Large scale multiple sequence alignment of intraspecies samples: conserved sequence analysis of influenza A virus HA segment and its application in rapid typing

Qianyu Lin[1] , Xiang Ji[2], Feng Wu[2], Lan Ma[1,2,3*]

[1]Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, China

[2]Tsinghua Shenzhen International Graduate School, Tsinghua University, China

[3]Shenzhen Bay Laboratory, Shenzhen, China

*** Correspondence:**
Lan Ma
malan@sz.tsinghua.edu.cn

Keywords: Influenza$_1$, Multiple Sequence Alignment$_2$, Conserved sequence$_3$, rapid detection$_4$.

Abstract

Background The high mutation rate of influenza A virus hemagglutinin segment brings great challenges to its long-term effective testing and subtyping.

Method We analyzed the conserved sequences of hemagglutinin subtype H1-H9 by breadth first, and designed primers for HA subtyping based on conserved sequences.

Results Our conserved sequence searching method get high specificity conserved sequences on H1-H9 subtypes respectively. And PCR experiments show that primers based on conserved sequences can be used in influenza A virus HA subtyping.

Conclusions Conserved sequences based primers are expected to be the long-term effective influenza A virus HA subtyping tools.

1 Introduction

Influenza is a common widespread infectious disease that has caused huge human deaths in the past century(1). The annual influenza epidemic makes 3 to 5 million cases of severe illness and 0.29 to 0.65 million respiratory deaths(2). Influenza A (short in “influenza” in following text) virus is RNA virus. The surface antigens of

influenza are hemagglutinin (HA) and neuraminidase (NA), 18 different HA and 11 different NA are known so far. Influenza's subtype is defined according to HA and NA subtypes. The influenza virus invade and infect the cells via HA's binding to cells' sialic acid-containing receptors(3). Specific binding with hemagglutinin to reduce the infectious capacity of influenza virus is the main idea of existing treatment and prevention of influenza diseases. Thus, fast and accurate identification of HA subtype is important for influenza's diagnosing and treatment.

However, the high mutation rate of influenza virus results in a large number of single-nucleotide polymorphism (SNP) and antigen drift(4, 5) accumulation. The antigen drifts result in antigen's protein sequence and structure changing which may reduce the antibody's specific binding ability with antigen, and finally improve the virus spreading effectiveness in the old strains vaccinated or immunized population. Unlike antigen drift, SNPs won't influence the antibody function, but they can reduce the nucleic acid amplification testing's specificity and sensitivity for the reason of mutation bases mismatch to primers.

Taking conserved sequences as reference is also necessary in antibodies developing for influenza diseases treatment. HA protein, which is one of the main antigens(6) on the surface of influenza virus, is the key in virus binding and entering host cells. Current antibodies developing for influenza disease treatment usually focuses on HA protein, so far, the antibodies targets to HA stalk(7) and HA head(8) have made much progress. However, due to the low degree of sequence conservation of the target binding sites, the existing influenza vaccines and antibodies are sensitive to antigenic drift(9), which makes these vaccines and antibodies have bad performance in long-term effect. Obviously, the long-term effect influenza products need to be robust against mutations, and conserved sequence is a good choice both in nucleic acid amplification testing primers and vaccines designing(10, 11).

Sequence alignment, including pairwise sequence alignment(12, 13) and multiple sequence alignment (MSA), are main methods in conserved sequences searching and plays an important role in bioinformatics. Different MSA algorithms have different performance in different sequence dataset. Clustal(14) is the most widely used tool for MSA. It calculates distance matrix by pairwise alignment, builds guide tree and makes progressive alignment following the guide tree. T-coffee(15) has a higher accuracy result than other methods, but more applicable for small sample size and short length sequence data, but the alignment speed is too slow in large scale dataset. Muscle(16) has the fastest speed compared with the first two methods, but it has a high memory requirement and not fit for long length sequences dataset.

2 Materials and methods

2.1 Influenza virus dataset

The sequence data used in this paper, including nucleotide sequence data and protein sequence data, are from NCBI influenza database(17). We select sequence type "Protein" and "Nucleotide", define search set as Type in "A", Host in "any", Country/ Region in "any", Protein in "HA", select Subtype from H1 to H9 respectively with N in "any", and set minimum sequence length to 560/1680 for protein/nucleotide sequence, collection date from 1918 to 2018. The sample numbers of each subtype are show in Table 1.

