Peertechz

Research Article

Comparative analysis of Hemagglutinin of 2013 H3N2 Influenza A virus Indicates its Evolution from 1968 H3N2 Pandemic Influenza A virus

Swatantra Kumar[1†], Vimal K Maurya[1†], Sneham Tiwari[2†], Amit K Banerjee[3†], Neelima Arora[3], Sai V Chitti[2], Debadatta Nayak[4], Anil Khurana[4], Raj K Manchanda[4], Srinivasulu Gadugu[5] and Shailendra K Saxena[1,2]*

[1]Department of Centre for Advance Research, Faculty of Medicine, King George's Medical University, Lucknow 226003, India

[2]CSIR-Centre for Cellular and Molecular Biology, Hyderabad 500007, India

[3]CSIR-Indian Institute of Chemical Technology, Hyderabad 500007, India

[4]CCRH, Ministry of Ayush, Janakpuri, New Delhi 110058, India

[5]Department of Medicine, JSPS Government Medical College, Hyderabad 500013, India

[†]These authors contributed equally to this work as first author

Abstract

Emergence of influenza A H3N2 is alarming. Strain 2013 H3N2 has H1 subtype Haemagglutinin (HA) gene segment replaced with a H3 HA gene segment. The vulnerability of the humans to H3 may be directly proportional to the HA alterations. Therefore, we studied this strain and analyzed its sequence and structural divergence and compared it with 2009 H3N1, 2009 H1N1 and its ancestor 1968 H3N2, including 11 other strains from 1918 till 2013. Our analysis showed a maximum sequence similarity between 2013 H3N2 and 1968 H3N2. The amino acid sequence variation of 2013 H3N2 was, 15% with 1968 H3N2, 27% with 2009 H3N1, 60% with 2009 H1N1. Phylogenetic distance of 0.01709 of 2013 H3N2, 0.01118 of 2012 H3N2, 0.01422 of 1968 H3N2, from the origin explained evolution of 2013 H3N2 strain. Glycosylation analysis indicates that influenza 2013 H3N2 has been found similar to 2009 H3N1 and 1968 H3N2 with five similar sites. Antigenic analysis shows that 2013 H3N2 contains different antigenic sites explaining evolution. Comparison of RMSD and hydrogen bonds displayed minimal difference in influenza 2013 H3N2 and 1968 H3N2. Protein disorder regions were found overlapping the antigenic sites in influenza 2013 H3N2 sequences stating the destabilization of antigenic epitopes. Our analyses show the evidence of the evolution of 2013 H3N2 from 1968 H3N2 pandemic, expressing that 2013 H3N2 might be highly virulent strain, tough to be targeted by drugs potently being responsible for pandemics in near future.

Abbreviations

WHO: World Health Organization; HA: Hemagglutinin; NA: Neuraminidase; RMSD: Root Mean Square Deviation; PONDR: Predictor of Naturally Disordered Region; BLAST: Basic Local Alignment Search Tool.

Introduction

Influenza virus is a significant human pathogen, which is evolving with a rapid rate in human race. Influenza A viruses are responsible for sporadic pandemics that cause higher mortality rates than seasonal influenza epidemics. The most

severe pandemic occurred in 1918 [1], followed by pandemics in 1957, 1968, 2009 and 2012, which infected millions of lives [2]. 2009 H1N1 (H1N1pdm09) pandemic followed with the emergence of a strain in which the H1 subtype haemagglutinin (HA) gene segment was replaced with a swine-derived H3 HA gene segment. 2012 H3N2 has been reported to prevailed as 2013 H3N2. The emergence of influenza A viruses is alarming. Globally influenza has infected quite variably. In Asia, 2009 H1N1 viruses are found to be the predominant subtype which co-circulates with H3N2 strains, whereas in many other Asian countries, H3N2 viruses are the predominant subtype, co-circulated with 2009 H1N1. 2009 H1N1 and H3N2 viruses are detected in Caribbean, Central and South America, whereas H3N2 is predominant in North America and 2009 H1N1 and H3N2 are found co-circulating in Africa [3]. H3N2 strains were reported to be predominant during 2011-2012 in northern Italy [4].

According to WHO- Global Influenza Surveillance and Response System (GISRS), H3N2 is expected to be a causal factor of pandemics in the coming years, globally (Figure 1). Hence our study focuses on this high risk viral strain. Based on FluNet reports (on 9 March 2013) from 83 laboratories from different countries between 17 to 30 March, about 51751 specimens were tested by WHO GISRS laboratories, out of them 10548 were positive for influenza viruses, of which 52.4% were typed as influenza A and 47.6% as influenza B. Out of the total influenza A viruses, 56.1% were 2009 H1N1 [influenza A(H1N1)

pdm09] and 43.9% were H3N2 [3]. Therefore, it becomes an urgency to study the virulence, antigenicity and glycosylation of emerging H3N2 and H1N1 strains. Comparison with the previous three influenza seasons shows that the activity of novel H3N2 strain was high among people aged ≥65 years reporting excess mortality in this age group [5]. H3N2 virus has been thought as the dominant subtype till date [6] which will be a potent candidate in future too.

Understanding the ecological and evolutionary drivers of viral diseases from data on disease incidence as well as viral genetic and antigenic variation plays an important role in tracking influenza infections [7]. Influenza A virus evolution in human beings is driven by emerging mutations allowing the virus to escape antibody neutralization [8]. H3N2 infection is highly variable and this variability directs to poor vaccine effectiveness as the employed vaccine strain may not be able to match well with the circulating virus. Antigenic and molecular characterization studies of H3N2 viruses highlight that number, nature and location of signature HA mutations that play an important role in antigenic drift of the virus [9]. Extreme importance of selection and interference effects in influenza evolution pathway has also been reported [10]; also evolutionary and ecological processes interact in a way that they inflict various quantitative restrictions and modifications in the phylogenetic trajectories of H3N2 viruses [7]. Also, H5N1 influenza viruses pose a pandemic threat. Even five amino acid substitutions may be enough for mammal-to-mammal

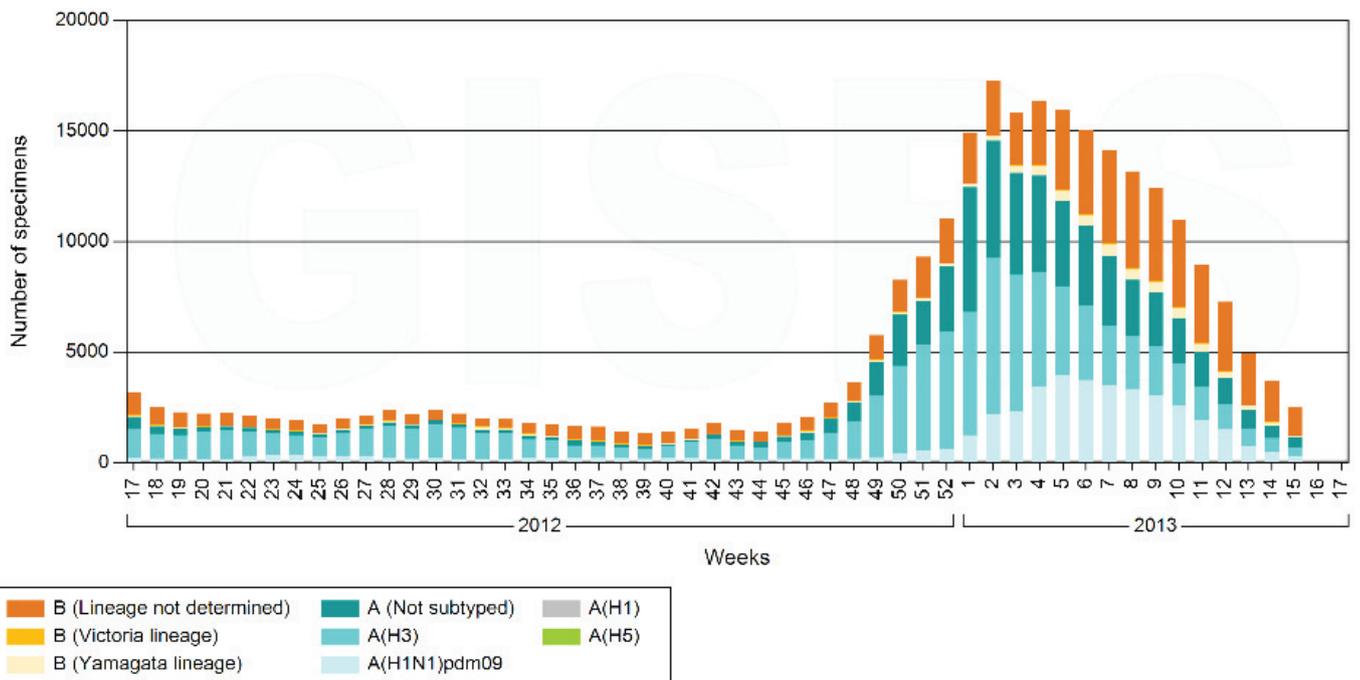## Global circulation of influenza viruses

Figure 1: Report of global circulation of influenza viruses. World Health Organization- Global Influenza Surveillance and Response System (GISRS) report of global circulation of influenza viruses in last one year (Source WHO, influenza virus activity in the world (Data source: FluNet www.who.int/flunet, GISRS, World Health Organization 2020).

002

transmission by respiratory droplets. H5N1 viruses carry two of these substitutions which might cause respiratory droplets transmission. Influenza A virus is highly infectious respiratory pathogen attacking natural hosts, like birds, lower mammals, and humans. It is an enveloped RNA virus containing eight segments of negative-sense RNA, encoding 10 proteins; Hemagglutinin (HA), Neuraminidase (NA), Matrix 1 (M1), M2, Nucleoprotein (NP), Non-Structural proteins (NS1 and NS2), and a Polymerase complex (PA, PB1, and PB2). NP is important as it has many important functions in virus life cycle, and it contains regions that are highly conserved [11]. The two important envelope proteins NA and HA are involved in virulence, having discrete functions in entry and release of progeny viruses during replication cycle [12,13]. HA, one of the major glycoprotein on the surface of influenza virus is of primary importance in the epidemiology of the influenza virus. It is the primary antigen responsible for viral binding to host receptors, enabling entry into the host cell through endocytosis and subsequent membrane fusion, and so it determines the virulence, host ranges, cell tropism and transmissibility of influenza viruses [14,15]. Accumulation of point mutations in the HA proteins leads to antigenic drift. Drift variants emerge due to a positive selection of spontaneous mutants that can escape from existing host's antibodies. The HA protein is the major antigenic component of the virus [16]. HA is a 200-kDa homotrimer made up of globular head and a stalk region. This homotrimer has an essential role in the viral binding and uptake into the host, that is why HA poses to be the main target of the host neutralizing antibodies [17]. Glycosylation sites in the stalk region [18] are mandatory for conformation of the HA molecule. Various influenza A viral strains, which are seasonally isolated from successive strains differ mainly in the sequences of amino acids of HA [19]. Serological studies propose that the infection is carried by vectors like swine, avians, etc. which are loaded with potent virus which can transmit between swine and other mammalian hosts [20]. Studies demonstrate that even turkey hens may also be affected by influenza strains which suggest that turkeys can also serve as useful animal models for studying host tropism and influenza A pathogenesis [21].

