Journal of Proteomics & Bioinformatics - Open Access

JPB/Vol.3 Issue 2

Research Article OPEN ACCESS Freely available online doi:10.4172/jpb.1000121

Comparative Sequence Analysis on Different Strains of Swine Influenza Virus Sub-type H1N1 for Neuraminidase and Hemagglutinin

Deepak Kumar Sharma[1], Anil Kumar Rawat[2], Shipra Srivastava[2], Rajeev Srivastava[2] and Ajay Kumar[1]*

[1]Department of Biotechnology, Institute of Biomedical Education and Research, Mangalayatan University, Aligarh (U.P), India
[2]Biotechnology and Bioinformatics Division, BIOBRAINZ, 566/29 J, Jai Prakash Nagar, Alambagh, Lucknow 226005, India

Abstract

The swine flu is an infectious disease of swine and human, causing a huge amount of death to both. The aim of this study was to analyse the mutation possibility of swine influenza virus sub-type A/Swine/Nebraska/(H1N1) from swine of Nebraska. The H1N1 amino acid sequences of neuraminidase (GenBank Acc. No: ABR28650) and hemagglutinin (GenBank Acc. No: ABR28647) were analyzed for mutations using BLASTP and ClustalW programs. Our in silico analysis predicted that hemagglutinin and neuraminidase of swine influenza virus are sensitive to mutations at positions 225, 283 and 240, 451 respectively. These mutations were significant for its pathogenic nature because they are involved in change in polarity or hydrophobicity. Domain and motif search shows that mutations were detected in NA (T240A, G451S) and HA (I283V) at a predicted site of N-myristoylation. Secondary structure analysis predicted that no structural conformation changes were observed in HA and NA at positions 225, 283 and 240, 451 respectively. The program PROTMUTATION was developed in Perl CGI programming using Needleman-Wunsch algorithm for global sequence alignment. This program was used to monitor the mutations and predicts the trend of mutations.

Keywords: Neuraminidase; Hemagglutinin; Mutation; Swine flu; H1N1; Influenza A virus; PROTMUTATION

Introduction

Swine flu viruses are causing a huge amount of death to both human and swine. The World Health Organization (WHO) figures show that worldwide more than 209 countries and overseas territories or communities have reported laboratory confirmed cases of pandemic influenza H1N1 2009, including at least 15174 deaths (WHO, 5 February 2010). Pathogenicity of virulence can be change in viruses while circulating in the poultry population due to high rate of mutation (Stech et al., 1999). The H1N1 subtype is pathogenic swine viral that has been documented to cause an outbreak of respiratory disease in both human and swine. Influenza was first described as a disease of swine in 1918 (Koen, 1919). The first influenza A virus was isolated from swine in 1930 (Brockwell-Staats et al., 2009). A swine influenza virus was isolated from a human in 1974, from 1974 to 2005; there were 43 confirmed cases of transmission of influenza A virus from pigs to humans reported, with six fatalities (Brockwell-Staats et al., 2009). Since there are no unique clinical symptoms to differentiate swine influenza from seasonal influenza in humans, this number is probably having a small fraction of the actual cases. Most of these cases were the result of direct exposure to swine or were human-to-human transmission within a family cluster (Myers et al., 2007).

Swine play an important role in the ecology of influenza A viruses because they are susceptible to viruses of both the avian and mammalian lineages. The cells of the swine respiratory tract contain receptor sialyloligosaccharides possessing both N-acetylneuraminic acid-α2, 3- galactose, which is the preferred receptor for avian influenza viruses and N-acetylneuraminic acid-α2, 6-galactose, which is the preferred receptor for mammalian influenza viruses (Ito et al., 1998; Rogers et al., 1983). This has led to the proposal that swine serve as a “mixing vessel” for influenza viruses of different lineages, providing a place for reassortment and host adaptation to take place (Scholtissek, 1990).

Swine influenza A virus belong to the viral family of *Orthomyxoviridae*. They are RNA viruses with a segmented genome that is comprised of eight negative-sense, single-stranded RNA segments. These eight segments encode eleven proteins (Brockwell-Staats et al., 2009). The polymerase complex includes the PB2, PB1 and PA proteins as well as the nucleoprotein (NP). There are two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA) (Brockwell-Staats et al., 2009).