Table 1.

Subtype	Nucleotide	Protein
H1	23543	19916
H2	613	624
H3	19358	21743
H4	1868	1896
H5	5658	6325
H6	1745	1788
H7	2090	2203
H8	139	141
H9	3623	3718

2.2 Conserved sequence searching

In this research, we took amino acid sequences in place of nucleotide conserved sequences to process conserved sequence searching, for the reason of that the replacement can reduce the sequence length by two thirds with the similar information content, which can greatly improve the calculation efficiency.

Our algorithm of conserved sequences searching is based on breadth first search. Adding a new amino acid at the end of the current conserved sequences to generate new candidate conserved sequences, and making selection of these sequences by recalculating their conserved probability in global dataset.

2.2.1 Protein conserved sequence

The protein sequence dataset is directly download from NCBI database. In other situation which only nucleotide sequence dataset available in database, like SARS-CoV-2(18) case, amino acid sequences can translated from nucleotide sequence.

20 amino acids are defined as length 1 conserved sequence string seeds. These seeds are the first 20 strings of queue q and their conservative rate are 100%. q_i is the i th string in the queue and d_j is the j th sample sequence of the current target subtype. q_i and d_j are match when q_i is d_j's substring and $f(q_i, d_j)=1$. For the current string q_i in the queue, we use Eq.1 to calculate q_i's Conservative probability.

$$p_i = \frac{\sum_{j=1}^{n} f(q_i, d_j)}{n}, f(q_i, d_j) = \begin{cases} 1 \; if \; q_i \wedge d_j \; match \\ 0 \; otherwise \end{cases} \quad (1)$$

When $p_i > 99\%$ (the threshold set in this study is 99%), we consider q_i as a conserved sequence of the target subtype, and new amino acid char will be added to the end of the string to extend the q_i, and new strings will be added to the end of the queue. The function we used in $f(q_i, d_j)$ is "findstr" in Matlab. Besides, KMP algorithm(19) can be used in $f(q_i, d_j)$ optimization.

2.2.2 Nucleotide conserved sequence

We got the nucleotide conserved sequence candidates by locating the protein conserved sequences into the corresponding positions of nucleotide sequences. We used "multialign" function in Matlab to statistic the SNPs and the sequences with too much SNPs will not be considered as nucleotide conserved sequences.

2.3 PCR

To test our conserved sequences based primers functions of clinical HA testing and subtyping by using PCR experiments. Influenza viruses are RNA viruses. In this study, we used their cDNA as substitute templates.

2.3.1 Influenza virus template plasmid

Template concentration is diluted to 100ug/mL(20). The plasmid (Sino Biological, China) information is shown in Table 2.

Table 2. Influenza virus template plasmid

Subtype	Description	Catalog Number
H1	H1N1 (A/Beijing/262/1995) Hemagglutinin	VG11068-UT
H2	H2N2 (A/Guiyang/1/1957) Hemagglutinin	VG40119-UT
H3	H3N2 (A/Hong Kong/1/1968) Hemagglutinin	VG40116-UT
H4	H4N6 (A/Swine/Ontario/01911-1/99) Hemagglutinin	VG11706-UT
H5	H5N1 (Anhui/1/2005) Hemagglutinin	VG11048-UT
H6	H6N2 (A/chicken/Guangdong/C273/2011) Hemagglutinin	VG40398-UT
H7	H7N9 (A/Hangzhou/1/2013) Hemagglutinin	VG40105-UT
H8	H8N4 (A/pintail duck/Alberta/114/1979) Hemagglutinin	VG11722-UT
H9	H9N2 (A/Chicken/Hong Kong/G9/97) Hemagglutinin	VG40036-UT

2.3.2 PCR primers design

PCR primers were screened from nucleotide conserved sequences. We select the primer pairs with similar annealing temperature (△tm<1°C). SNPs are unavoidable, so we use degenerate bases (n < 5). Other principles follow the general requirements of primer design. The results are shown in Table 3.