Global pandemic of flu 2009 was caused by 2009 H1N1, leading to over 318,925 cases and about 3,917 deaths globally. 2009 H1N1 was reported to be a triple reassortment product of North American swine, avian influenza, human influenza and classical swine influenza virus [22,23]. H3N2 strains have been reported to cause infections, and mostly children and middle-aged adults are found to be on major risk from this variant H3N2 strain [24]. The conditions in Asian countries are set for emergence of novel influenza strains, as they provide more chance to the virus to mutate. Asian tropical countries have short winter spans and high temperature, moisture and rainfall, which are all supporting conditions for seasonal influenza epidemics. This creates a sense of urgency to develop drugs which can target all approaching novel seasonal influenza strains [25]. Local outbreaks of H3N2 are reported in Pakistan, China, Hong Kong, Japan etc. Triple-reassortant swine influenza viruses comprise of internal genes derived from swine (matrix, non-structural and nucleoprotein), human [PB1] and avian (polymerase acidic and PB2) forming a triple-reassortant internal gene (TRIG) cassette, whereas external genes HA and NA mutate extensively creating novel strains with different HA and NA combinations [26]. Upon infection the virus takes over the host cell machinery by virus-driven hijacking of host proteins. There is a novel mechanism reported by which influenza virus attacks the host cells by interaction between viral non-structural protein 1 (NS1) which has histone mimic sequence, and human PAF1 transcription elongation complex (hPAF1C) in infected host cells. This histone mimic has the ability to help influenza in inducing gene expression selectively for antiviral response suppression [27].

In 2012 we faced the H3N2 infection (which prevailed in 2013 too) which poses as a very potent strain capable of causing future virus attacks too. Hence to gain insight into the possible origin virus outbreak in near future, we performed sequential and structural comparative analysis of hemagglutinin of influenza viral strains from 1918 to 2013 including past pandemics of 1977, 1968 and 1918 including H1NI, H3NI, H3N2 and new evolving H5N1. Therefore, in this paper, we have analyzed HA of influenza A/Ontario/001/2013(H3N2) and compared it with influenza A/Georgia/3058/2012(H3N2) which emerged in 2012 (prevailing in 2013 also) influenza A/California/07/2009(H1N1) (H1N1pdm09), influenza A/Perth/16-RGcH5-3/2009(H3N1), influenza A/Hong Kong/1/1968(H3N2), influenza A/South/Carolina/1/18(H1N1),influenzaA/NewCaledonia/20/1999(H1N1), influenza A/Kawasaki/173/2001(H1N1), influenza A Fujian/411/2002(H3N2), influenza A/Norway/807/2004(H3N2), influenza A/Vietnam/1203/2004(H5N1), influenza A/Wisconsin/67/2005(H3N2), influenza A/Indonesia/5/2005(H5N1),influenzaA/Brisbane/10/2007(H3N2), Influenza A/Brisbane/59/2007(H1N1), influenza A/California/7/2004(H3N2), influenza A/Qingdao/2157/2009 (H3N2) strains for sequence variation as well as structural and antigenic divergence for a better understanding of viral pathogenesis and antigenicity, so that it might help in the appropriate drug designing.

Materials and methods

Sequences retrieval

For comparison between circulating H1 and H3 strains, we downloaded HA sequences of various strains from 1918 till reported 2013 strains from nucleotide database from National Center for Biotechnology Information (Table 1) (http://www.ncbi.nlm.nih.gov/nuccore). We included the sequences which were derived from human hosts and were important in causing pandemics in their era. Each sequence involves represents the viral strain of that particular year in which influenza caused pandemics. Sequences taken for this study involve influenza A/Ontario/001/2013(H3N2) (AccessionNo.AGD98956),influenza A/Georgia/3058/2012(H3N2) (Accession No. CY130194.1), influenza A/California/07/2009(H1N1) (H1N1pdm09) (Accession No. AFM72832.1), influenza A/Perth/16-RGcH5-3/2009(H3N1) (Accession No. CY110923.2), influenza A/Hong Kong/1/1968(H3N2) (Accession No. CY112249.1), influenza A/South/Carolina/1/1918(H1N1) (Accession No. AF117241.1), influenza A/New Caledonia/20/1999(H1N1) (Accession No. ABW80979.1), influenza A/Kawasaki/173/2001(H1N1)

Table 1: NCBI nomenclature and properties of all 16 strains.

Strains	Short forms	Accession no	% Helix	% Strands	% Turns
A/South Carolina/1/1918(H1N1)	1918 HIN1	AF117241.1	6.1	46.3	96.3
A/Hong Kong/1/1968(H3N2)	1968 H3N2	CY112249.1	5.6	36.1	85.8
A/New Caledonia/20/1999(H1N1)	1999 H1N1	ABW80979.1	5.7	43.7	82.8
A/Kawasaki/173/2001(H1N1)	2001 H1N1	BAK86315.1	4.4	40.7	75.9
A/Fujian/411/2002(H3N2)	2002 H3N2	AFD64223.1	6.2	47.1	100.4
A/Norway/807/2004(H3N2)	2004 H3N2	ABI22074.1	5.3	33.9	73.1
A/Vietnam/1203/2004(H5N1)	2005 H5N1	EU122404.1	4.9	45.8	102.9
A/Wisconsin/67/2005(H3N2)	2005 H3N2	ABW80978.1	5.5	38.8	78.8
A/Indonesia/5/2005(H5N1)	2005 H5N1	EU146622.1	5.6	39.5	82.4
A/Brisbane/10/2007(H3N2)	2007 H3N2	ABW23353.1	5.8	38.6	92.4
A/Brisbane/59/2007(H1N1)	2007 H1N1	AET50439.1	4.7	51.1	85.6
A/California/07/2009(H1N1)	2009 H1N1	AFM72832.1	5.2	28.8	74.8
A/Perth/16-RGcH5-3/2009(H3N1)	2009 H3N1	CY110923.2	5.8	42.9	85.8
A/Qingdao/2157/2009(H3N2)	2009 H3N2	ADL39170.1	5.5	33.3	84.3
A/Georgia/3058/2012(H3N2)	2012 H3N2	CY130194.1	5.7	41.4	85.1
A/Ontario/001/2013(H3N2)	2013 H3N2	AGD98956	5.9	41.3	92.4

Compilation of all the 16 strains from the ancestor 1918 till the 2013 influenza A. Table displays the NCBI nomenclature, short names, and accession numbers of all the strains. The comparison between morphological properties of protein models like number of helix, strands and turns are also included.

(Accession No. BAK86315.1), influenza A/Fujian/411/2002(H3N2) (Accession No. AFD64223.1), influenza A/Norway/807/2004(H3N2) (Accession No. ABI22074.1), influenza A/Vietnam/1203/2004(H5N1) (Accession No. EU122404.1), influenza A/Wisconsin/67/2005(H3N2) (Accession No. ABW80978.1), influenza A/Indonesia/5/2005(H5N1) (Accession No. EU146622.1), influenza A/Brisbane/10/2007(H3N2) (Accession No. ABW23353.1), influenza A/Brisbane/59/2007(H1N1) (Accession No. AET50439.1), and influenza A/Qingdao/2157/2009 (H3N2) (Accession No. ADL39170.1).

Sequence divergence and phylogenetic analysis

Virus strains were selected for further analysis based on their Hemagglutinin (HA) gene. All HA sequences of the included viral strains were aligned using an interactive multiple sequence alignment tool Clustal W (http://www.ebi.ac.uk/Tools/clustalw2/index.html) [28]. Based on amino acid composition all the sequences were involved for analysis. We analyzed the sequence divergence of 1918 HIN1, 1968 H3N2, 1999 H1N1, 2001 H1N1, 2002 H3N2, 2004 H3N2, 2005 H5N1, 2005 H3N2, 2005 H5N1, 2007 H3N2, 2007 H1N1, 2009 H1N1, 2009 H3N1, 2009 H3N2, 2012 H3N2 and 2013 H3N2 strains by aligning their HA amino acid sequences by using ClustalW. Mega multiple Sequence alignment software was also used for the alignment. MEGA is an integrated tool for conducting sequence alignment. The alignment shows comparison of all eight genome segment alignment datasets (PB2, PB1, PA, HA, NP, NA, M, and NS) but we involved comparison of HA only as it is the most important gene which has a crucial role in influenza evolution.

Trees give important relationship graphs which are used to portray composite interrelationships. The topology and the lengths of the various branches of the phylogram provide qualitative and the quantitative values, respectively which are important factors in learning about the various strains involved in the study. For phylogenetic analysis all the strains were aligned together by ClustalW guide tree software which uses the neighbor-joining algorithm to construct trees from the distance matrix. Phylogeny uses an alignment directly entered into the input box in a supported Alignment format Clustal. The tree file (dnd format) obtained from the ClustalW software was visualized with FigTree v1.4.0, which was used to visualize the phylogenetic tree obtained from ClustalW phylogeny. The phylogenetic distance of viral strains from the origin were visualised by Fig tree v1.4.0 software (by Andrew Rambaut, Institute of Evolutionary Biology, University of Edinburgh) (http://tree.bio.ed.ac.uk/software/figtree/) which is a graphical viewer of phylogenetic trees for a better show of the phylogeny. The aligned files the evolutionary history of viral strains and inferring the evolution of strains. The strains appear in the tree according to their evolutionary track.

Differentiation in glycosylation pattern and antigenic vitiations

To determine the variation in the sites of viral attachment to host cells, a comparison of amino acid sequences of HA glycosylation sites of 1918 HIN1, 1968 H3N2, 1999 H1N1, 2001 H1N1, 2002 H3N2, 2004 H3N2, 2005 H5N1, 2005 H3N2, 2005 H5N1, 2007 H3N2, 2007 H1N1, 2009 H1N1, 2009 H3N1, 2009 H3N2, 2012 H3N2 and 2013 H3N2 strains was performed with NetNGlyc 1.0 software (http://www.cbs.dtu.dk/services/NetNGlyc/) [29]. The NetNglyc server predicts N-Glycosylation sites in proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequons. The derived glycosylation sites from all the included strains were tabulated and compared for their glycosylation motifs. The antigenic divergence between influenza A strains from 1918 till 2013 was implemented by the CTL epitope prediction method. The amino acid sequences of HA were evaluated CTLPred software

(http://www.imtech.res.in/raghava/ctlpred/) which uses the consensus approach. CTLPred directly predicts CTL epitopes which are important in subunit vaccine design and can be a potent drug target. The method is based on Artificial Neural network and support vector machine techniques. The methods allow the consensus and combined predictions. The predicted antigenic sites were tabulated and compared for HA from all the strains included in the study. All the sequences were compared for the various antigenic sites which was helpful in further understanding the re-emergence of influenza infections.