In view of these outbreaks that occurred due to mutations in influenza A virus as discuss in case of Bird Flu virus subtype H5N1 (Anwar et al., 2006). We initiated our in silico study to analyze the amino acid sequences of neuraminidase and hemagglutinin proteins of swine for the amino acid mutation at different positions compared to the other strains of the same sub-type (H1N1). The great genetic variability in influenza A virus lead to the difficulties in diagnosis, treatment, and prevention of influenza in humans. Therefore, it is significant to analyze these proteins for their mutation and phylogenetic analysis comparing with other strains of influenza virus.

Materials and Methods

The amino acid sequence of neuraminidase (NA) and hemag-

***Corresponding author:** Ajay Kumar, Department of Biotechnology, Institute of Biomedical Education and Research, Mangalayatan University, Aligarh (U.P), India, Tel: 91-09412883081; E-mail: akibmerbz@gmail.com

Received January 05, 2010; **Accepted** February 13, 2010; **Published** February 15, 2010

Citation: Sharma DK, Rawat AK, Srivastava S, Srivastava R, Kumar A (2010) Comparative Sequence Analysis on Different Strains of Swine Influenza Virus Sub-type H1N1 for Neuraminidase and Hemagglutinin. J Proteomics Bioinform 3: 055-060. doi:10.4172/jpb.1000121

Copyright: © 2010 Sharma DK, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

glutinin (HA) of swine influenza virus subtype H1N1 of A/Swine/Nebraska/(H1N1) were retrieved from protein sequence database situated at NCBI (http://www.ncbi.nlm.nih.gov/). These protein sequences have GenBank accession number ABR28650 for NA and AAB29091 for HA. Amino acid sequences of NA and HA of swine influenza virus sub-type H1N1 strains were used for screening of 98-99% similar sequences available in non-redundant (nr) database situated at NCBI using BLASTP (http://www.ncbi.nlm.nih.gov/BLAST/).

The protein sequences of selected strains (Table 1) was then aligned by using multiple sequence alignment tool ClustalW 1.83 (http://www.ebi.ac.uk/tools/clustalw2/index.html) situated at European Molecular Biology Laboratory. ClustalW (Thompson et al., 1994) program calculate the best match for the selected sequences and line them up, so that the identities, similarities and differences can be seen. These alignments were then analyzed for differences in their amino acid at specific positions. A program PROTMUTATION (http://www.biobrainz.com/tools/protmutation.htm) was developed in Perl CGI programming. It is an interactive, user-friendly program for identifying mutations by comparing protein sequences of two different strains using Needleman-Wunsch algorithm for global sequence alignment to point out mutations in different strains. Hydrophobicity values

Proteins	GenBank Accession No. Accession No.	Strains
Neuraminidase	ABR15833	A/swine/Tennessee/19/1977(H1N1)
	ABY51218	A/swine/Tennessee/7/1978(H1N1)
	ABS49924	A/swine/Iowa/1/1977(H1N1)
	ABR28617	A/swine/Kentucky/1/1976(H1N1)
	ABV29527	A/swine/Tennessee/37/1977(H1N1)
Hemagglutinin	ABW71503	A/swine/Tennessee/2/1978(H1N1)
	ABY51204	A/swine/Tennessee/4/1978(H1N1)
	ABR28581	A/swine/Tennessee/82/1977(H1N1)
	ABW86574	A/swine/Tennessee/5/1978(H1N1)
	ABW86585	A/swine/Tennessee/8/1978(H1N1)
	ABR28735	A/swine/Wisconsin/641/1980(H1N1)
	ABR28757	A/swine/Wisconsin/8/1980(H1N1)
	ABU80276	A/swine/Wisconsin/663/1980(H1N1)
	ABR15852	A/swine/Tennessee/31/1977(H1N1)
	ABR28713	A/swine/Wisconsin/11/1980(H1N1)
	ABR28625	A/swine/Minnesota/27/1976(H1N1)
	ABR29605	A/swine/Iowa/3/1985(H1N1)
	BAH02160	A/swine/Niigata/1/1977(H1N1)
	ABR15819	A/swine/Iowa/4/1976(H1N1)
	ABR28537	A/swine/Tennessee/10/1977(H1N1)
	ABR28603	A/swine/Illinois/1/1975(H1N1)
	ABW36333	A/swine/Ontario/2/1981(H1N1)
	ABR28702	A/swine/Wisconsin/1/1971(H1N1)
	ABS49921	A/swine/Iowa/1/1977(H1N1)
	ABX58646	A/swine/Iowa/2/1987(H1N1)
	ABU80232	A/swine/Tennessee/86/1977(H1N1)
	ABR28669	A/swine/Ontario/7/1981(H1N1)
	ABX58657	A/swine/Tennessee/3/1978(H1N1)
	ABR28614	A/swine/Kentucky/1/1976(H1N1)
	ABR28658	A/swine/Ontario/4/1981(H1N1)
	ABR28636	A/swine/Minnesota/5892-7/1979(H1N1)
	ABW36344	A/swine/Ontario/3/1981(H1N1)
	ABQ45436	A/swine/Tennessee/15/1976(H1N1)
	ABR15863	A/swine/Tennessee/49/1977(H1N1)
	ABU80410	A/swine/Tennessee/48/1977(H1N1)
	ABS49932	A/swine/Ontario/1/1981(H1N1)

