

# Sequence Relationships Similarity of Swine Influenza a (H1N1) Virus

Patsaraporn Somboonsak, Mud-Armeen Munlin

**Abstract**—In April 2009, a new variant of Influenza A virus subtype H1N1 emerged in Mexico and spread all over the world. The influenza has three subtypes in human (H1N1, H1N2 and H3N2) Types B and C influenza tend to be associated with local or regional epidemics. Preliminary genetic characterization of the influenza viruses has identified them as swine influenza A (H1N1) viruses. Nucleotide sequence analysis of the Haemagglutinin (HA) and Neuraminidase (NA) are similar to each other and the majority of their genes of swine influenza viruses, two genes coding for the neuraminidase (NA) and matrix (M) proteins are similar to corresponding genes of swine influenza. Sequence similarity between the 2009 A (H1N1) virus and its nearest relatives indicates that its gene segments have been circulating undetected for an extended period. Nucleic acid sequence Maximum Likelihood (MCL) and DNA Empirical base frequencies, Phylogenetic relationship amongst the HA genes of H1N1 virus isolated in Genbank having high nucleotide sequence homology.

In this paper we used 16 HA nucleotide sequences from NCBI for computing sequence relationships similarity of swine influenza A virus using the following method MCL the result is 28%, 36.64% for Optimal tree with the sum of branch length, 35.62% for Interior branch phylogeny Neighbor – Join Tree, 1.85% for the overall transition/transversion, and 8.28% for Overall mean distance.

**Keywords**—Sequence DNA, Relationship of swine, Swine influenza, Sequence Similarity

## I. INTRODUCTION

SEQUENCE alignment is the most basic analysis used in the comparative study of molecular sequence (nucleic acid and proteins). The 2009 Swine influenza A (H1N1) is a new influenza virus causing illness in people. Pandemic H1N1 2009 viruses first detected in people in the United States in April 2009 and commonly called “Swine flu” because laboratory testing showed that many of the genes in this new virus were very similar to influenza viruses that normally occur in pigs (swine) in North America. But further study has shown that this new virus is very different from what normally circulates in North American pigs. It has two genes from flu viruses that normally circulate in pigs in Europe and Asia and bird (avian) genes and human genes.

P. S. Author is with the Faculty of Information Science and Technology Mahanakorn University of Technology, Nongchock, 10530 Bangkok Thailand phone: 086-034-7785; fax: 02-915-9628; e-mail: psomboonsak@yahoo.com).

M. M. Author, is with the Faculty of Information Science and Technology, Mahanakorn University of Technology, Nongchock, 10350 Thailand (e-mail: mmunlin@gmail.com).

Influenza viruses can be passed from human to pig, and from pig to human. This appears to be a particular characteristic in pig-human inter species transmission of influenza A. Human infection with flu viruses from pigs are most likely to occur when people are in close proximity to infected pigs, such as in pig barns and livestock exhibits housing pigs at fairs. Recent reports of widespread transmission of swine-origin influenza probably in much the same way that regular seasonal influenza A (H1N1) viruses among human population in United States and elsewhere highlight this ever-present threat to global public health. On June 11, 2009, the World Health Organization (WHO) signaled that a pandemic of 2009 H1N1 flu was underway. A few days later, the Centers for Disease Control and Prevention in the United States confirmed that these human influenza cases were caused by the same new influenza A (H1N1) virus.

In veterinary diagnostic laboratories, the detection of type A swine influenza virus infection has been routinely carried out by virus isolation in embryonated chicken eggs or Madin–Darby canine kidney (MDCK) cells with subsequent subtype determination by hemagglutinin and neuraminidase inhibition tests using mono specific antiserum to each subtype by Webster [14]. The virus was isolated from these samples using MDCK cells as described previously Meguro et al. [15]. Detection and subtyping of swine influenza H1N1, H1N2 and H3N2 viruses in clinical samples using two multiplex RT-PCR assays Y.K. Choi et al. [16]. A rapid detection and subtyping method is necessary to obtain detailed information on the prevalence of different subtypes of influenza A virus and to establish effective control measures for the swine industry. Identification and subtyping are also important for tracking prevalent strains in a region of the country. Because swine are often viewed as ‘mixing vessels’ for both avian and human subtypes of influenza virus Eric C.J. Claas [17] Brockwell et al. [18]. Detection of influenza A (H1N1) virus by real-time RT-PCR M Panning et al. [19]. A novel real-time RT-PCR for influenza A (H1N1) virus was set up ad hoc and validated following industry-standard criteria. The lower limit of detection of the assay was 384 copies of viral RNA per ml of viral transport medium (95% confidence interval: 273-876RNA copies/ml). Specificity was 100% as assessed on a panel of reference samples including seasonal human influenza A virus H1N1 and H3N2, highly pathogenic avian influenza A virus H5N1 and porcine influenza A virus H1N1, H1N2 and H3N2 samples.