Protein disorder

The disorder predictor used in this work is the PONDR® (Predictor of Natural Disordered Regions) (http://www.pondr.com/cgi-bin/PONDR/pondr.cgi) to predict the distorted protein motifs in 1918 HIN1, 1968 H3N2, 1999 H1N1, 2001 H1N1, 2002 H3N2, 2004 H3N2, 2005 H5N1, 2005 H3N2, 2005 H5N1, 2007 H3N2, 2007 H1N1, 2009 H1N1, 2009 H3N1, 2009 H3N2, 2012 H3N2 and 2013 H3N2 strains. [30-32]. It is known that sequence helps in structure determination, so it might determine change in structure as well. Hence this information was used as a principle. A series of neural network predictors (NNPs) were developed that use amino acid sequence data to predict disorder in a given region collectively called as Predictors of Natural Disordered Regions. PONDR® functions from primary sequence data alone. The predictors use neural networks that use sequence information from 21 amino acids. Fractional composition of particular amino acid was calculated which are used as inputs for the predictor. The neural network, then gives a value for the central amino acid and if the value exceeds a threshold of 0.5 the residue is considered disordered. The VL-XT predictor integrates three neural networks: the VL1 predictor and the N- and C- terminal predictors (XT). Output for the VL1 predictor starts and ends 11 amino acids from the termini. The XT predictors give output predictions up to 14 amino acids from their respective ends, by analysing the data together, disordered regions were predicted.

Primary and secondary structural divergence

The alignment of the nucleotide and amino acid sequences of 1918 HIN1,1968 H3N2,1999 H1N1,2001 H1N1, 2002 H3N2, 2004 H3N2, 2005 H5N1, 2005 H3N2, 2005 H5N1, 2007 H3N2, 2007 H1N1, 2009 H1N1, 2009 H3N1, 2009 H3N2, 2012 H3N2 and 2013 H3N2 strains were obtained by ClustalW tool. The sequences were submitted to the job box of ProtParam tool from EXPASY (Expert Protein Analysis System) software to know the percentage of the amino acids present in the sequence. The result of percentage of amino acids was then tabulated. The amino acid sequences were then visualized by using Weblogo Version 2.8.2 (http://weblogo.berkeley.edu/logo.cgi). To determine the secondary structural divergence in HA proteins, the amino acid sequences of HA of all the strains were submitted to protein modelling tool, Modeller (Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints) (http://salilab.org/modeller/) which made the models on the basis of 3SDY as template [33], taken from RCSB Protein Data Bank - RCSB PDB . The modelled structures were visualized by RasMol 2.7.5 and UCSF Chimera.

The properties of the secondary structures of all proteins were compared for the variation in number of hydrogen bonds, helix, strands, turns and RMSD values with the tools within the model viewer tool and the reply log. All the values were tabulated which helped in further analysis of the morphological difference between all the strains.

Results and discussion

Human body gives a response to the unwanted entry of the virus inside it. Majorly HA and NA influenza antigens are targeted by immune system, hence during subtyping the influenza A viruses, antigenic distinctions of HA and NA proteins play a major part. HA and NA proteins. HA has important role in mediating sialic acid receptor binding of the virus on a host cell which initiates the infection. Upon entering the host cell, HA protein modifies them self to escape the host immune response. If HA is targeted, virus can be neutralized through blocking binding of the virus to the host receptors. Here we have analyzed 2012 H3N2 (which still prevailing 2013 H3N2 and has been found 97% similar to 2012 strain) along with 1968 H3N2, 2009 H3N1, 2009 H1N1 and 11 other strains from ancestors to 2013 (Table 1).

Sequence divergence analysis

The most abrupt changes in antigenic specificity occurs through the HA genes. Therefore, sequence of HA of 2013 H3N2 was compared with other influenza strains for predicting its extent of variability with them. Our results exhibited (Figure 2) the overall sequence variation of H3N2 2013 as 3% with H3N2 2013, 15% with 1968 H3N2, 27% with 2009 H3N1, 60% with 2009 H1N1, 59% with 1918 H1N1, 61% with 1999 H1N1, 60 % with 2001 H1N1, 5 % with 2002 H3N2, 4 % with 2004 H3N2, 4% with 2005 H3N2, 60% with 2005 H5N1, 4 % with 2007 H3N2, 60% with 2007 H1N1, 61 % with 2005 H5N1, 3 % with 2009 H3N2. The result clearly suggests that the 2013 H3N2 has been found much more similar to the 1968 H3N2 strain which caused major pandemic and killed millions of people, it was less similar with 2009 H3N1, which was an extremely dangerous strain, taking millions of lives. This strain showed to be also similar to 2012 H3N2, 2002 H3N2, 2004 H3N2, 2005 H3N2, 2007 H3N2, 2009 H3N2. Henceforth 2013 H3N2 which may be a potent strain to cause infections in near future too. It may prove to be highly virulent and can be a causal factor a big pandemic.

Protein BLAST analysis of HA of 2013 H3N2 revealed that the closest relatives of it are A/Georgia/3058/2012(H3N2), A/Sao Paulo/49663/2012(H3N2), A/Texas/JMM_43/2012(H3N2), A/Boston/DOA38/2011(H3N2), A/Ohio/3083/2012(H3N2), A/Singapore/GP3680/2010(H3N2), A/Czech Republic/114/2012(H3N2), A/Australia/46/2009(H3N2), A/Egypt/N11354/2009(H3N2).

This observation suggests that somehow this viral strain has infected different regions of the world like Sao Paulo, Texas, Boston, Ohio, Sao Paulo, Singapore, Czech Republic, Australia, Korea, and Qingdao etc. So it may be well hypothesized that this virus is getting triple reassorted and may be transported to different regions through carriers, like swine or avian vectors, which makes it more virulent and more potent for pandemics.
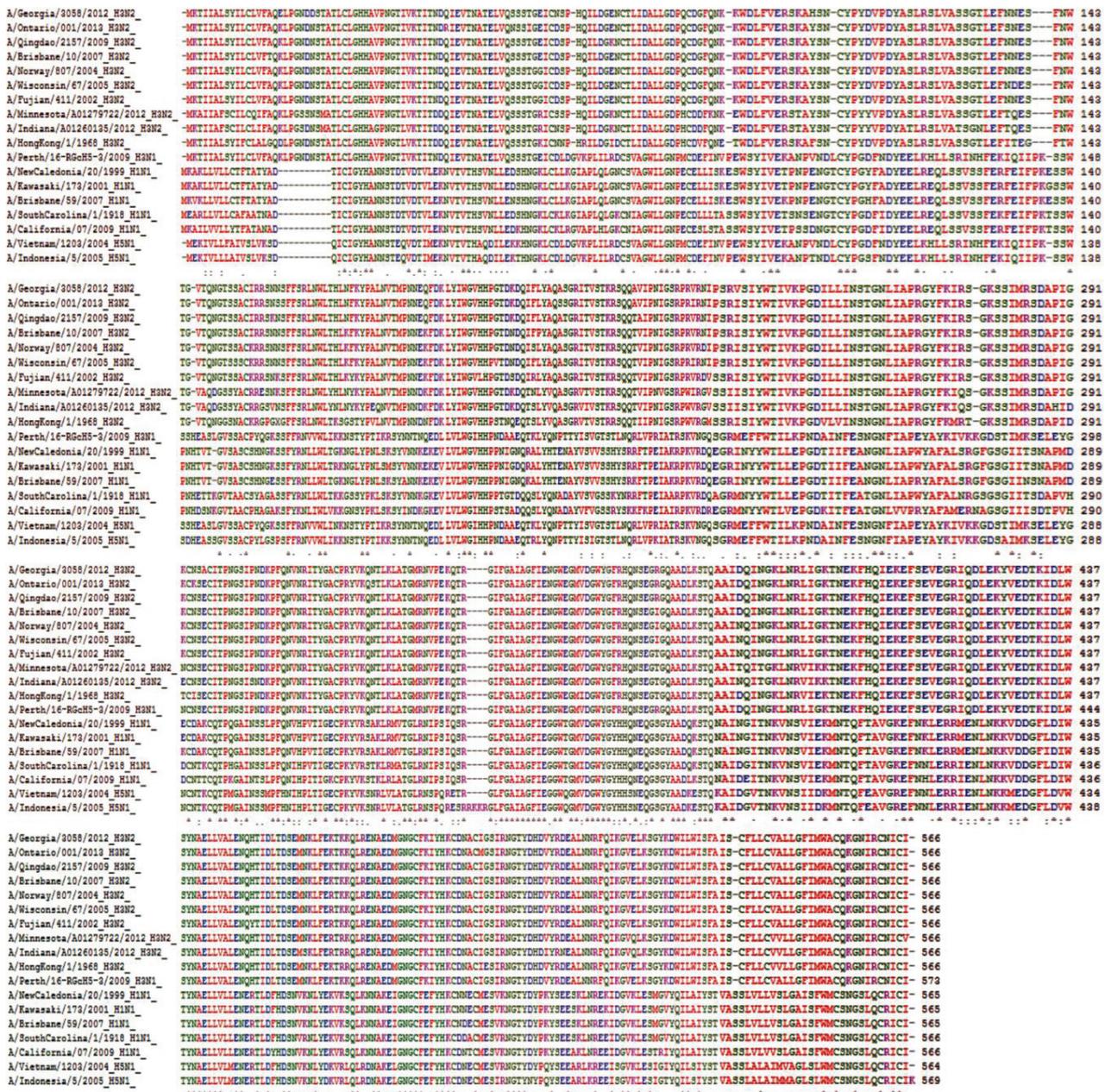
Figure 2: Clustal W alignment of amino acid sequence. Amino acid sequence alignment of the sequences of HA of 2013 H3N2 with 15 other strains included in the study from ancestor till 2013. Clustal W alignment exhibits non-conservative substitutions (" "), conservative substitutions (": ") and semi-conservative substitutions ("."). Conserved regions are represented as ("*").