Table 3. Primer design results

subtype	protein sequence	primer sequence	Fragment length(bp)
H1	NVTVTHS	(5'-3')AATGTRACWGTRACMCACTCW	1534
	SFWMCSN	(3'-5')ATTRGARCACATCCARAARCT	
H2	YHHSNDQ	(5'-3')TAYCAYCACAGCAATGAYCAR	481
	YQILAIYAT	(3'-5')TGTAGCRTADATDGCAAGDATTTGRTA	
H3	ITPNGSI	(5'-3')ATYACTCCAAATGGAAGCATY	532
	AEDMGN	(3'-5')ATTKCCCATRTCYTCAGC	
H4	CYPFDV	(5'-3')TGYTAYCCATTTGATGTG	1243
	QGYKDI	(3'-5')RATGTCYTTGTATCCYTG	
H5	VTVTHA	(5'-3')GTBACKGTYACACAYGCY	1219
	LMENERTLD	(3'-5')RTCYAGAGTTCTYTCATTTTCCATGAG	
H6	WYGYHHE	(5'-3')TGGTAYGGMTAYCAYCATGAR	349
	CFEFWHKC	(3'-5')RCAYTTRTGCCARAATTCAAARCA	
H7	FYAEMK	(5'-3')TTCTATGCRGARATGAAR	790
	GNVINW	(3'-5')CCARTTWATSACATTVCC	
H8	EGMCYP	(5'-3')GAGGGRATGTGYTAYCCT	175
	SINWLTKK	(3'-5')CTTYTTRGTYARCCARTTRATGCT	
H9	GWYGFQHS	(5'-3')GGTTGGTATGGDTTCCAGCATTCA	556
	AFLFWAM	(3'-5')CATGGCCCAGAAYARGAAGGC	

2.3.3 PCR experiment

We used E.coli DH5α Competent Cells(Takara, 9057) for the plasmid resuspension to culture bacteria and use SanPrep Column Plasmid Mini-Preps Kit(Sangon, B518191) extract plasmid. cDNA of HA segment are separated by restriction enzyme(Takara) digestion. Finally, we use NANODROP 2000(Thermo Scientific) to measure the cDNA concentration and dilute it to 50ng/μL.

The PCR was performed in a 50μL system with 5μL 10X PCR Buffer(Takara，R001B), 0.25μL TaKaRa Taq(Takara，R001B)，1μL dNTP Mixture(Takara，R001B)，0.2μL of forward primer(Sangon) and 0.2μL of reverse primer(Sangon), 0.1μL of cDNA, 43μL of nuclease-free water. Each primer pair test 9 subtypes in the same condition at one time. The PCR reaction conditions are shown in Table 4. Agarose gel electrophoresis images were collected under gel imager (LIUYI, WD-9413B).

Table 4. PCR reaction conditions

Subtype	Step1	Step2			Step3	cycles
		S1	S2	S3		
H1	94℃ 5min	94℃ 30s	56℃/2min	72℃/90s	72℃ 7min	30
H2			58.5℃/1min	72℃/70s		
H3			59℃/1min	72℃/30s		
H4			58.5℃/2min	72℃/70s		
H5			59℃/2min	72℃/70s		
H6			58℃/30s	72℃/30s		
H7			50℃/1min	72℃/30s		
H8			59℃/30s	72℃/70s		
H9			65℃/1min	72℃/30s		

3 Results

3.1 Conserved sequences

The protein conserved sequences (length≥5) were mapped to the spatial structure of influenza virus HA protein, and the visualization results are shown in Figure 1.