Recently, there have been several methods for analysis DNA H1N1 viruses, from both Biology and Computer

science, such as Artificial Intelligence methods, Data mining methods. In this paper we used the current data of DNA H1N1 form NCBI 2009 analysis with data mining methods. Actually, sequence analysis is still an approach and many methodologies have been proposed. The methods most frequently used for DNA analyses and prediction region are based on machine learning methods, such as neural networks, decision trees and others.

## II. LITERATURE REVIEW

The alignment of molecular sequence was first described by Saul B. Needleman, Christlan D. Wunsch [1], their algorithm performs a global alignment on two sequences for the case when the alignment is penalized solely by the matches and mismatches, and gaps have no penalty. The corresponding dynamic programming algorithm takes cubic time. To find the alignment with the highest score, a two-dimensional array  $F$  is allocated. A better dynamic programming algorithm with quadratic running time for same problem was first introduced by David Sankoff [2]. Smith and Waterman determining similar regions between two nucleotide and compares segments of all possible lengths with calculate optimize the similarity measure. The algorithm is a general local alignment method also based on dynamic programming. With sufficiently similarity sequences, there is no difference between local and local alignments. McClure et al. [3] Since then the theory and art of sequence alignment reconstruction has witnessed a proliferation of alignment algorithms aiming at improving computational feasibility and performance, on the one hand, and the biological relevance and quality of the deduced alignments, on the other (for reviews, see McClure et al. [3] Hirose et al. [4] Ramana M. Idury, Michael S. Waterman [5] Dan Gusfield et al. [6] Thompson et al. [7] Edgar et al. [8] Multiple methods offer significantly better alignment quality and reduced computation cost. A benefit of this approach is that it permits the rapid alignment of even hundreds of sequences. A major limitation is that the final alignment depends on the order in which sequence are joined. Thus, it is not guaranteed to provide the most accurate alignments. DNA sequences in comparative work Saiki et al. [9] Kocher et al. [10]. Thompson et al. [7], there are access the program with ClustalW algorithm for performing progressive multiple sequence alignment. Prior to the 1960, most systematic studies utilized morphological character as evidence for relationships. In biology, phylogenetic is the study of evolutionary relatedness among various groups which is discovered through molecular sequencing data and matrices. The research have increased Pietro and Nick Goldman [11] Erpenbeck et al. [12] Sudhir Kumar et al. [13], there are access the program for comparative analysis of homologous nucleotide sequence either from multi gene families or from different species with relationships and patterns of nucleotide and protein evolution (Mega 4.1). The application used Neighbor-Joining (NJ) method to infer phylogenetic tree inference.

## III. MATERIALS AND METHODS

Influenza viruses used in this study are listed in Table1, all complete nucleotide sequence of influenza A virus data were collected as part of the influenza nucleotide sequencing for the period 2009-2010. The NCBI Influenza virus Sequence Database was used nucleotide sequences to analyze the genetic evolution of the new influenza A (H1N1) virus. All sequence data were downloaded from the NCBI.

We used nucleotide sequences to analyses the relationships of the influenza A virus. This set of nucleotide sequences included hemagglutinin (HA). All alignment showed in fig. 1 and fig. 2.

Sequence relationship using NJ method generated a tree in MEGA, researcher used sixteen H1N1 viruses of swine and in this study are listed Phylogenetic tree topologies related the homologies for influenza viruses that have been circulating.