Phylogenetic analysis

Phylogenetic analysis of 2013 H3N2, along with 2012 H3N2, 2009 H1N1, 2009 H3N1, 1968 H3N2 and 11 other included strains was performed. Two outliers A/Minnesota/A01279722/2012(H3N2) and A/Indiana/A01260135/2012(H3N2) of swine origin were also included. The phylogenetic tree displayed that 2013 H3N2 was falling close to 2012 H3N2 and 1968 H3N2 and was less distant to 2009 H3N1 as compared to 2009 H1N1 which was falling visibly far (Figure 3). 2004 H5N1 and 2005 H5N1 were seen to fall in separate cluster and so were the outliers, showing their difference from the reference 2013

H3N2 strain.

The phylogenetic analysis henceforth makes it very clear that the novel 2013 H3N2 strain being very similar to 1968 H3N2 pandemic strain, might be a very potent strain to cause great pandemics in near future like it's ancestor. Phylogenetic distance of 0.01709 of 2013 H3N2, 0.01118 of 2012 H3N2 0.01422 of 1968 H3N2, from the origin sets them extremely close, pointing towards the similarity of these strains. Whereas phylogenetic distance of 0.05734 of 2009 H3N1 and 0.08416 of 2009 H1N1 from origin makes them a far relative of this novel strain. The whole scenario can be understood as 2013 H3N2 is
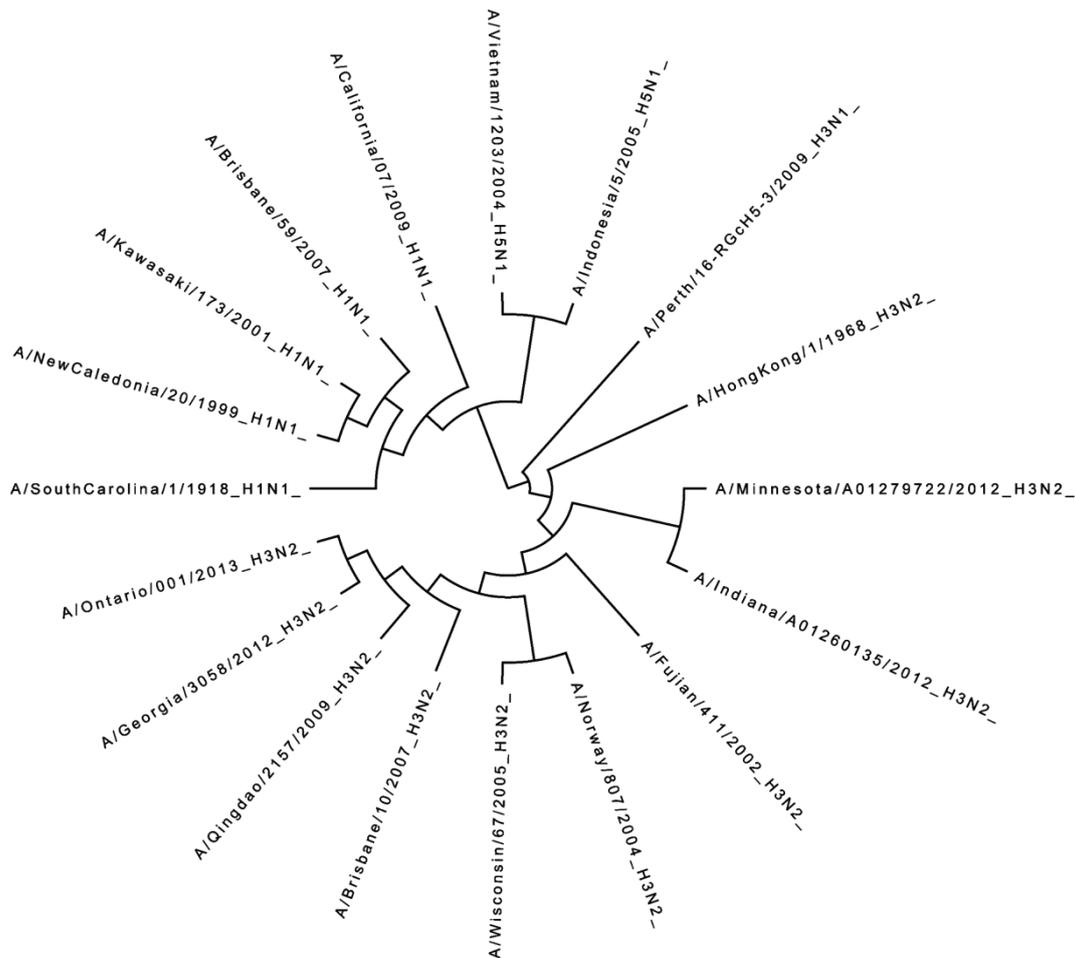
Figure 3: Phylogenetic representation. Phylogram tree showing similarity between 2013 H3N2, 2012 H3N2, 2009 H3N1, 2009 H1N1 and 1968 H3N2 showing the proximity of 2013 H3N2, with 1968 H3N2. 2004 H5N1 and 2005 H5N1 are the outliers involved which help in understanding the close relationship of rest of the strains involved in this study.

most similar to 1968 H3N2, less similar to 2009 H3N1 and least similar to 2009 H1N1. Hence, 2013 H3N2 may stand as a highly virulent candidate for causing pandemics in coming future.

N-Glycosylation sites

Our glycosylation analysis showed that 2013 H3N2 showed extremely different and variant glycosylation sites with one similar site NGTY when compared to 2009 H1N1. Whereas, it displayed five similar glycosylation sites NSTA, NATE, NVTM, NGSI, NGTY when compared with 1968 H3N2 and five similar glycosylation sites NSTA, NGTI, NATE, NGSI, NGTY when compared to 2009 H3N1. It displayed ten similar sites NGTI, NATE, NCTL, NESF, NWTG, NGTS, NVTM, NSTG, NGSI, and NGTY when compared to 2012 H3N2. Five similar glycosylation sites were observed when compared to 2002 H3N2, 2004 H3N2, 2005 H3N2 and 2007 H3N2. The number of glycosylation sites was found to be more in 2013 H3N2 strain with two new glycosylation epitopes NSSI and NSSF which were not found in other 15 strains included in this study.

The results indicate that with reference to glycosylation, 2013

H3N2 is highly similar to 2012 H3N2 but presence of two new glycosylation sites indicate its evolution and severe prevalence with modification which might help the 2013 H3N2 strains to escape host immunity as well as the earlier designed drugs, which may not be able to target these new sites and hence will the help the virus to escape the medications employed to the newly infected patients. However, five similar sites are found with both 2009 H3N1 and 1968 H3N2 but least similar to 2009 H1N1 viral strain (Table 2), explaining the reassortment and the fact that the novel strains are the mixture of the ancestors. 2013 H3N2 may have properties of both 2009 H3N1 and 1968 H3N2 which may make it a very potent candidate for leading to huge pandemic conditions in coming future.

Glycosylation of HA represents the quality of the pathogen to escape the host defense through co-evolution with the host and identification of the host receptor [34,35]. Also it has been hypothesized that the level of glycosylation is inversely related to virulence, as glycosylation increases, the severity of the disease decreases because of improved recognition ability of the virus [36,15]. Though number of glycosylation sites was found to be more in 2012 H3N2 as compared to 2009 H3N1 and 1968 H3N2 which states that this strain might be less virulent

007

Table 2: Comparison of N-glycosylation sites of HA of all 16 strains.

A/South Carolina/1/1918(H1N1)		A/Hong Kong/1/1968(H3N2)		A/New Caledonia/20/1999(H1N1)		A/Kawasaki/173/2001(H1N1)	
Position	Sequence	Position	Sequence	Position	Sequence	Position	Sequence
27	NNST	24	NSTA	27	NNST	27	NNST
28	NSTD	38	NGTL	28	NSTD	28	NSTD
40	NVTV	54	NATE	40	NVTV	40	NVTV
104	NGTC	97	NETW	71	NCSV	71	NCSV
304	NSSL	181	NVTM	104	NGTC	104	NGTC
498	NGTY	301	NGSI	142	NHTV	142	NHTV
557	NGSL	499	NGTY	176	NLSK	176	NLSM
				303	NSSL	303	NSSL
				497	NGTY	497	NGTY
				556	NGSL	556	NGSL

A/Fujian/411/2002(H3N2)		A/Norway/807/2004(H3N2)		A/Vietnam/1203/2004(H5N1)		A/Wisconsin/67/2005(H3N2)	
Position	Sequence	Position	Sequence	Position	Sequence	Position	Sequence
24	NSTA	24	NSTA	27	NSTE	24	NSTA
38	NGTI	38	NGTI	39	NVTV	38	NGTI
54	NATE	54	NATE	170	NSTY	54	NATE
79	NCTL	79	NCTL	181	NNTN	79	NCTL
138	NESF	138	NESF	209	NPTT	142	NWTG
142	NWTG	142	NWTG	302	NSSM	149	NGTS
149	NGTS	149	NGTS	496	NGTY	160	NNSF
160	NKSF	160	NNSF	555	NGSL	181	NVTM
181	NVTM	181	NVTM			262	NSTG
262	NSTG	262	NSTG			301	NGSI
301	NGSI	301	NGSI			499	NGTY
499	NGTY	499	NGTY				

A/Indonesia/5/2005(H5N1)		A/Brisbane/10/2007(H3N2)		A/Brisbane/59/2007(H1N1)		A/California/07/2009(H1N1)	
Position	Sequence	Position	Sequence	Position	Sequence	Position	Sequence
26	NNST	24	NSTA	27	NNST	27	NNST
27	NSTE	38	NGTI	28	NSTD	28	NSTD
39	NVTV	54	NATE	40	NVTV	40	NVTV
100	NPTN	79	NCTL	71	NCSV	104	NGTC
170	NSTY	138	NESF	104	NGTC	293	NTTC
181	NNTN	142	NWTG	142	NHTV	304	NTSL
209	NPTT	149	NGTS	176	NLSK	498	NGTY
302	NSSM	160	NNSF	303	NSSL	557	NGSL
500	NGTY	181	NVTM	497	NGTY		
559	NGSL	262	NSTG	556	NGSL		
		301	NGSI				
		499	NGTY				

A/Perth/16-RGcH5-3/2009(H3N1)		A/Qingdao/2157/2009(H3N2)		A/Georgia/3058/2012(H3N2)		A/Ontario/001/2013(H3N2)	
Position	Sequence	Position	Sequence	Position	Sequence	Position	Sequence
24	NSTA	38	NGTI	38	NGTI	24	NSTA
38	NGTI	54	NATE	54	NATE	38	NGTI
54	NATE	79	NCTL	79	NCTL	54	NATE
180	NSTY	138	NESF	138	NESF	61	NSSI
191	NNTN	142	NWTG	142	NWTG	79	NCTL
219	NPTT	149	NGTS	149	NGTS	138	NESF
308	NGSI	181	NVTM	160	NNSF	142	NWTG
506	NGTY	262	NSTG	181	NVTM	149	NGTS
		301	NGSI	262	NSTG	160	NSSF
		499	NGTY	301	NGSI	181	NVTM
				499	NGTY	262	NSTG
						301	NGSI
						499	NGTY

Sites in green are the similar sites between 1968 H3N2 and 2013 H3N2. Blue site NGTY is the site which is common in all the strains down the evolutionary pathway.

as compared to its ancestors, but it is very important to state that many of its glycosylation sites are similar to the ancestors which makes it no less virulent than the ancestral strains and highlights it high potential to cause huge influenza pandemics in the future.