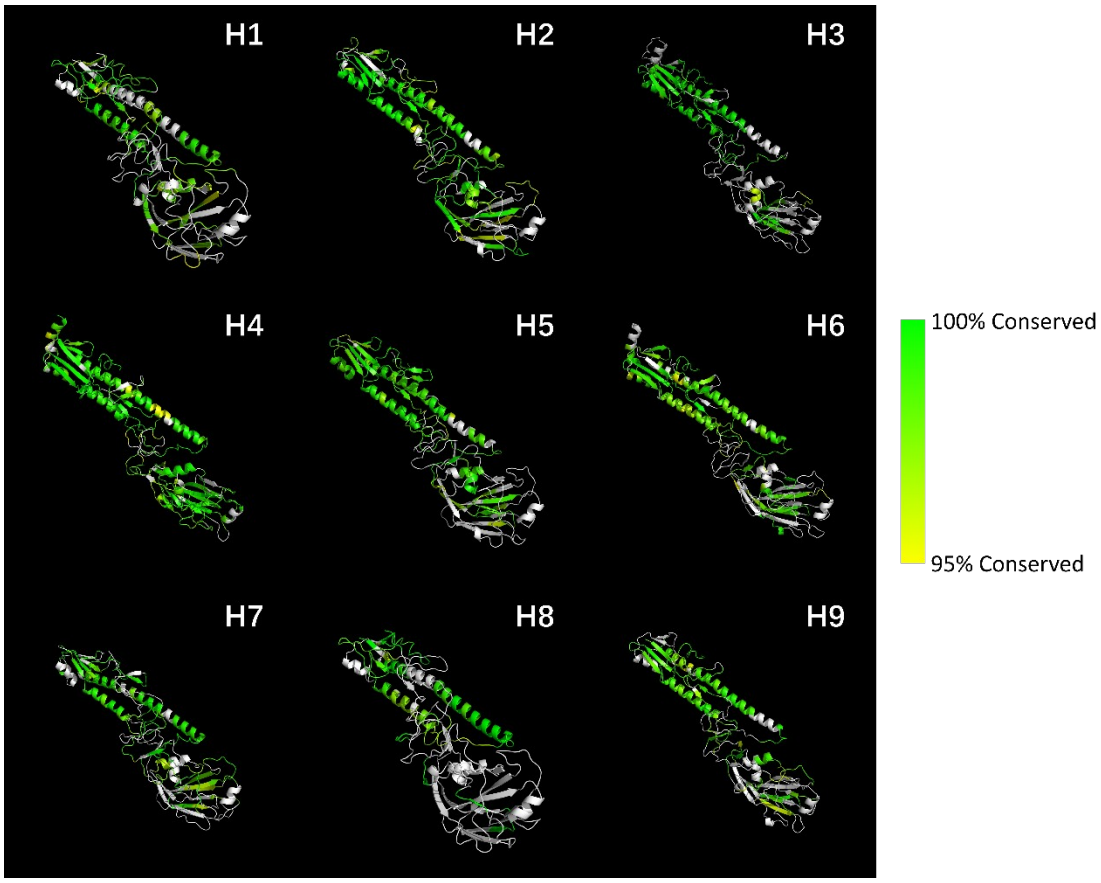

Figure.1. Visualization of HA protein conserved sequences mapping on spatial structure. Influenza virus HA protein is homotrimer, we have removed duplicate parts for better show. The structures of H1(1ruz)(21)、H2(2wr0)(22)、H3(4uo0)(23)、H4(5xl1)(24)、H5(1jsm)(25)、H6(5t08)(26)、H7(1ti8)(27)、H9(1jsd)(25) are from Protein Data Bank database(28). The structure of H8 subtype is from H1 structure's homologous modeling(29). The color from Green(RGB:010) to Yellow(RGB:110) represents different conservative rates from 100% to 95%, conservative rates lower than 95% are colored in white.

As the results show, in subtype H1-H9, conserved regions of influenza virus HA protein are concentrated in HA stem while HA head is less conserved. It's accordance with the fact that virus antigen drift is more likely to happen on antigen binding site.

The protein conserved sequences are calculated from each subtypes of HA protein datasets separately. Focusing on subtype identification, only conserved sequences with unique specificity for a single subtype will be considered in primers designing. Figure 2 shows the most matching bases of nucleotide conserved sequences in different subtypes.

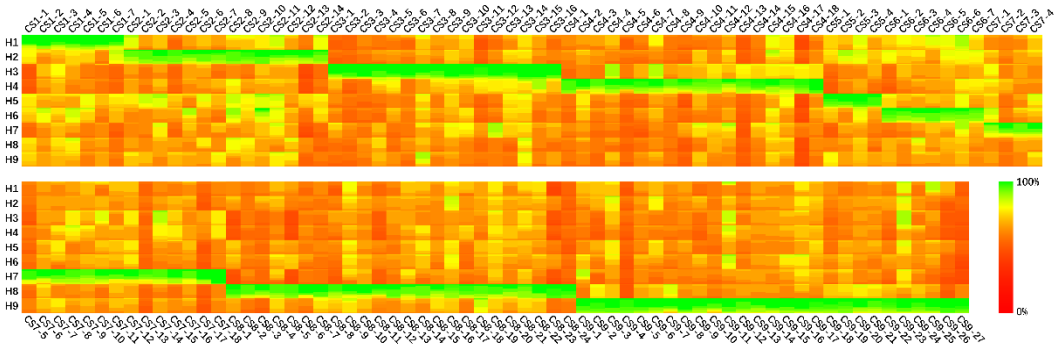

Figure.2. Nucleotide conserved sequences matching in H1-H9 subtypes. The matching of each nucleotide conserved sequence with each subtype is shown in 100 x 100 pixel squares, and each pixel represents an equal proportion of samples. 100% base matching is expressed in green (RGB010) and 0% base matching in red (RGB100). Row names are labeled by HA subtype, and column names are labeled by conserved sequences name, which named after CS(conserved sequence)+HA subtype+order.