### A. Sequence similarity analysis of the HA gene of influenza A (H1N1)

Sequence similarity is the fraction of aligned positions in a sequence alignment at with identical sequence characters or conservative substitutions are located. Positions with gap are usually not scored. A DNA sequence consists of four DNA base whose character code are: A, G, C and T nucleotide. Sequences generate alignment using a built-in CLUSTALW implementation for the complete sequence or data in any rectangular region. Genetic Analysis software used MEGA4.1 MEGA, version 4.1, Tamura et al. [20]. NJ method as the use of the Maximum Composite Likelihood method (MCL) distance leads to a much higher accuracy.

### B. Phylogenetic analysis of the HA gene of influenza A (H1N1)

Phylogenetic analysis is a powerful tool to study the relationships among sequences. Form such relationships the origins. The first step in detailed Phylogenetic analysis of nucleotide sequences is the alignment of the sequences and comparison of homologous character shared between organisms Fig. 5. As the secondary comparing the sequence structure models with calculate optimize sequence homologies and region.

TABLE I  
ACCESSION INFLUENZA A VIRUS

	Viruses H1N1	Region	Gene	GenBank No.
1	Influenza A virus A/Australia/1/2009	Australia	HA	CY055526
2	Influenza A virus A/San Diego/INS42/2009	San Diego	HA	CY056172
3	Influenza A virus A/Mexico City/001/2009	Mexico	HA	CY050198
4	Influenza A virus A/Thailand/CU-B5/2009	Thailand	HA	GQ866951
5	Influenza A virus A/Wisconsin/629-D00859/2009	Wisconsin	HA	CY063299
6	Influenza A virus A/California/VRDL14/2010	California	HA	CY063211
7	Influenza A virus A/Athens/INS163/2009	Athens	HA	CY062891
8	Influenza A virus A/Madrid/INS186/2009	Madrid	HA	CY063035
9	Influenza A virus	Vienna	HA	CY062987



TABLE IV  
MAXIMUM COMPOSITE LIKELIHOOD ESTIMATE

	A	T	C	G
A	-	3.33	2.83	<b>17.2</b>
T	<i>4.99</i>	-	<b>14.74</b>	3.06
C	<i>4.99</i>	<b>14.26</b>	-	3.06
G	<b>28</b>	3.33	2.83	-

Each entry shows the probability of substitution from one base (row) to another base (column) instantaneously. Only entries within a row should be compared. Rates of different transitional substitutions are shown in bold and those of transversion substitutions are shown in italics.

The nucleotide frequencies are 0.35 (A), 0.234 (T/U), 0.19 (C), and 0.216 (G). The transition/ transversion rate ratios are  $k_1 = 5.61$  (purines) and  $k_2 = 4.28$  (pyrimidines). The overall transition/transversion bias is 1.85.

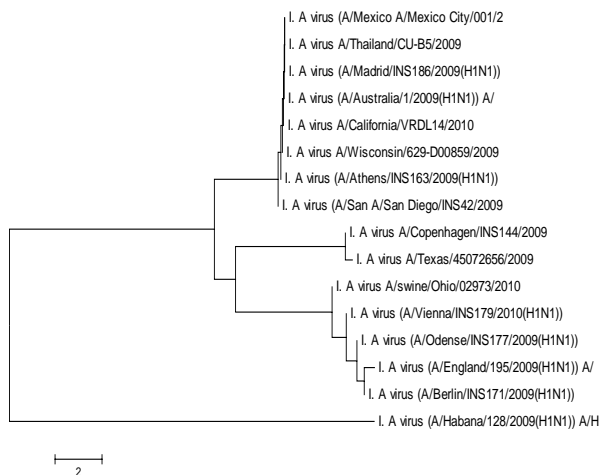


Fig. 4 Displaying a neighbor joining (NJ) tree of H1N1 viruses

Sequence relationships among the 16 HA nucleotides. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length is 36.64%. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the MCL in TABLE IV and are in the units of the number of base substitutions per site in TABLE III and The summary sequence relationships each methods shown in TABLE V.