Table 1: Selected strains of NA and HA, which are screened using BLASTP containing GenBank accession number.

were obtained from the tool ProtScale (Gasteiger et al., 2005) at ExPASy server (http://www.expasy.org/tools/protscale.html) choosing Kyte & Doolittle hydrophobicity scale (Kyte and Doolittle, 1982). An evolutionary distance matrix was generated from multiple sequence alignment of selected homologous sequences and phylogenetic tree was then drawn using the Neighbour joining method (Saitou and Nei, 1987) by MEGA (Molecular Evolutionary Genetics analysis) version 4.0 (Tamura et al., 2007). Secondary structures of the proteins were predicted using program SOPMA (Geourjon and Deléage, 1995). Domains or motifs among these protein sequences of A/Swine/Nebraska/(H1N1) were searched using ScanProsite at Expasy Server (http://www.expasy.ch/tools/scanprosite/). Motifs with high probability of occurrence were also included in the search.

Results and Discussion

Sequence analysis

On comparing our sequence of neuraminidase after multiple alignment with the selected sequences of the same subtype having 98-99% similarity, it was found that at position 240 of the sequence the hydrophilic threonine was replaced by hydrophobic alanine and at position 451, hydrophobic glycine was replaced by hydrophilic serine (Table 2). The mutation at position 240 may be important in the sense that here hydrophilic Threonine is being replaced by hydrophobic Alanine, which may help the protein to attain more stable conformation, while at position 451 hydrophobic Glycine is being replaced by hydrophilic Serine, which may help the protein to attain less stable conformation, since glycine is the smallest amino acid, it fits into tight places inside a folded protein and disrupts α-helix formation. In hemagglutinin, mutations were found at position 225, where hydrophilic arginine was substituted by same type of amino acid hydrophilic Lysine and position 283, where more hydrophobic isoleucine was replaced by less hydrophobic valine.

The program PROTMUTATION reports all the mutations along with their specific position in the sequence. PROTMUTATION program accepts input sequences by pasting two sequences in raw format in the corresponding text boxes as shown in Figure 1.

Output of PROTMUTATION produces result instantly in a file containing global alignment of both inputted protein se-

Base	Hatay/2004	Change in properties	Sec. structure	Hydrophobicity (Kyte
Neuraminidase				
240	T→A	Hydrophilic, Polar →Hydrophobic	Strand	0.622 → 0.900
451	G →S	Hydrophobic → Hydrophilic, Polar	Coil	-0.167 → -0.211
Hemagglutinin				
225	R →K	Hydrophilic →Hydrophilic	Coil	-2.289 → -2.222
283	I →V	Hydrophobic →Hydrophobic	Strand	1.044 → 1.011

Table 2: Amino acid mutations in neuraminidase and hemagglutinin at specific position and change in properties, secondary structure and hydrophobicity.

BIOBRAINZ
Life Sciences

PROTMUTATION
A Protein Mutation Finder

Please Enter Protein Sequence 1 (Plain Text Format)

Please Enter Protein Sequence 2 (Plain Text Format)

Submit Reset

Figure 1: Snapshot of PROTMUTATION Tool.

quences. The results of PROTMUTATION were checked with results produced by manually analyzing mutations after multiple alignments and with known one tool, the results were similar in all cases, thus, it was predicted that the program PROTMUTATION produces more than 90% accurate results. This program will be greatly helpful to swine flu researchers who are interested in finding out the rate of mutation in Influenza viruses, as the virus continuously undergoes antigenic shift and antigenic drift.

Phylogenetic analysis

Phylogenetic trees of NA and HA sequences with selected strains (Table 1) from swine were shown in Figure 2 and Figure 3. It is evident from phylogenetic analysis of protein sequences that isolates were having close relationship.