One more interesting point which was observed by us for the first time is the conservation of glycosylation site NGTY in

all the strains across the evolutionary path of virus from 1918 till 2012. The conservation of NGTY makes it a good drug target as it is present all through the evolutionary track. This might act as a signature sequence which might help several drugs to target the virus very carefully.

Antigenic divergence

Antigenic analysis shows that the 2013 H3N2 shared eight top similar CTL epitopes with 2012 H3N2 and nine similar CTL epitopes with 2009 H3N2. Surprisingly 2013 H3N2 shared no

similar antigenic sites with the ancestor 1968 H3N2 or 2009 H3N1. It was observed by us that it had one new epitope VASSGTLEF which was not similar to any other strains included in the study (Table 3). Even when we analyzed the 2012 H3N2 strains, we observed two new epitopes QELPGNDDS and SEVEGRIQD which were not found in any other strains included in the study and hence may be very important as a potent drug target which will specifically target these newly assorted influenza strains.

This result suggests that 2013 H3N2 is very different in

Table 3: Comparison of antigenicity of HA of all 16 strains.

A/South Carolina/1/1918(H1N1)		A/Hong Kong/1/1968(H3N2)		A/New Caledonia/20/1999(H1N1)		A/Kawasaki/173/2001(H1N1)	
Position	Sequence	Position	Sequence	Position	Sequence	Position	Sequence
42	TVTHSVNLL	398	RVIEKTNEK	42	TVTHSVNLL	42	TVTHSVNLL
210	QNADAYVSV	240	GMSSRISIY	513	REKIDGVKL	513	REKIDGVKL
158	ASSFYRNLL	406	KFHQIEKEF	206	ALYHTENAY	137	GSSWPNHTV
242	GRMNYYWTL	514	RFQIKGVEL	209	HTENAYVSV	206	ALYHTENAY
437	TYNAELLVL	28	TLCLGHHAV	418	RRMENLNKK	209	HTENAYVSV
457	SNVRNLYEK	333	TGMRNVPEK	436	TYNAELLVL	418	RRMENLNKK
336	RNIPSIQSR	404	NEKFHQIEK	157	KSSFYRNLL	436	TYNAELLVL
509	ESKLNREEI	510	ALNNRFQIK	335	RNIPSIQSR	157	KSSFYRNLL
544	LVSLGAISF	187	DNFDKLYIW	408	AVGKEFNKL	169	RKNGLYPNL
558	GSLQCRICI	251	IVKPGDVLV	543	LVSLGAISF	335	RNIPSIQSR

A/Fujian/411/2002(H3N2)		A/Norway/807/2004(H3N2)		A/Vietnam/1203/2004(H5N1)		A/Wisconsin/67/2005(H3N2)	
Position	Sequence	Position	Sequence	Position	Sequence	Position	Sequence
361	GMVDGWYGF	172	HLKFKYPAL	156	KSSFFRNVV	172	HLKFKYPAL
399	RLIGKTNEK	361	GMVDGWYGF	74	AGWLLGNPM	361	GMVDGWYGF
407	KFHQIEKEF	399	RLIGKTNEK	81	PMCDEFINV	399	RLIGKTNEK
515	RFQIKGVEL	407	KFHQIEKEF	158	SFFRNVVWL	407	KFHQIEKEF
553	WACQKGNIR	515	RFQIKGVEL	240	GRMEFFWTI	515	RFQIKGVEL
28	TLCLGHHAV	553	WACQKGNIR	425	KMEDGFLDV	553	WACQKGNIR
34	HAVPNGTIV	28	TLCLGHHAV	435	TYNAELLVL	28	TLCLGHHAV
334	TGMRNVPEK	34	HAVPNGTIV	512	REEISGVKL	34	HAVPNGTIV
405	NEKFHQIEK	161	NSFFSRLNW	241	RMEFFWTIL	161	NSFFSRLNW
344	TRGIFGAIA	334	TGMRNVPEK	274	KKGDSTIMK	334	TGMRNVPEK

A/Indonesia/5/2005(H5N1)		A/Brisbane/10/2007(H3N2)		A/Brisbane/59/2007(H1N1)		A/California/07/2009(H1N1)	
Position	Sequence	Position	Sequence	Position	Sequence	Position	Sequence
156	SPSFFRNVV	172	HLKFKYPAL	42	TVTHSVNLL	42	TVTHSVNLL
525	ESIGTYQIL	361	GMVDGWYGF	513	REKIDGVKL	210	QNADAYVFV
561	SLQCRICIK	399	RLIGKTNEK	206	ALYHTENAY	240	REGRMNYYW
274	KKGDSAIMK	407	KFHQIEKEF	209	HTENAYVSV	217	FVGSSRYSK
74	AGWLLGNPM	515	RFQIKGVEL	418	RRMENLNKK	239	DREGRMNYY
81	PMCDEFINV	553	WACQKGNIR	436	TYNAELLVL	242	GRMNYYWTL
158	SFFRNVVWL	28	TLCLGHHAV	335	RNIPSIQSR	539	SSLVLVVSL
240	GRMEFFWTI	34	HAVPNGTIV	408	AVGKEFNKL	437	TYNAELLVL
429	KMEDGFLDV	161	NSFFSRLNW	543	LVSLGAISF	544	VVSLGAISF
439	TYNAELLVL	334	TGMRNVPEK	557	GSLQCRICI	336	RNIPSIQSR

A/Perth/16-RGcH5-3/2009(H3N1)		A/Qingdao/2157/2009(H3N2)		A/Georgia/3058/2012(H3N2)		A/Ontario/001/2013(H3N2)	
Position	Sequence	Position	Sequence	Position	Sequence	Position	Sequence
166	KSSFFRNVV	282	SIMRSDAPI	373	NSEGRGQAA	282	SIMRSDAPI
368	GMVDGWYGF	373	NSEGRGQAA	485	IYHKCDNAC	373	NSEGRGQAA
406	RLIGKTNEK	485	IYHKCDNAC	89	DPQCDGFQN	485	IYHKCDNAC
414	KFHQIEKEF	89	DPQCDGFQN	245	RVSIYWTIV	89	DPQCDGFQN
522	RFQIKGVEL	245	RISIYWTIV	17	QELPGNDDS	245	RISIYWTIV
560	WACQKGNIR	17	QKLPGNDNS	25	STATLCLGH	17	QKLPGNDNS
28	TLCLGHHAV	25	STATLCLGH	55	ATELVQSSS	25	STATLCLGH
34	HAVPNGTIV	55	ATELVQSSS	90	PQCDGFQNK	90	PQCDGFQNK
341	TGMRNVPEK	90	PQCDGFQNK	219	TVSTKRSQQ	128	VASSGTLEF
412	NEKFHQIEK	219	TVSTKRSQQ	416	SEVEGRIQD	219	TVSTKRSQQ

Blue colored antigenic site VASSGTLEF is the site which is common in all the strains down the evolutionary pathway.

terms of antigenicity, which nominates it as a quite virulent strain which might be of great risk in future pandemic break too, because the virus is changing its antigenic sites which need the development of new site targeting drugs in order to combat with the virus. Also, this finding may be correlated with extreme penetrance of 2013 H3N2, as it has novel antigenic sites and hence humans may lack immunity against it. Although the epitope sequences from ancestor 1968 H3N2 strain are present in sequence of 2012 H3N2, but these are not selected as epitopes, instead all new epitopes are observed according to CTL epitope prediction, which suggests that antibodies against these novel epitopes may not be available to give resistance against the virus and people may be still affected. These novel sites may be of great threat as studies about them is in embryonic stages and would need a raging speed of development of drugs to target them. These new sites make the virus more unpredictably virulent. This highlights that it might prove as a source of big pandemic potential in the coming years.

Primary structure analysis

Individual amino acid sequences of 2013 H3N2 were observed to be more similar to 2012 H3N2 and 1968 H3N2, but lesser with 2009 H3N1 and least with 2009 H1N1. It was seen that there was quite a similar percentages of amino acids present in their sequences, even hydrophilic amino acids were observed to be quite similar all through the strains (Table 4). It was seen in weblogo analysis that in the starting of the sequence the amino acids were mostly conserved but loads of variation was seen in the middle portion of the amino acid sequence (Figure 4). Every time when a new viral strain emerges it is reported that viral genome reasserts and results as a mixture of genes from humans and other hosts. During this analysis we saw amino acids of 2013 H3N2 were more similar to the ancestor 1968 strain rather than 2009 HINI or 2009 H3N1 pandemic strain which highlights the evolution which is still linked to the ancestor strain pointing that 2013 H3N2 strain is equally virulent as 1968 strain and might also be responsible for future outbreaks too.

Middle HA region of these all influenza strains are subjected to maximum number of mutations during the reassortment of the already existing virus and introduction of novel strain. This similarity between 2012 H3N2 and 1968 H3N2 strain, and lesser similarity with 2009 H3N1 indicates the hypothesis that 2013 viral strain, possesses the similar amino acids as ancestor strain, and hence may be similarly functional and equally dangerous as huge pandemic causing 1968 H3N2 strain. Hence

Table 4: Comparison of percentage of individual amino acids present in the sequences of all 16 strains.