TABLE V  
SUMMARY SEQUENCE RELATIONSHIPS EACH METHODS

Methods test	Result (%)
Neighbor-Joining(NJ)	36.64
Phylogenetic relationships	35.59
Overall mean distance	8.28
Optimal tree of branch length	36.64
Transition/Transversion	1.85
Maximum composite likelihood estimate of the pattern of nucleotide substitution(MCL).	28.00

Phylogenetic relationships amongst the HA genes of Habana/HM176611 isolates. The evolutionary history was inferred using the Minimum evolution method. The optimal tree with the sum of branch length is 35.59%. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the MCL in Fig. 5 and are in the units of the number of base substitutions per site. The NJ algorithm was used to generate the initial tree.

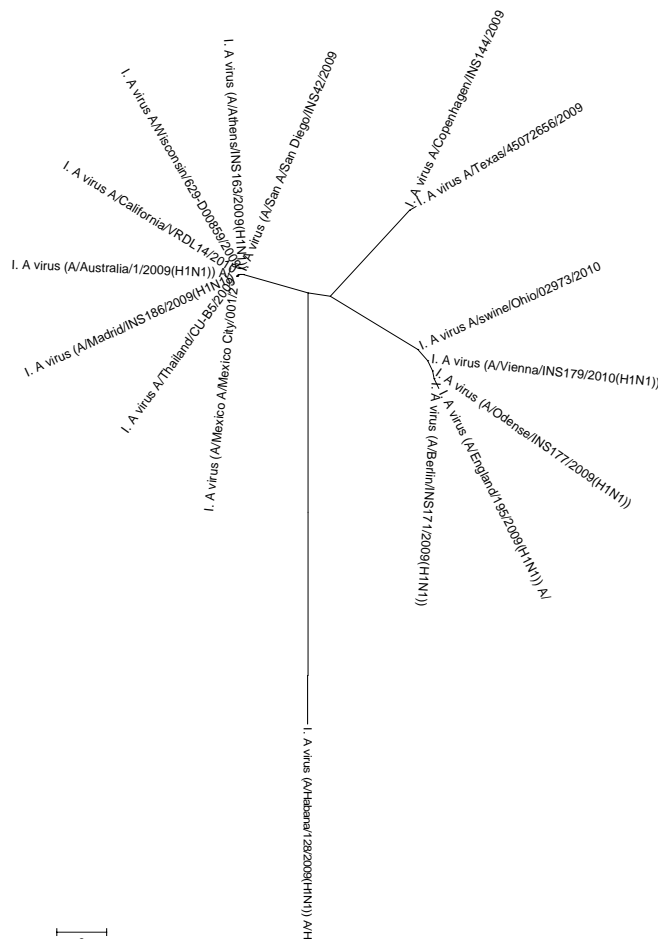


Fig. 5 Phylogenetic relationships

V.CONCLUSION

We can conclusion from this study that for H1N1 nucleotide isolated with tools. In this paper we calculate the sequence similarity by NJ and phylogenetic tree. We have analyzed the new nucleotide sequence HA of swine influenza virus present sequence similarity nucleotide. Our findings allow the relationships of the influenza A viruses into country clusters among the currently circulating influenza A viruses.

## VI. FUTURE WORKS

We will find the method with used partially gene to compute the good result in order to speedup a computer. Since Mega4 has some problem sometime inaccuracy, it takes a long time to calculate. To solve this, we plan to cluster DNAs before applying the tools. The predictive aim the source country of the disease and their similarity.

## ACKNOWLEDGMENT

This conference was financially supported by Mahanakorn University of Technology. The authors wish to thank Associate Prof. Dr. Thanaruk Theeramunkong for excellent technical assistance.