Analysis of the sequence comparison shows that mutations were observed in HA (2sites) and NA (2 sites). Thus, we can say that HA and NA of swine influenza virus sub-type H1N1 are

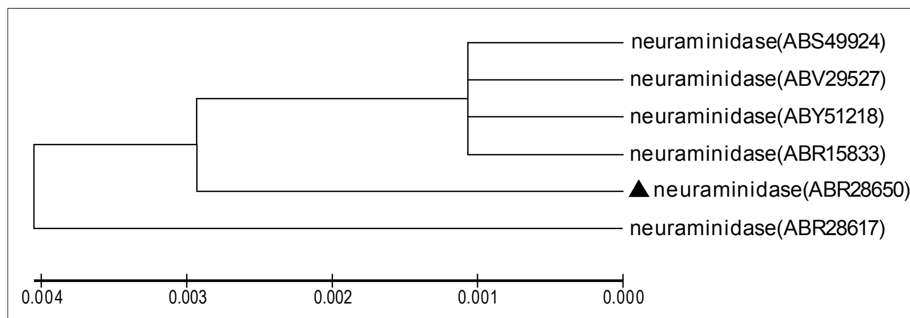

Figure 2: Neighbour-joining tree of test NA sequence with selected strains obtained from BLASTP search using MEGA (Molecular Evolutionary Genetics Analysis) version 4.0 software. It shows relationship between test NA sequence with closely related representative strains.

Secondary structure and hydrophobicity prediction

In order to better understand what sort of changes this mutation possibly might have on the A/Swine/Nebraska/(H1N1) strain, secondary structure and hydrophobicity in this mutation site are examined. The results of secondary structure prediction of neuraminidase and hemagglutinin amino acids are situated at http://www.biobrainz.com/tools/secstr.htm. Secondary structure prediction of NA of A/Swine/Nebraska/(H1N1) strain and all the other strains show that no structure changes were detected at position 240 and 451 and are strand and coil respectively. For HA, no secondary structure change is detected at position 225 and 283 of A/Swine/Nebraska/(H1N1) and all the other strains and are found coil and strand respectively. Hydrophobicity is evaluated using ProtScale software. This software displays the polarity of the initial amino acids and amino acids after the mutation presented in Table 2.

Domain/motif search

Different domains that were found in the NA and HA proteins are given in the Tables 3 and 4. While in NA, mutations (T240A, G451S) were found at predicted functional sites of N-myristoylation GScfAI and GVnsST at positions 236 - 241 and 447-452 respectively was shown in Table 3 (Prosite Documentation, http://www.biobrainz.com/tools/pro.htm). The signature for N-myristoylation sites is G - {EDRKHPFYW} - x (2) - [STAGCN] - {P}, at position 5, small-uncharged residues (Ala, Ser, Thr, Cys, Asn and Gly) are allowed and serine is favored (Towler et al., 1988; Grand, 1989). The mutation from T to A at position 240 and G to S at position 451 do not account for any difference in N-myristoylation site. For HA, mutation (I283V) was found at a predicted functional site (GIviSD) of N-myristoylation at position 281 – 286 was shown in table 4 (Prosite Documentation, http://www.biobrainz.com/tools/pro.htm). In the signature sequence of N-myristoylation site at position 3 and 4, most, if not all, residues are allowed (Towler et al., 1988; Grand, 1989). Therefore mutation from I to V at position 283 does not account for any difference in N-myristoylation site.