Amino acids	A/South Carolina/1/1918(H1N1)	A/Hong Kong/1/1968(H3N2)	A/New Caledonia/20/1999(H1N1)	A/Kawasaki/173/2001(H1N1)	A/Fujian/411/2002(H3N2)	A/Norway/807/2004(H3N2)	A/Vietnam/1203/2004(H5N1)	A/Wisconsin/67/2005(H3N2)
Ala (A)	6.50%	4.90%	5.30%	5.30%	5.50%	5.50%	5.00%	5.30%
Arg (R)	3.50%	4.80%	3.40%	3.50%	4.90%	4.80%	3.90%	4.80%
Asn (N)	7.40%	7.40%	8.30%	8.10%	7.80%	8.10%	8.30%	8.10%
Asp (D)	4.40%	5.70%	3.20%	3.40%	5.30%	5.30%	4.60%	5.30%
Cys ©	2.80%	3.20%	2.80%	2.80%	3.20%	3.20%	2.70%	3.20%
Gln (Q)	3.00%	4.10%	2.80%	2.80%	4.20%	4.20%	3.50%	4.20%
Glu (E)	6.20%	5.10%	7.10%	6.90%	4.80%	4.80%	7.10%	4.80%
Gly (G)	7.80%	8.00%	7.60%	7.60%	7.60%	7.60%	6.70%	7.40%
His (H)	2.30%	1.90%	2.50%	2.50%	1.60%	1.90%	2.50%	1.90%
Ile (I)	5.70%	8.10%	5.70%	5.70%	8.50%	8.70%	7.10%	8.80%
Leu (L)	9.00%	7.60%	9.00%	9.00%	8.00%	7.80%	8.50%	7.80%
Lys (K)	5.80%	5.30%	6.20%	6.00%	6.50%	6.40%	6.40%	6.40%
Met (M)	1.40%	1.80%	1.60%	1.80%	1.40%	1.40%	2.50%	1.40%
Phe (F)	3.40%	3.90%	3.50%	3.50%	3.70%	3.90%	3.40%	4.10%
Pro (P)	3.40%	3.50%	3.70%	3.70%	3.40%	3.50%	3.40%	3.50%
Ser (S)	8.70%	6.90%	8.50%	8.50%	7.40%	7.20%	7.80%	7.20%
Thr (T)	6.50%	7.10%	5.50%	5.50%	6.20%	6.00%	4.80%	6.00%
Trp (W)	1.90%	2.10%	1.90%	1.90%	1.90%	1.90%	1.80%	1.90%
Tyr (Y)	4.60%	3.20%	4.80%	4.80%	3.70%	3.40%	4.30%	3.40%
Val (V)	5.70%	5.30%	6.50%	6.50%	4.40%	4.40%	5.90%	4.40%

Amino acids	A/Indonesia/5/2005(H5N1)	A/Brisbane/10/2007(H3N2)	A/Brisbane/59/2007(H1N1)	A/California/07/2009(H1N1)	A/Perth/16-RGcH5-3/2009(H3N1)	A/Qingdao/2157/2009(H3N2)	A/Georgia/3058/2012(H3N2)	A/Ontario/001/2013(H3N2)
Ala (A)	5.10%	5.30%	5.30%	6.00%	5.40%	5.70%	5.80%	5.50%
Arg (R)	4.00%	4.60%	3.20%	3.40%	3.70%	4.80%	4.80%	4.90%
Asn (N)	8.30%	8.10%	8.50%	7.20%	8.40%	8.00%	8.10%	7.60%
Asp (D)	4.40%	5.30%	3.20%	4.60%	4.90%	5.30%	5.30%	5.30%
Cys ©	2.60%	3.20%	2.80%	2.70%	3.10%	3.20%	3.20%	3.20%
Gln (Q)	3.30%	4.20%	2.80%	2.50%	4.00%	4.40%	4.40%	4.20%
Glu (E)	7.00%	4.90%	7.10%	6.20%	6.50%	4.80%	4.90%	4.90%
Gly (G)	6.70%	7.40%	7.60%	7.10%	7.00%	7.40%	7.40%	7.40%
His (H)	2.50%	1.90%	2.70%	2.70%	2.10%	1.90%	2.10%	1.90%
Ile (I)	7.20%	8.80%	5.80%	6.50%	8.40%	8.70%	8.50%	8.80%
Leu (L)	8.60%	7.60%	8.80%	8.10%	8.00%	7.80%	7.80%	7.80%
Lys (K)	6.90%	6.40%	6.70%	7.40%	7.00%	6.50%	6.00%	6.40%
Met (M)	2.60%	1.40%	1.60%	1.20%	1.60%	1.40%	1.40%	1.60%
Phe (F)	3.20%	4.10%	3.50%	3.40%	3.80%	4.10%	4.10%	4.10%
Pro (P)	3.50%	3.70%	3.70%	3.50%	3.50%	3.50%	3.50%	3.50%
Ser (S)	8.10%	7.10%	8.50%	8.30%	6.50%	6.90%	7.10%	7.60%
Thr (T)	4.90%	6.20%	5.10%	6.40%	5.60%	6.20%	5.80%	5.70%
Trp (W)	1.80%	1.90%	1.80%	1.80%	2.10%	1.90%	1.90%	1.90%
Tyr (Y)	4.20%	3.40%	4.60%	4.80%	3.70%	3.40%	3.20%	3.40%
Val (V)	5.10%	4.40%	6.50%	6.40%	4.90%	4.20%	4.60%	4.20%

It displays the intensity of an amino acid, coming number of times (in percentage) in sequence of all 16 strains.
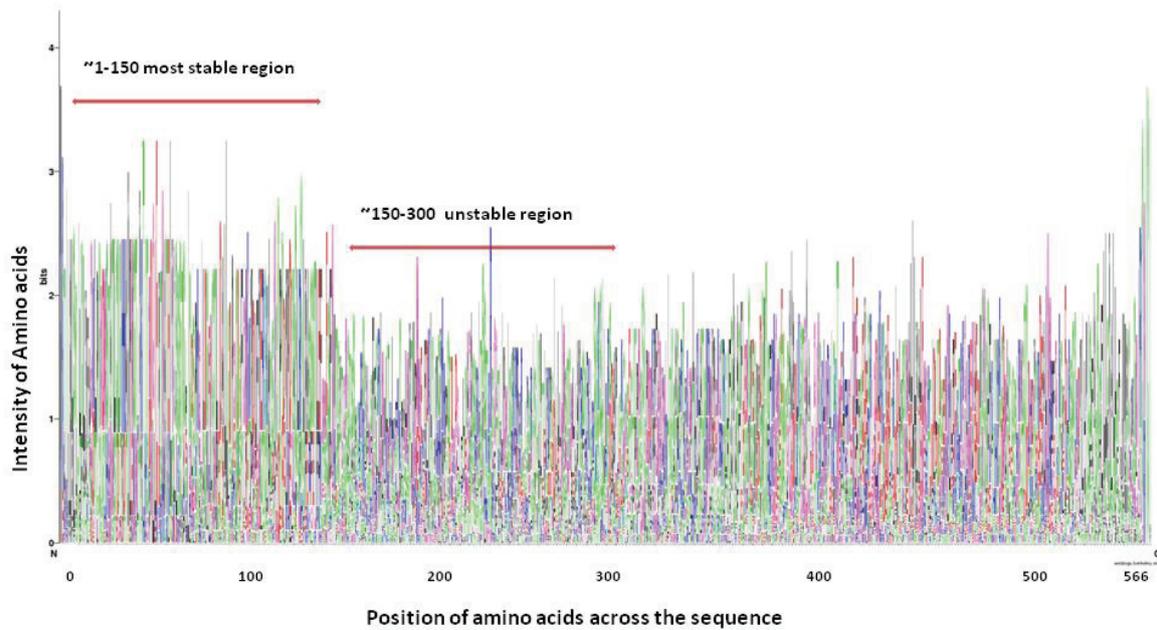
Figure 4: Weblogo display of amino acids. Weblogo showing the frequency of amino acids in each position along the sequence when amino acids of 2013 H3N2, 2012 H3N2, 1968 H3N2, 2009 H3N1 and 2009 H1N1 were compared. The higher the peak of an amino acid at the position, higher is the stability and conservation of that amino acid at that particular position.

this novel strain may be of huge pandemic potential in coming future also.

Protein disorder

Protein disordered regions prediction revealed uneven disordered regions in all the strains all through the evolutionary track. The majority of disordered regions were different among all the H1 and H3 strain isolates. There were three common disordered positions found in 2013 H3N2 and 1968 H3N2 and lesser similarity with 2009 H3N1 namely IEVNATEL, AADLKS and CIS which were found in relatively different positions (Table 5). More importantly two common disordered regions "STRSSQQ" in 2013 H3N2 and 1968 H3N2 strain fall into the R region (108–261 amino acids) of HA residues, which is importantly known for the presence of receptor binding sites and neutralizing antibodies epitopes [2].

Disordered protein sequences may destabilize the receptor binding domains which can affect their efficacy for attachment of viral proteins and transcription factors and making the virulent strain more virulent. The protein disorder analysis reflects that the number of protein disorder sites have decreased in 2013 strain when compared to the ancestors.

During our analysis we saw that the protein disordered region and the antigenic sites of 2013 H3N2 were overlapping. We found that the antigenic site SIMRSDAPI, RISIYWTIV and TVSTKRSQQ were falling in the protein disordered regions too. This indicates that the antigenic sites may not be stable and are disordered suggesting that with these disordered antigenic epitopes, people might still get the infection as even if the drugs are made which target these epitopes, might not be able to attach to these epitopes because of the changed dynamics of these disordered epitopes.

Secondary structure modelling and analysis

Secondary structure comparison of the HA proteins of 2013 H3N2, 2012 H3N2, 1968 H3N2, 2009 H1N1 and 2009 H3N1 was done along with 11 more strains taken from the ancient most till 2013. The comparison suggests that the structure of 2013 H3N2 strain (H-bonds: 238) shows more similarity with 1968 H3N2 (H-bonds: 231) and less similar with 2009 H3N1 (H-bonds: 227) and 2009 H1N1 (H-bonds: 218).

However, structural comparison by comparing the difference between RMSD values shows that there is more similarity between 2013 H3N2 and 2012 H3N2 with RMSD value of 2.797 and with 1968 H3N2 strain with RMSD value of 2.531, lesser with 2009 H3N1 with RMSD value of 3.127 and least with 2009 H1N1 with RMSD value of 3.226. But surprisingly 2012 H3N2 structure was found to be most similar with 2009 H3N2 with RMSD value of 1.869. Also, the difference between RMSD values of HA of 2012 H3N2 was 3.297 with 1918 H1N1, 3.246 with 1999 H1N1, 3.119 2001 H1N1, 2.938 with 2002 H3N2, 2.732 with 2004 H3N2, 2.615 with 2005 H3N2, 2.487 with 2007 H3N2, 3.022 with 2007 H1N1 and 1.869 with 2009 H3N2.

Minimal difference in number of hydrogen bonds and RMSD values, directly depict 2013 H3N2 as more similar with 1968 H3N2 and less similar with 2009 H3N1 and 2009 H1N1. Number of helix, coils and turns also show the similarity (Table 1). These results suggest that 2013 H3N2 might be potent infectious strain causing epidemics globally in the coming future. Also similar RMSD of 2013 H3N2 with 2009 H3N2 as compared to ancestor 1968 H3N2, portrays that during the evolutionary track virus might me more similar to the predecessor relatives, having

Table 5: Comparison of protein disordered regions.