## REFERENCES

- [1] Saul B. Needleman, Christlan D. Wunsch, "A General Applicable to the Search for Similarities in the Amino Acid Sequence of two Proteins", *J Mol. Biol.*, Vol. 48, 1970, pp. 443-453.
- [2] David Sankoff, "Simultaneous solution of the RNA folding alignment and protosequence problems", *Siam J. Appl Math.*, Vol. 45, 1985, pp. 810-825.
- [3] McClure MA, Vasi TK, Fitch WM, "Comparative analysis of multiple protein sequence alignment methods", *Mol Biol Evol.*, Vol. 4, 1994, pp. 571-592.
- [4] Makoto Hirosawa, Yasushi Totoki, Masaki Hoshida, Masaro Ishikawa, "Comprehensive study on iterative algorithms of multiple sequence alignment", *Comput. Appl. Biosci.*, Vol. 11, 1995, pp. 13-18.
- [5] Ramana M. Idury, Michael S. Waterman, "A New Algorithm for DNA Sequence Assembly", *Computer Biology*, Vol. 2, 1995, pp. 291-306.
- [6] Dan Gusfield, Jens Stoye, "Linear time algorithms for finding and representing all the tandem repeats in a string", *Computer and Stem Sciences*, Vol. 69, 2004, pp. 525-546.
- [7] Thompson JD, Higgins DG, Gibson TJ, "CLUSTALW improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice", *NucleicAcids Res.*, Vol. 22, 1994, pp. 4673-80.
- [8] Edgar, Batzoglou, "Multiple sequence alignment", *Curr Opin Struct Biol.*, Vol. 3, 2006, pp. 368-373.
- [9] Saiki, Gelfand, Stoffel, Scharf, Higuchi, Horn, Hullis, Erlich, "Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase", *Science*, Vol. 239, 1988, pp. 487-491.
- [10] Kocher T., Thomas K., Meyer, Edwards V., PÅÅbo, Villablanca X., Wilson, "Dynamics of mitochondrial DNA evolution in animals Amplification and sequencing with conserved primers", *Proc. Natl. Acad. Sci. USA*, Vol. 86, 1989, pp. 6196-6200.
- [11] Pietro Liò, Nick Goldman, "Models of Molecular Evolution and Phylogeny", *Genome Res.*, Vol. 8, 1998, pp. 1233-1244.
- [12] Erpenbeck, Breeuwer, Soest, "Implications from a 28S rRNA gene fragment for the phylogenetic relationships of halichondrid sponges", *Zoological Systematics and Evolutionary Research*, Vol. 43, 2005, pp. 93-99.
- [13] Sudhir Kumar, Masatoshi Nei, Joel Dudley, Koichiro Tamura, "MEGA A biologist-centric software for evolutionary analysis of DNA and protein sequences", *Bioinformatics*, Vol. 9, 2008, pp. 299-306.
- [14] Robert G. Webster, "Influenza AN Emerging Disease", *Emerging infectious diseases*, Vol. 4, 1998, pp. 436-441.
- [15] Hidenori Meguro, David Bryant, Anne E. Torrence, Peter F. Wright, "Canine Kidney Cell Line for Isolation of Respiratory Viruses", *Clinical Microbiology*, Vol. 9, 1997, pp. 175-179.
- [16] Y.K.Choi, Goyal, Kang, Farnham, Joo, "Detection and subtyping of swine influenza H1N1 H1N2 and H3N2 viruses in clinical samples using two multiplex RT-PCR assays", *Virological Methods*, Vol. 102, 2002, pp. 53-59.
- [17] Eric C.J. Claas, "Pandemic influenza is a zoonosis as it requires introduction of avian-like gene segments in the human population", *Veterinary microbiology*, Vol. 74, 2000, pp. 133-139.
- [18] Brockwell Stats C, Webster RG, Webby RJ, "Diversity of InfluenzaViruses in Swine and the Emergence of a Novel Human Pandemic InfluenzaA (H1N1)", *Influenza Other Respi Viruses*, Vol. 3, 2009, pp. 207-213.
- [19] M Panning, M Eickmann, O Landt, M Monazahian, S Ölschläger, S Baumgarte, U Reischl, J J Wenzel, H H Niller, S GÜnther, B Hollmann, D Huzly, J F Drexler, A Helmer, S Becker, B Matz, A M Eis-Hübinger, C Drosten, "Detection of influenza A(H1N1) virus by real-time RT-PCR", *Eurosurveillance*, Vol. 14, 2009, pp. 1-6.
- [20] Tamura K, Dudley J, Nei M, Kumar S, "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0", *Molecular Biology and Evolution*, Vol. 24, 2007, pp. 1596-1599.