Conserved sequences are cut into proper length to fit for PCR primers designing, figure 3 shows the designed primers' matching in H1-H9 subtypes.

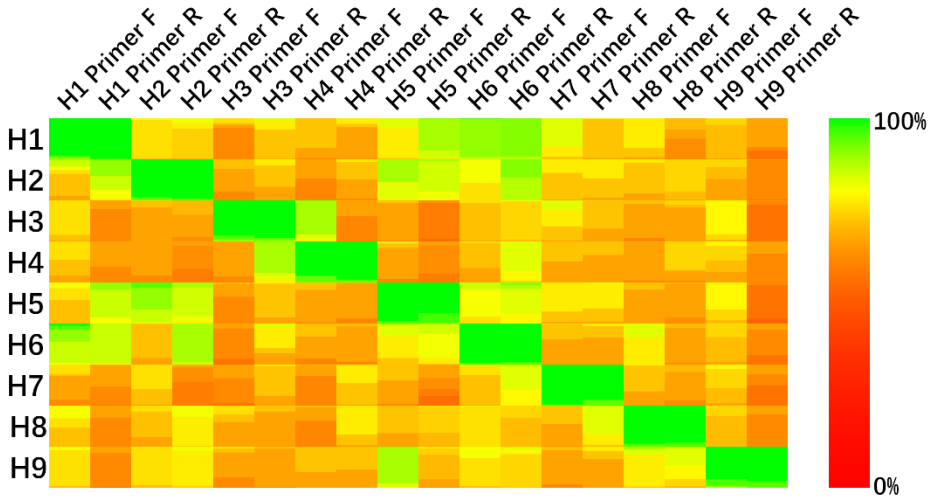

Figure.3. Designed primers' matching in H1-H9 subtypes. Each square contains 100 x 100 pixels, each pixel inside the squares represents an equal proportion of samples. HA subtype is labeled on the left of the figure, 100% base matching is expressed in green (RGB010) and 0% base matching in red (RGB100). Row names are labeled by HA subtype and column names are labeled by primer names.

As figure 3 shows, the conserved sequences based PCR primers have good expected specificity and sensitivity performance in HA subtype identification, and robust in global influenza virus strains.

3.2 PCR experiment

We proved the feasibility of nucleotide conserved sequences in influenza virus hemagglutinin subtype identification via PCR experiments. The primers were selected from calculated nucleotide conserved sequences set, designed follow the general primer design rules. To improve the sensitivity of primer pairs in matching target subtype, we used degenerate bases to smooth the influences of SNPs.