A/South Carolina/1/1918(H1N1)	A/Hong Kong/1/1968(H3N2)	A/New Caledonia/20/1999(H1N1)	A/Kawasaki/173/2001(H1N1)
1M	47 DDQIEVTNATEL 58	1M	1M
173 SSYPKLSK 180	206 EQTSLYVQASGRVTVSTRRSQQTIIP NIGSRPWVRGMSSRI 246	123 SSVSSFE 129	123 SSVSS 127
205 QQSLYQNAD 213	278 TGKSSI 283	206 ALYHT 210	145 VTGVS 149
226 RFTPEIAARPKVR V238	289 P	225 RFTPEIAKRPKVRDQEG 241	225 RFTPEIAKRPKVRDQEG 241
327 TKLRMATGL 335	293 CIS 295	291 CDAKC 295	291 CDAKC 295
379 AADQKS 384	380 AADLKS 385	327 KLRMVTGL 334	327 KLRMVTGL 334
412 KEFNNLERRIEN 423	452 TIDLTDSEMNKLFE KTRRQLRENA 475	378 AADQKSTQN 386	378 AADQKSTQN 386
518 DGVK 521		411 KEFNKLERRMENLN 424	411 KEFNKLERRMENLN 424

A/Fujian/411/2002(H3N2)	A/Norway/807/2004(H3N2)	A/Vietnam/1203/2004(H5N1)	A/Wisconsin/67/2005(H3N2)
50 IEVTNATELVQSSSTGGICD 69	50 IEVTNATELVQSSSTGGICD 69	34 TIMEK 38	50 IEVTNATELVQSSSTGGICD 69
212 AQASGRITV STKRSQQTVI PNIGSRPRVR DVSSRIS 247	210 LYAQASGRITV STKRSQQTVIPNIGSRPRVR DIPSRIS 247	142 EASL 145	221 STKRSQQTVI PNIGSRPRIR NIPSRIS 247
278 SGKSSIMR 285	278 SGKSSIMR 285	212 TYISVGTST LNQRLVPRIA TRS 233	278 SGKSSIMR 285
380 AADLKST 386	333 A	329 LATGL 333	333 A
461 NKLFERTKKQLREN 474	384 K	412 FNNLERRIENLN 423	384 K
	452 TIDLTDSEM NKLFERTKKQ LREN 474	412 FNNLERRIENLN 423	452 TIDLTDSEM NKLFERTKKQ LREN 474
		463 KVRLQL 468	
		512 REEISGVK 519	
		536 ASSLAL 541	

A/Indonesia/5/2005(H5N1)	A/Brisbane/10/2007(H3N2)	A/Brisbane/59/2007(H1N1)	A/California/07/2009(H1N1)
34 TIMEKN 39	50 IEVTNATELVQSSSTGEICDSPH 72	120EQLSSVSSFE129	91 SSWSYIV 97
145 SG 146	220 V STKRSQQTVI PNIGSRPRVR NIPSRIS 247	226FTPEIAKRPKVRDQEG 241	121 QLSSVSS 127
205 RLYQNPTTYISIGTST LNQRLVPKIATRS 233	278 SGKSSIMR 285	327 KLRMVTGL 334	173 NSYPK 177
329 LATGLRNSPQRE SRRKKRGL 348	333 A	378 AADQKSTQN386	
416 FNNLERRIENLN 427	381 AD 382	411 KEFNKLERRMENL 424	228 KPEIAIRPKVRDR EGRMN 245
467 KVRLQL 472	452 TID 454		322 KYVKSTKLRLATGL 335
515 KREEISGVKL 524			517 IDGVKL 522
537 STVASSLALAI 547			

A/Perth/16-RGcH5-3/2009(H3N1)	A/Qingdao/2157/2009(H3N2)	A/Georgia/3058/2012(H3N2)	A/Ontario/001/2013(H3N2)
50 IEVTNATELVQ SSSTGEI 67	50 IEVTNATELVQ SSSTGEIC 68	50 IEVTNATELVQSSSTGEICNSPH 72	48 DRIEVTNATELVQ NSSIGEICDS 70
152 EASL 155	221 STKRSQQTAIPNIGSRPRVR NIPSRIS 247	221 STKRSQQAVIPNIGSRPRVRNIPSRVS 247	220 VSTKRSQQAVIPNIGSRPRIRNIPSRIS 247
222 TYISVGTSTLNQRLVPRIATR 243	278 SGKSSIMR 285	278 SGKSSIM 284	278 SGKSSIMRSDA 288
302 S	333 A	333 A	332 LAT 334
340 A	380 AADLKS 385	380 AADLKS 385	380 AADLKS 385
387 AADLKS 392	452 TID 454	452 TID 454	452 TID 454
459 T ID 461			

Comparison of protein disordered regions as predicted by PONDR software in HA of all 16 strains. The Table displays the position and length of the disordered proteins in all the strains involved in this study.

all the modification to escape immunity, and yet be similar to ancestor to hold the potency to cause severe pandemic in the coming future (Figure 5).

Developing countries may be prone to higher risk for infection, because of high population of vectors, supporting environment for vectors to flourish, more import of population which might indirectly import the virus as well and can recombine in a way which can evolve and can be resistant to the already reported drugs and may create panic all over the world. Continuous monitoring of circulating virus characteristics is very important and essential tool to understand the epidemiological and virological features of influenza viruses. This may help in matching them with seasonal vaccine strains for regulation of efficient vaccination strategies [4]. Pre-existing neutralizing antibody in the host system provide initial defence against influenza virus, hence annual vaccinations are needed to be provided to maintain protective levels of antibody in the system. Several live attenuated virus vaccines have been designed to combat with these reassorting viruses [16] but these clever frequently altering strains, pose a big challenge for enhanced designing of vaccines/drugs which may be able to efficiently target the virus [37,38]. Amino acid sequence analysis of the latest strain can reveal motifs, receptor
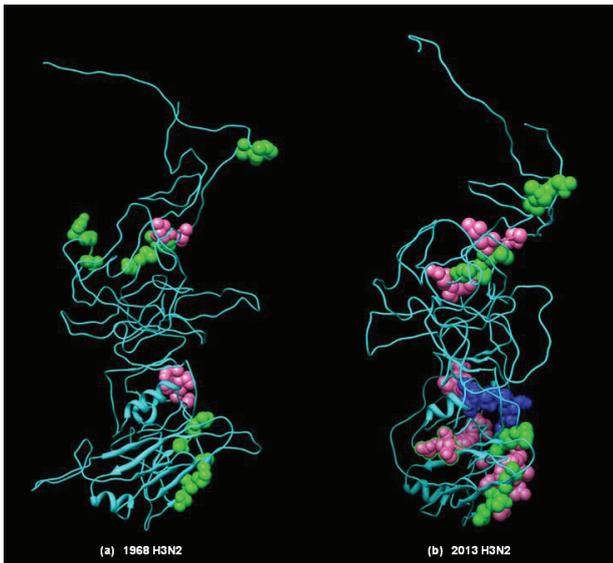
012

Figure 5: Secondary structures of HA. Comparative display of glycosylation and antigenic sites in secondary structures of HA of 1968 H3N2 (a) and 2013 H3N2 (b). Green spheres show similar glycosylation sites, whereas pink spheres show dissimilar glycosylation sites. Blue spheres display novel antigenic sites in 2013 H3N2 (b) which are not found in any other strains of this study.

specificity, potential glycosylation sites which can act as drug targets as they are sensitive to drugs like oseltamivir [39] zanamivir, and amantadine [40]. Novel prophylactic strategies have been designed against influenza viruses. Very promising ones are Toll like receptors (TLRs) which may be capable in interrupting the transmission of the virus by aiding cytokine production by host immune cells in mammals [41]. At molecular level, it is possible to modify influenza viruses, generating mutant or reassortants which can be further analyzed for their characteristics like pathogenicity, virulence, hosts and transmissibility, which may aid in developing vaccines to combat future pandemic influenza viruses [42]. U.S. influenza Vaccine Effectiveness (Flu VE) Network is trying to monitor overall vaccine effectiveness (VE) on factors like age, site, race, health, and days from illness onset in patients. It reports that VE is estimated to be 47% efficient against influenza A (H3N2) virus infections, but the efficiency lowers down in adults of 65 years age or more [43]. There are several proposed aspects to combat against influenza viruses. Antiviral medications may be used for treatment in patients, like statins which control elevated cholesterol and also have antiviral properties but are very less efficient [44]. Vaccination should be provided to immunocompromised patients having chronic diseases such as Cirrhotic patients in order to reduce the morbidity and mortality of influenza [45]. The matrix 2 protein (M2e) and the internal nucleoprotein (NP) of influenza viruses remain highly conserved and hence can serve as potent antigens for universal influenza vaccine development which may provide high immunogenicity. Studies on cloning of high-affinity human monoclonal antibodies against the influenza virus might also be a good tool [46]. Vaccination with recombinant NM2e fusion protein may be quite promising universal influenza vaccine [47]. Prokaryotically expressed influenza A virus NP elicits cross-protection against influenza virus in mice models,

giving high immune response and protective efficacy [48]. Antibody C05 is reported to recognize conserved motifs of HA receptor-binding sites, which has the ability to neutralize several influenza strains [49]. Action of a single intrinsic immune effector Interferon-inducible transmembrane (IFITM3) protein can very potently restrict virus replication, in mouse and humans [50]. All the analysis done above, including sequence homology, phylogeny, antigenic and N-glycosylation sites analysis show the similarity of the novel viral strain 2013 H3N2 with influenza 1968 H3N2, lesser with 2009 H3N1 and least with 2009 H1N1. Influenza A H3N2 may be correlated with extreme penetrance of novel influenza and hence humans may lack immunity against it which highlights that it might prove as a potent source of pandemics in the coming future. This influenza strain has auxiliary pandemic potential and should be taken into consideration by the sense of urgency.