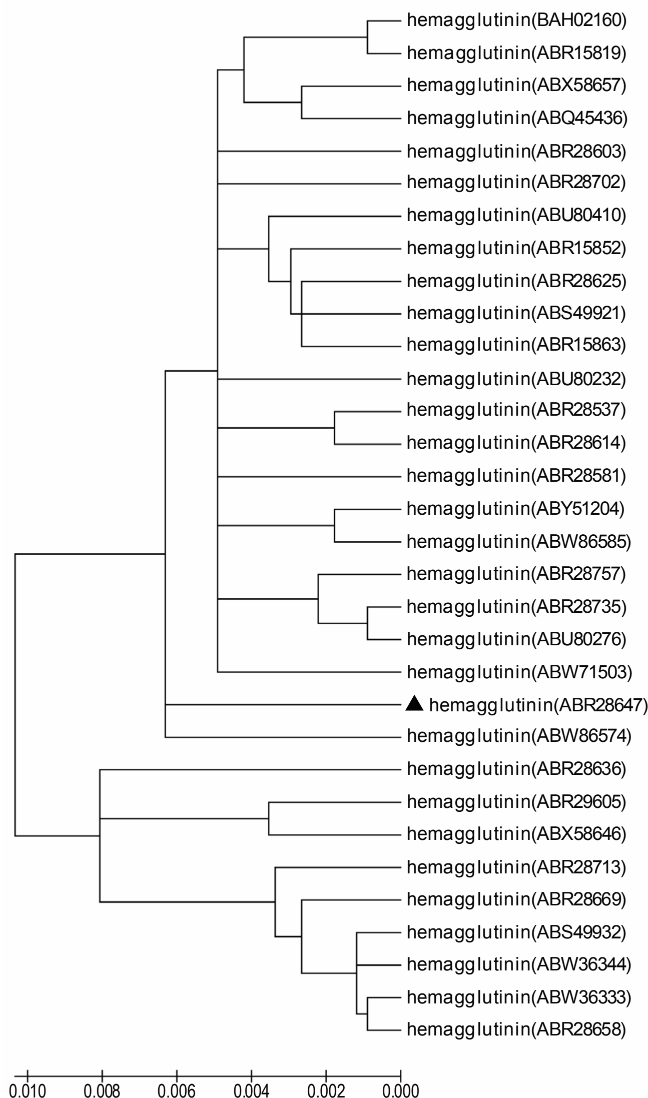

Figure 3: Neighbour-joining tree of test HA sequence with selected strains obtained from BLASTP search using MEGA (Molecular Evolutionary Genetics Analysis) version 4.0 software. It shows relationship between test HA sequence with closely related representative strains.

sensitive towards mutations. Then we find that these mutations involve in change polarity or hydrophobicity. Furthermore, not only the polarity or hydrophobicity is significantly altered by most mutations but also the propensity of each amino acid residue to stabilize the secondary structure. In this work, no structural conformation changes were observed in HA (at position 225 and 283) and in NA (at position 240 and 451). All the important domains of NA and HA protein sequences were tracked (Table 3 and Table 4). In NA and HA mutations were found at predicted functional sites of N-myristoylation. These mutation were not affecting N-myristoylation according to our in silico study but in future if mutation occurs at this position the virus may become more lethal to humans.

Neuraminidase		
Site	**Position**	**Domain**
N-myristoylation site	27 - 32	GNiiSL
	137 - 142	GAlIND
	236 - 241	GScfAI
	356 - 361	GVwiGR
	440 - 445	GSsiSF
	447 - 452	GVnsST
Casein kinase II phosphorylation site	44 - 47	ShpE
	110 - 113	SkgD
	125 - 128	ShlE
	148 - 151	TvkD
	172 - 175	SrfE
	196 - 199	SgpD
	369 - 372	SgfE
	381 - 384	TetD
	456 - 459	SwpD
N-glycosylation site	50 - 53	NQSV
	58 - 61	NNTW
	63 - 66	NQTY
Protein kinase C phosphorylation site	105 - 107	SiR
	148 - 150	TvK
	215 - 217	TiK
	218 - 220	SwR
	252 - 254	SyK
	285 - 287	TgK
	299 - 301	SnR
	350 - 352	SfR
	366 - 368	SsR
	388 - 390	SmK

Table 3: Predicted domains/motifs in neuraminidase representing the functional site name, its position on the sequence and the sequence of the site using prosite tool situated at expasy server.

Hemagglutinin		
Site	**Position**	**Domain**
N-glycosylation site	27 – 30	NNST
	28 – 31	NSTD
	40 – 43	NVTV
	104 – 107	NGTC
	293 – 296	NTTC
	304 – 307	NTSL
	498 – 501	NGTY
	557 – 560	NGSL
Casein kinase II phosphorylation site	35 – 38	TvlE
	100 – 103	SnsD
	126 - 129	SsfE
	201 – 204	TstD
	249 – 252	TliE
	316 – 319	TigE
	398 – 401	SviE
	491 – 494	TcmE
N-myristoylation site	80 – 85	GNpeCE
	148 – 153	GVtaAC
	277 – 282	GSgsGI
	281 – 286	GIviSD
	301 – 306	GAinTS
	345 – 350	GLfgAI
	348 – 353	GAiaGF
	356 – 361	GGwtGM
	360 – 365	GMidGW
Protein kinase C phosphorylation site	145-147	Tnr
	220-222	SsK
	298 – 300	TpK
	326 – 328	StK
	393 – 395	TnK
	495 – 497	SvK
	524 – 526	StR

Table 4: Predicted domains/motifs in hemagglutinin representing the functional site name, its position on the sequence and the sequence of the site using Prosite tool situated at expasy server.

