures where superimposed using the Matchmaker software[27], by first creating a pairwise sequence alignment with the Needleman and Wunsch algorithm and BLOSUM-62 matrix, and then fitting the aligned residue pairs. RMSD between 165 aligned atom pairs was 1.111 A, yielding a highly matched structural alignment. The aligned crystallographic structures were analyzed with the emphasis on the interface between HA protein and neutralizing antibody chains: H and L chains from 1QFU. Volume analysis was conducted with VolSurf as implemented in Tripos SybylX (Tripos Inc., St. Louis, MO, USA) packages. UCSF Chimera was used for hydrogen bond network analysis, contact/clash

estimation and electrostatic surface generation.

Calculation of $p_{epitope}$ value and vaccine effectiveness

All calculations were based on data shown in Pan *et al.*[24] The *P*-value which is used to measure antigenic distance between circulating and vaccine strain[21] was calculated for every epitope of H1 (A-E) using equation 8 in reference [24]:

P-value= Number of substitutions in the epitope / Number of amino acids in the epitope.

The calculation was based on the assumption that the only difference between the hypothetical strain and vaccine strain are the five amino acid substitutions present in epitope E of H1. The $P_{epitope}$ value is the highest of all *P*-values[24]. The P-value for epitope E was taken as the $p_{epitope}$ value for further calculations. As shown in chart on Figure 2 in reference [24] based on epidemiological data for various circulating and vaccine strains from 1982 to 2008 taken from Table II in reference [24], vaccine effectiveness is a decreasing linear function (R^2=0.68) of $p_{epitope}$ described as:

vaccine effectiveness = $-1.19p_{epitope} + 0.53$.

The predicted value of vaccine effectiveness calculated by Pan *et al*[24] was 52.7%, for $p_{epitope}$ = 0.

Phylogenetic analysis of A(H1N1)pdm09 isolates

Two hundred and twenty one Taiwan origin A(H1N1) hemagglutinin nucleotide sequences deposited from 2009 through 2011 were downloaded from EpiFlu database. The sequences were aligned in Geneious Alignment Tool and duplicated sequences were removed. The 12 unique sequences were then aligned with the five sequences from our study. After extraction, merged regions comprising nucleotides 183 - 291 and 364-1189 were used to create a

1phylogenetic tree based on Neighbor-joining method with Tamura Nei genetic distance model

2using Geneious Tree Builder.

Acknowledgements

We acknowledge the authors and laboratories originating and submitting the sequences from GISAID's EpiFlu™ Database, on which this research is based. The list is detailed in Table 3.

KL and BS were supported by a PBS grant of NCBiR (National Center for Research and Development)

References

1. Dawood FS, Jain S, Finelli L, Shaw MN, Lindstrom S, Garten RJ, et al. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Eng J Med 2009; 360:2605-15.

2. Naffakh N, van der Werf S. April 2009: an outbreak of swine-origin influenza A (H1N1) virus with evidence for human-to-human transmission. Microbes Infect 2009; 11:725 – 28.

3. Smith GJD, Vijaykrish D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 2009; 459:1122-26.

4. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, et al. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 2009; 325:197-201.

5. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. Nature 2009; 459:931-39.

6. Shinde V, Bridges CB, Uyeki TM, Shu B, Balish A, Xu X, et al. Triple- reassortant swine influenza A(H1) in humans in the United States, 2005-2009. N Eng J Med 2009; 360:2616-25

7. Skehel JJ, Wiley DC. Receptor binding and membrane fusion: the influenza hemagglutinin. Annu Rev Biochem 2009; 69:531-64.

18. Barr IG, Cui L, Komadina N, Lee RT, Lin RT, Deng Y, et al. A new pandemic influenza 2A(H1N1) genetic variant predominated in the winter 2009 influenza season in Australia, New 3Zealand and Singapore. Euro Surveill 2010; 15:(42):pii=19692.

4

59. Gupta V, Earl DJ, Deem MW. Quantifying influenza vaccine efficacy and 6antigenic distance. Vaccine 2006; 24:3881-3888.

7

810. http://www.nimr.mrc.ac.uk/who-influenza-centre/annual-and-interim-reports/

9

1011. Morens DM, Taubenberger JK, Fauci AS. The 2009 pandemic influenza virus: what next? 11mBio 2010; 1(4):e00211-10.

12

1312. Pajak B, Stefanska I, Łepek K, Donevski S, Romanowska M, Szeliga M, et al. Rapid 14differentiation of mixed influenza A/H1N1/v virus infections with seasonal and pandemic 15variants by multitemperature single-stranded conformational polymorphism analysis. J Clin 16Microbiol 2011; 49(6):2216–21.