Acknowledgements

References

1. Kobasa D, Takada A, Shinya K, Hatta M, Halfmann P, et al. (2004) Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. Nature 431: 703-707. **Link:** http://bit.ly/3cbzu0M

2. Masoodi TA, Shaik NA, Shafi G, Munshi A, Ahamed AK, et al. (2012) Comparative analysis of hemagglutinin of 2009 H1N1 influenza A pandemic indicates its evolution to 1918 H1N1 pandemic. Gene 491: 200-204. **Link:** http://bit.ly/3a6ZJ6Y

3. World Health Organisation, influenza virus activity in the world (2013) Source: Laboratory confirmed data from the Global Influenza Surveillance and Response System (GISRS). **Link:** http://bit.ly/2HYr9zT

4. Pariani E, Amendola A, Ranghiero A, Anselmi G, Zanetti A (2013) Surveillance of influenza viruses in the post-pandemic era (2010-2012) in Northern Italy. Hum Vaccin Immunother 9: 2013. **Link:** http://bit.ly/3ciZVlw

5. European Centre for Disease Prevention and Control (ECDC) (2013) Weekly influenza surveillance overview. Main surveillance developments in week Surveillance Report **Link:** http://bit.ly/3cepl3N

6. Bragstad K, Emborg H, Kolsen Fischer T, Voldstedlund M, et al. (2013) Low vaccine effectiveness against influenza A(H3N2) virus among elderly people in Denmark in 2012/13--a rapid epidemiological and virological assessment. Euro Surveill 18. **Link:** http://bit.ly/32r0qW0

7. Ratmann O, Donker G, Meijer A, Fraser C, Koelle K. (2012) Phylodynamic inference and model assessment with approximate bayesian computation: influenza as a case study. PLoS Comput Biol 8: e1002835. **Link:** http://bit.ly/2SXfmbk

8. Thangavel RR, Reed A, Norcross EW, Dixon SN, Marquart ME, et al. (2011) "Boom" and "Bust" cycles in virus growth suggest multiple selective forces in influenza A evolution. Virol J 8: 180. **Link:** http://bit.ly/38Zz3ER

9. Ann J, Papenburg J, Bouhy X, Rhéaume C, Hamelin MÈ, et al. (2012) Molecular and antigenic evolution of human influenza A/H3N2 viruses in Quebec, Canada, 2009-2011. J Clin Virol 53: 88-92. **Link:** http://bit.ly/2HW4BzJ

10. Illingworth CJ, Mustonen V (2012) Components of selection in the evolution of the influenza virus: linkage effects beat inherent selection. PLoS Pathog 8: e1003091. **Link:** http://bit.ly/2wFHfMf

11. Li Z, Watanabe T, Hatta M, Watanabe S, Nanbo A, et al. (2009) Mutational analysis of conserved amino acids in the influenza A virus nucleoprotein. J Virol 83: 4153-4162. **Link:** http://bit.ly/38W9hRY

12. Saxena SK, Mishra N, Saxena R, Swamy ML, Sahgal P, et al. (2009) Structural and antigenic variance between novel influenza A/H1N1/2009 and influenza A/H1N1/2008 viruses. J Infect Dev Ctries 4: 1-6. **Link:** http://bit.ly/2SXBouC

13. Cohen J, Enserink M (2009) Infectious diseases. As swine flu circles globe, scientists grapple with basic questions. Science 324: 572-573. **Link:** http://bit.ly/380flb1

14. Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature 289: 366-373. **Link:** http://bit.ly/2T3d7U5

15. Sun S, Wang Q, Zhao F, Chen W, Li Z (2012) Prediction of biological functions on glycosylation site migrations in human influenza H1N1 viruses. PLoS One 7: e32119. **Link:** http://bit.ly/381GwCe

16. Gallaher WR (2009) Towards a sane and rational approach to management of Influenza H1N1 2009. Virol J 6: 51. **Link:** http://bit.ly/2T9p6xS

17. Igarashi M, Ito K, Yoshida R, Tomabechi D, Kida H, et al. (2009) Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. PLoS One 5: e8553. **Link:** http://bit.ly/38W9xQW

18. Hai R, Krammer F, Tan GS, Pica N, Eggink D, et al. (2012) Influenza viruses expressing chimeric hemagglutinins: globular head and stalk domains derived from different subtypes. J Virol 86: 5774-5781. **Link:** http://bit.ly/39bNRAF

19. Sarkar M, Agrawal AS, Sharma Dey R, Chattopadhyay S, Mullick R, et al. (2011) Molecular characterization and comparative analysis of pandemic H1N1/2009 strains with co-circulating seasonal H1N1/2009 strains from eastern India. Arch Virol 156: 207-217. **Link:** http://bit.ly/2VtWoLf

20. Medina RA, Manicassamy B, Stertz S, Seibert CW, Hai R, et al. (2010) Pandemic 2009 H1N1 vaccine protects against 1918 Spanish influenza virus. Nat Commun 1: 28. **Link:** https://go.nature.com/38WabOm

21. Ali A, Yassine H, Awe OO, Ibrahim M, Saif YM, et al. (2013) Replication of swine and human influenza viruses in juvenile and layer turkey hens. Vet Microbiol 163: 71-78. **Link:** http://bit.ly/32qpWKU

22. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, et al. (2009) Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 325: 197-201. **Link:** http://bit.ly/390KTid

23. Saxena SK, Mishra N, Saxena R, Saxena S (2009) Swine flu: influenza A/H1N1 2009: the unseen and unsaid. Future Microbiol 4: 945-947. **Link:** http://bit.ly/2vgndYt

24. Patel J (2012) Study finds children and middle-aged adults are most at risk from variant H3N2 influenza. Expert Rev Anti Infect Ther 11: 1169. **Link:** http://bit.ly/2SZuAwx

25. Graham-Rowe D (2011) Epidemiology: Racing against the flu. Nature 480: S2-S3. **Link:** http://bit.ly/381kMq7

26. Ma W, Gramer M, Rossow K, Yoon KJ (2006) Isolation and genetic characterization of new reassortant H3N1 swine influenza virus from pigs in the midwestern United States. J Virol 80: 5092-5096. **Link:** http://bit.ly/2TdDTrD

27. Marazzi I, Ho JS, Kim J, Manicassamy B, Dewell S, et al. (2012) Suppression of the antiviral response by an influenza histone mimic. Nature 483: 428-433. **Link:** http://bit.ly/2TiUfir

28. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680. **Link:** http://bit.ly/38WaMj4

29. Gupta R, Jung E, Brunak S (2004) Prediction of N-glycosylation sites in human proteins. **Link:** http://bit.ly/2VmI3ju

30. X Li, Romero P, Rani M, Dunker AK, Obradovic Z (1999) Predicting protein disorder for N-,C-, and internal regions. Genome Inform Ser Workshop Genome Inform 10: 30-40. **Link:** http://bit.ly/390rEVL

31. Romero P, Obradovic Z, Li X, Garner E, Brown C, et al. (2001) Sequence complexity of disordered protein. Proteins 42: 38-48. **Link:** http://bit.ly/3c5WeiN

32. Romero P, Obradovic Z, Dunker AK (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family. Genome Inform Ser Workshop Genome Inform 8: 110-124. **Link:** http://bit.ly/2w2TG4y

33. Ekiert DC, Friesen RH, Bhabha G, Kwaks T, Jongeneelen M, et al. (2011) A highly conserved neutralizing epitope on group 2 influenza A viruses. Science 333: 843-850. **Link:** http://bit.ly/37YDVci

34. Eichler R, Lenz O, Garten W, Strecker T (2006) The role of single N-glycans in proteolytic processing and cell surface transport of the Lassa virus glycoprotein GP-C. Virol J 3: 41. **Link:** http://bit.ly/32qpTii

35. Roedig JV, Rapp E, Höper D, Genzel Y, Reichl U (2011) Impact of host cell line adaptation on quasispecies composition and glycosylation of influenza A virus hemagglutinin. PLoS One 6: e27989. **Link:** http://bit.ly/37TyKdy

36. Suzuki Y (2011) Positive selection for gains of N-linked glycosylation sites in hemagglutinin during evolution of H3N2 human influenza A virus. Genes Genet Syst 86: 287-294. **Link:** http://bit.ly/2TetSu5

37. Hai R, García-Sastre A, Swayne DE, Palese P (2011) A reassortment-incompetent live attenuated influenza virus vaccine for protection against pandemic virus strains. J Virol 85: 6832-6843. Link; http://bit.ly/32pzBSb

38. Chen Z, Wang W, Zhou H, Suguitan AL Jr, Shambaugh C, et al. (2010) Generation of live attenuated novel influenza virus A/California/7/09 (H1N1) vaccines with high yield in embryonated chicken eggs. J Virol 84: 44-51. **Link:** http://bit.ly/32s7Bgv

39. Wong ZX, Jones JE, Anderson GP, Gualano RC (2011) Oseltamivir treatment of mice before or after mild influenza infection reduced cellular and cytokine inflammation in the lung. Influenza Other Respi Viruses 5: 343-350. **Link:** http://bit.ly/2w5He49

40. Siddique N, Naeem K, Ahmed Z, Abbas MA, Farooq S, et al. (2012) Isolation, identification, and phylogenetic analysis of reassortant low-pathogenic avian influenza virus H3N1 from Pakistan. Poult Sci 91: 129-138. **Link:** http://bit.ly/2w92DJO

41. St Paul M, Mallick AI, Read LR, Villanueva AI, Parvizi P, et al. (2012) Prophylactic treatment with Toll-like receptor ligands enhances host immunity to avian influenza virus in chickens. Vaccine 30: 4524-4531. **Link:** http://bit.ly/2waVwjJ

42. Neumann G, Ozawa M, Kawaoka Y (2012) Reverse genetics of influenza viruses. Methods Mol Biol 865: 193-206. **Link:** http://bit.ly/391UntB

43. Centers for Disease Control and Prevention (2013) Interim adjusted estimates of seasonal influenza vaccine effectiveness - United States. Morb Mortal Wkly Rep 62: 119-123. **Link:** http://bit.ly/2VpDntc

44. Kumaki Y, Morrey JD, Barnard DL (2012) Effect of statin treatments on highly pathogenic avian influenza H5N1, seasonal and H1N1pdm09 virus infections in BALB/c mice. Future Virol 7: 801-818.

45. Sayyad B, Alavian SM, Najafi F, Mokhtari Azad T, Ari Tabarestani MH, et al. (2012) Efficacy of influenza vaccination in patients with cirrhosis and inactive carriers of hepatitis B virus infection. Iran Red Crescent Med J. 14: 623-630. **Link:** http://bit.ly/2HUHPbg

014

46. Wrammert J, Smith K, Miller J, Langley WA, Kokko K, et al. (2008) Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. Nature 453: 667-671. **Link:** http://bit.ly/2w5O0GW

47. Wang W, Huang B, Jiang T, Wang X, Qi X, et al. (2012) Robust immunity and heterologous protection against influenza in mice elicited by a novel recombinant NP-M2e fusion protein expressed in E. coli. PLoS One 7: e52488. **Link:** http://bit.ly/32CTfKF

48. Huang B, Wang W, Li R, Wang X, Jiang T, et al. (2012) Influenza A virus nucleoprotein derived from Escherichia coli or recombinant vaccinia (Tiantan) virus elicits robust cross-protection in mice. Virol J 9: 322. **Link:** http://bit.ly/2T09SMR

49. Ekiert DC, Kashyap AK, Steel J, Rubrum A, Bhabha G, et al. (2012) Cross-neutralization of influenza A viruses mediated by a single antibody loop. Nature 489: 526-532. **Link:** http://bit.ly/37TzLCo

50. Everitt AR, Clare S, Pertel T, John SP, Wash RS, et al. (2012) IFITM3 restricts the morbidity and mortality associated with influenza. Nature 484: 519-523. **Link:** http://bit.ly/2wOaaOq