Although we have not yet been predicted any mutation that may lead to an outbreak of swine flu rather we can in principle monitor the mutations along the time course, and predict the trend of mutations. Thus, further mutational analysis would have to be carried out to map a specific amino acid change in a protein causing the change in pathogenicity of the virus.

Conclusions

Our study revealed that HA and NA are prone to mutations, thus we conclude that HA and NA might be an important proteins involved in the pathogenesis of swine influenza virus. We also found that these mutations involve change in polarity or hydrophobicity. Furthermore, it is not only the polarity or hydrophobicity significantly altered by most mutations but also the propensity of each amino acid residue to stabilize the secondary structure. Mutations at N-myristoylation site in NA (T240A, G451S) and HA (I283V) do not make any difference but if another mutation occurs at this point then it might be fatal. Secondary structure prediction and prosite documentation files of these proteins are available at http://www.biobrainz.com/tools/secstr.htm and http://www.biobrainz.com/tools/pro.htm. The program PROTMUTATION available at http://www.biobrainz.com/tools/protmutation.htm can be used to predict mutations in new strains.

Acknowledgements

The authors are grateful to Prof. Ashok Kumar (Dean, I.B.M.E.R, Mangalayatan University Aligarh, U.P, India) for providing necessary facilites and encouragement. The authors are also thankful to all faculty members of the Institute of Biomedical Education and Research, Mangalayatan University Aligarh, U.P, India for their generous help and suggestions during the course of experimental work and manuscript preparation.

References

1. Anwar T, Lal SK, Khan AU (2006) In silico analysis of genes nucleoprotein, neuraminidase and hemagglutinin: a comparative study on different strains of influenza A (Bird flu) virus sub-type H5N1. In Silico Biol 6: 161-8. » CrossRef » PubMed » Google Scholar

2. Brockwell-Staats C, Webster RG, Webby RJ (2009) Diversity of Influenza Viruses in Swine and the Emergence of a Novel Human Pandemic Influenza A (H1N1). Influenza Other Respi Viruses 3: 207-213. » CrossRef » PubMed » Google Scholar

3. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook. Humana Press pp571-607. » CrossRef » PubMed » Google Scholar

4. Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11: 681-4. » CrossRef » PubMed » Google Scholar

5. Grand RJA (1989) Acylation of viral and eukaryotic proteins. Biochem J 258: 625-638. » CrossRef » PubMed » Google Scholar

6. Ito T, Couceiro JN, Kelm S, Baum LG, Krauss S, et al. (1998) Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. J Virol 72: 7367-7373. » CrossRef » PubMed » Google Scholar

7. Koen JS (1919) A practical method for field diagnosis of swine diseases. Am J Vet Med 14: 468-470. » CrossRef » PubMed » Google Scholar

8. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105-32. » CrossRef » PubMed » Google Scholar

9. Myers KP, Olsen CW, Gray GC (2007) Cases of swine influenza in humans: a review of the literature. Clin Infect Dis 44: 1084-1088. » CrossRef » PubMed » Google Scholar

10. Rogers GN, Paulson JC (1983) Receptor determinants of human and animal influenza virus isolates:differences in receptor specificity of the H3 hemagglutinin based on species of origin. Virology 127: 361-373. » CrossRef » PubMed » Google Scholar

11. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-25. » CrossRef » PubMed » Google Scholar

12. Scholtissek C (1990) Pigs as the “mixing vessel” for the creation of new pandemic influenza A viruses. Med Principles Pract 2: 65-71. » CrossRef » PubMed » Google Scholar

13. Stech J, Xiong X, Scholtissek C, Webster RG (1999) Independence of evolutionary and mutational rates after transmission of avian influenza viruses to swine. J Virol 73: 1878-84. » CrossRef » PubMed » Google Scholar

14. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-9. » CrossRef » PubMed » Google Scholar

15. 16. Towler DA, Gordon JI, Adams SP, Glaser L (1988) The biology and enzymology of eukaryotic protein acylation. Annu Rev Biochem 57: 69-99. » CrossRef » PubMed » Google Scholar

16. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-80. » CrossRef » PubMed » Google Scholar

17. World Health Organization (2010) Influenza A (H1N1) Update 50. Available at http://www.who.int/csr/don/2010_02_5/en/index.html. » CrossRef » PubMed » Google Scholar