17

1813. Deem MW, Pan K. The epitope regions of H1-subtype influenza A with application to 19vaccine efficacy. Protein Eng Des Sel 2009; 22:543-546.

20

2114. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and 22updated classification. BMC Immunology 2008; 9:1-15.

23

115. Squires RB, Noronha J, Hunt V, Garcia-Sastre A, Macken C, Baumgarth N et al. Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. Influenza Other Respir Viruses 2012; 6(6):404-16.

516. Corti D, Voss J, Gamblin SJ, Codoni G, Macagno A, Jarrossay D, et al. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinin. Science 2011; 12:850-6.

917. Fluery D, Daniels RS, Skehel JJ, Knossow M, Bizebard T. Structural evidence for recognition of a single epitope by two distinct antibodies. Proteins 2000; 1:572-8.

1218. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinformatics 2006; 7:339.

1519. Gigant B, Fluery D, Bizebard T, Skehel JJ, Knossow M. Crystallization and preliminary X-ray diffraction studies of complexes between an influenza hemagglutinin and Fab fragments of two different monoclonal antibodies. Proteins 1995; 23:115-7.

1920. Knossow M, Skehel JJ. Variation and infectivity in neutralization of influenza, Immunology 2006; 119:1-7.

2221. Fleury D, Barrère B, Bizebard T, Daniels RS, Skehel JJ, Knossow M. A complex of influenza hemagglutinin with a neutralizing antibody that binds outside the virus receptor binding site. Nat. Struct. Biol 1999; 6(6):530-4.

122. Fluery D, Wharton SA, Skehel JJ, Knossow M, Bizebard T. Antigen distortion allows
2influenza virus to escape neutralization. Nat Struct Biol 1998; 5:119-23.
3
423. Zhou H, Pophale R, Deem MW. Chapter 10: Computer-associated
5vaccine design. Influenza: Molecular Virology, Caister Academic Press
62009; 173-191.
7
824. Pan K, Subieta KC, Deem MW. A novel sequence-based antigenic
9distance measure for H1N1, with application to vaccine effectiveness and
10selection of vaccine strains. Protein Eng Des Sel 2011; 24:291-299.
11
1225. Kaczanowski R, Trzeciak L, Kucharczyk K. Multitemperature single-strand conformation
13polymorphism. Electrophoresis 2010; 22:3539-45.
14
1526. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain terminating inhibitors.
16Proc Natl Acad Sci USA 1997; 74:5463-67.
17
1827. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-
19structure analysis with UCSF Chimera. Bioinformatics. 2006; 12(7):339.
20

Figure Legends

Figure 1

New genetic variants among A(H1N1)pdm09 isolates collected at Taiwan between 2009-2011 detected by MSSCP genotyping.

RT-PCR products of hemagglutinin gene obtained from pandemic Taiwan A(H1N1)pdm09 virus isolates, as well as reference seasonal (s) and pandemic (p) strains of influenza virus A(H1N1)pdm09 were denatured and ssDNA's were separated on a 9 % polyacrylamide gel using MSSCP method under optimum electrophoretic conditions. DNA bands were visualized with silver stain. Strains are indicated as follows: s – reference seasonal strain, p – reference pandemic strains. Taiwan isolates are described with symbols listed in Table 1. Note the presence of five distinct MSSCP electrophoretic profiles (arrows at the bottom of the figure) (samples number: 2010-03994, 2011-01219, 2010-01164, 2010-05270, 2010-05347) among samples, which based on RT-PCR assay, were classified as A(H1N1)pdm09 strain. Dividing lines indicate grouping of images from different parts of the same gel.

Figure 2

Schematic representation of genetic diversity of hemagglutinin (HA) sequence in five Taiwan isolates of A(H1N1)v pan09 strain.

Black arrows above and below A(H1N1)v pdm09 reference sequence indicate modified DNA codons. Red letters show nucleotide changes. Altered amino acids in HA protein sequence are also shown. Blue star over the description indicates amino acids localized in the epitope E of HA. Additionally, red arrows underline three point mutations, which are translated to amino acids in the epitope E of HA.

Figure 3

Structure of HA-Ab complex.

Hemagglutinin protein monomer (blue color) interacting which neutralizing antibody (red color). Amino acid 66, localized on the HA-Ab interface is marked in green.

Figure 4

Phylogenetic tree of nucleotide sequences of hemagglutinin gene of Taiwan isolates of influenza A(H1N1)pdm09 virus.

All sequences included in phylogenetic analysis were 935 nucleotides long. All Taiwan isolates deposited from 2009 through 2011 were retrieved from the EpiFlu Database. Each isolate (EPI210250 HA, EPI382332 HA, EPI382342 HA, EPI382415 HA, EPI190807 HA, EPI382397 HA, EPI382345 HA, EPI382352 HA, EPI382407 HA, EPI382412 HA, EPI382365 HA, EPI382375 HA) represents a number of identical sequences used in alignment. Isolates studied in this paper are described in Table 2. Numbers on branches indicate number of nucleotide substitutions per site.

1**Table 1**

2Sample information and clinical symptoms of flu infection among A(H1N1)pdm09 Taiwan

3patients.

4

Sample number	Age	Sex	Date of isolation	Geographical location of isolation	Clinical symptoms
2009-02826	5	Male	2009-07-27	Central Taiwan	cough, fever (>38°C), pneumonia
2009-00940	70	Male	2009-09-10	Northern Taiwan	cough, fever (>38°C)
2009-08542	45	Male	2009-09-28	Eastern Taiwan	cough, sore throat, dyspnea, pneumonia
2010-03994	4	Male	2010-08-11	Southern Taiwan	cough, fever (>38°C), dyspnea, pneumonia, respiratory failure
2010-00842	63	Male	2010-12-20	Northern Taiwna	dyspnea, sore throat, fever, pneumonia
2010-06031	42	Female	2010-08-18	Northern Taiwan	unavailable
2011-02054	60	Male	2011-01-20	Central Taiwan	myalgia, cough, dyspnea, pneumonia, sore throat, fever (>38°C)
2011-00623	56	Female	2011-02-09	Northern Taiwan	cough, fever (>38°C), dyspnea
2011-01219	53	Male	2011-02-10	Northern Taiwan	cough, fever (>38°C), pneumonia
2009-06078	10	Male	2009-08-27	Central Taiwan	cough, fever (>38°C), pneumonia
2009-04909	58	Female	2009-09-28	Eastern Taiwan	myalgia, dyspnea, cough, sore throat
2009-00937	83	Male	2009-09-09	Northern Taiwan	dyspnea, cough, fever (>38°C), pneumonia
2009-08575	22	Female	2009-09-20	Eastern Taiwan	myalgia, dyspnea, cough, sore throat, pneumonia
2010-01164	15	Female	2010-05-05	Northern Taiwan	unavailable
2010-05270	37	Male	2010-09-06	Northern Taiwan	cough, sore throat, fever (>38°C), dyspnea, pneumonia
2010-05347	56	Male	2010-08-12	Northern Taiwan	cough, fever (>38°C), pneumonia
2011-04512	38	Female	2011-01-17	Eastern Taiwan	myalgia, dyspnea, cough, sore throat, fever (>38°C)

2011-02068	4	Male	2011-01-24	Central Taiwan	cough, fever (>38°C), vomiting
2011-04611	55	Male	2011-02-08	Northern Taiwan	dyspnea, cough

1

Table 2

Genetic diversity of HA gene fragment in Taiwan A(H1N1)pdm09 isolates. Red indicates mutations localized in epitope E of HA protein. Blue indicates amino acids substitutions preserving physico-chemical properties. Bold black font indicates synonymous amino acid substitutions.

Sample number	Mutations in nucleotide sequence (ref›mut)	Changes in amino acid sequence (ref›mut) with their position number	Amino acid physico-chemical properties	Comments
2009-02826 2009-00940 2009-08542 2010-00842 2010-06031 2011-02054 2011-00623 2009-06078 2009-04909 2009-00937 2009-08575 2011-04512 2011-02068 2011-04611	-	-	-	Identical to A(H1N1)pdm09 reference sequence
2010-03994	TAC›AAC	Y›N (78)	hydrophobic› polar uncharged	-
	GTA›ATA	V›I (80)	hydrophobic› hydrophobic	-
2011-01219	**GCT›GCA**	**A›A (58)**	-	**Localized in epitope E of HA**
	GGC›GAC	G›D (59)	aliphatic uncharged› charged negative	-
	TGG›TGT	C›W (60)	sulhur-containing› hydrophobic	-
	ATC›ATA	I›I (61)	-	-
	CTG›ATG	L›M (62)	hydrophobic› hydrophobic	-
	CTC›ATC	L›I (70)	hydrophobic› hydrophobic	**Localized in epitope E and HA-Ab interface**
	AGC›AAC	S›N (74)	polar uncharged› polar uncharged	**Localized in epitope E of HA**
	TAC›AAC	Y›N (78)	hydrophobic› polar uncharged	-
2010-01164	AGA›AGG **GTA›GGA**	R›R (45) **V›G (47)**	- **hydrophobic› aliphatic uncharged**	- **Localized in epitope E of HA**

	TAC›AAC	Y›N (78)	hydrophobic› polar uncharged	-
2010-05270	**GCC›GAC**	**A›D (48)**	**hydrophobic› charged negative**	**Localized in epitope E of HA**
	GCT›GCA	**A›A (58)**	-	**Localized in epitope E of HA**
	GGC›GAC	G›D (59)	aliphatic uncharged› charged negative	-
	GAG›AAG	**E›K (66)**	**charged negative› charged positive**	**Localized in epitope E and HA-Ab interface**
	AGC›AAC	S›N (74)	polar uncharged› polar uncharged	**Localized in epitope E of HA**
	TAC›AAC	Y›N (78)	hydrophobic› polar uncharged	-
	GTA›ATA	V›I (80)	hydrophobic› hydrophobic	-
2010-05347	**ATT›ATC**	**I›I (57)**	-	**Localized in epitope E of HA**
	CCA›CCT	**P›P (65)**	-	**Localized in HA-Ab interface**
	GAG›AAG	**E›K (66)**	**charged negative› charged positive**	**Localized in epitope E and HA-Ab interface**
	TAC›AAC	Y›N (78)	hydrophobic› polar uncharged	-
	GTA›ATA	V›I (80)	hydrophobic› hydrophobic	-

1

1We acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiFlu™ Database on which this research is based. The list is detailed below.

3All submitters of data may be contacted directly via the GISAID website www.gisaid.org

4Segment ID	Segment	Country	Collection date	Isolate name	Originating Lab	Submitting Lab	Authors
5							
6EPI382375	HA	Taiwan	2010-Dec-01	A/Taiwan/66259/2010		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
7Chang,S.-C.; Lin,T.-Y."							
8EPI382365	HA	Taiwan	2010-Nov-01	A/Taiwan/3276/2010		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
9Chang,S.-C.; Lin,T.-Y."							
10EPI382412	HA	Taiwan	2011-Feb-01	A/Taiwan/90284/2011		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
11Chang,S.-C.; Lin,T.-Y."							
12EPI382407	HA	Taiwan	2011-Feb-01	A/Taiwan/90252/2011		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
13Chang,S.-C.; Lin,T.-Y."							
14EPI382352	HA	Taiwan	2010-Sep-01	A/Taiwan/90149/2010		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
15Chang,S.-C.; Lin,T.-Y."							
16EPI382345	HA	Taiwan	2010-Aug-01	A/Taiwan/90112/2010		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
17Chang,S.-C.; Lin,T.-Y."							
18EPI382397	HA	Taiwan	2011-Feb-01	A/Taiwan/90187/2011		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
19Chang,S.-C.; Lin,T.-Y."							
20EPI190807	HA	Taiwan	2009-May-19	A/Taiwan/T0724/2009		Other Database Import	
21EPI38241 5	HA	Taiwan	2011-Feb-01	A/Taiwan/65330/2011		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
22Chang,S.-C.; Lin,T.-Y."							
23EPI382342	HA	Taiwan	2010-May-01	A/Taiwan/90060/2010		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
24Chang,S.-C.; Lin,T.-Y."							
25EPI382332	HA	Taiwan	2010-Jan-01	A/Taiwan/90001/2010		Other Database Import	"Tsao,K.-C.; Shih,S.-R.; Chen,G.-W.; Huang,C.-G.;
26Chang,S.-C.; Lin,T.-Y."							
27EPI210250	HA	Taiwan	2009-Apr-08	A/Taiwan/693/2009	CDC	WHO CCRRI	

28