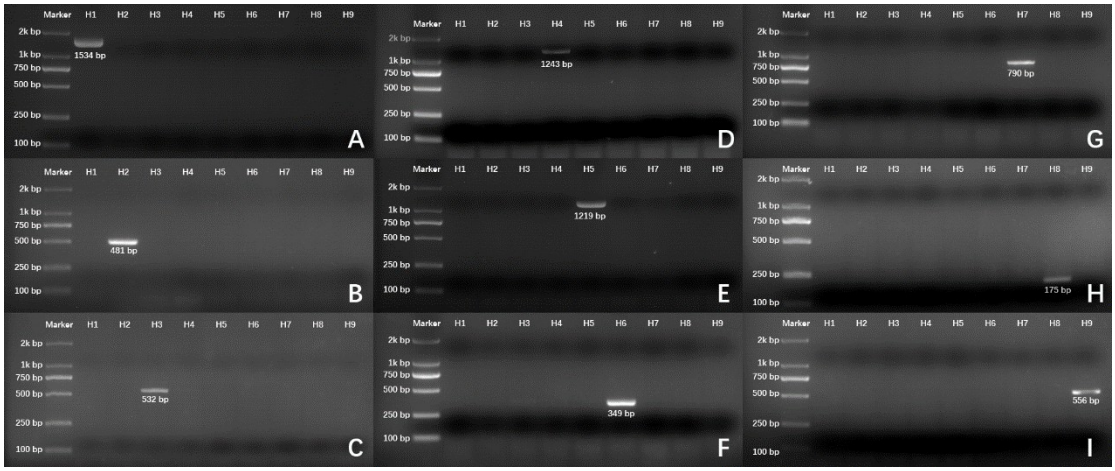

Figure.4. Agarose gel electrophoresis images of PCR experiment. (A-I) Using H1-H9 specific primers respectively for H1-H9 cDNA (lanes marked as labeled).

The results of length of the PCR amplification bands come up to our expectations. The 9 primer pairs all have good performance in H1-H9 subtypes identification and no PCR amplification bands on non-target subtypes lanes.

4 Discussion

The current MSA methods have bad performance in aligning datasets with long length sequence and large sample size. Full length MSA is unnecessary in many situations, such as the conserved sequences analysis and the primers designing. In this research, based on breadth first search strategy, the time cost of our conserved sequences searching method depends on conservative probability threshold setting and dataset sequence similarity. Taking the dataset containing m conserved protein sequences with L amino acids as example, the time complexity is $O(m \times L \times N)$, parallel computing can be used in candidate sequences conservative probability to improve computing efficiency. In addition, taking the conserved sequences as anchors(30) to separate the long sequences into several shorter segments, traditional MSA methods can be effectively optimized by conserved sequences searching algorithm.

In clinical practicing, the seasonal influenza testing relies on the several early samples sequences. Focusing on the new mutations of the current strain, the designed seasonal influenza testing primers is unavailable in long term testing to multi-strains(31). It's also undeniable that the negative effect of focusing on the long-term effectiveness of testing primers is the weakening of timeliness. Influenza primers designed based on

conserved sequences cannot distinguish between epidemic and seasonal influenza strains.

Different from testing primers, compared with immuring to single strain, the ideal vaccines are expected to help the vaccinated people to generate immunity for different seasonal strains. Using the conserved domain to induce immunity is a good choice, the recent reported research on universal vaccine(32) using HA stem replace HA head as target domain to reduce the negative effect of antigen drift on vaccine long term effect. However, as our results show, many mutations exist on HA stem, missense mutations can even change the protein structure to failure the antibody's specific binding. Selecting highly conserved sequences as target domain is significant for designing long term effect vaccine.

Considering the length of the high conserved sequences, mRNA vaccines(33, 34) and peptide vaccines(35, 36) have more potential in universal influenza vaccine research. Existing mRNA vaccine research(33) using conserved sequences from multiple segments, including HA stem, NA, M2 and NP, to strengthen the vaccine's effect on influenza virus antigen. Similarly, it's possible to design mRNA vaccine using conserved sequences from different HA subtypes to provide broad cross protection. Including the sequences from HA stem(36), multi-target is also feasible in peptide vaccine(35) designing.

In general, our conserved sequence searching method have good performance in large scale data dataset. Our results of conserved nucleotide sequences and amino acid sequences, not only in influenza testing and HA subtype identification, but also have high potential in future influenza research. Also, another important antigen on influenza virus surface with multiple subtypes, NA, is suitable with the same methods and procedure as HA.

5 Acknowledgement

We gratefully acknowledge Xuejuan Wang, Mingzhe Wang for helpful suggestions.

6 Financial Support

This work was supported by National Key R & D Plan in China (2016YFD0501103)

7 Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

8 Reference

1. Short KR, Kedzierska K, van de Sandt CE. Back to the Future: Lessons Learned From the 1918 Influenza Pandemic. Front Cell Infect Microbiol. 2018;8:343-.

2. Ask the expert: Influenza Q&A
https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)2018 [

3. Chiu W, Burnett R, Garcea R. Attachment and entry of influenza virus into host cells. Pivotal roles of hemagglutinin. Structural biology of viruses1997. p. 80-104.

4. Ketklao S, Boonarkart C, Phakaratsakul S, Auewarakul P, Suptawiwat O. Responses to the Sb epitope contributed to antigenic drift of the influenza A 2009 H1N1 virus. Arch Virol. 2020:1-10.

5. Kim H, Webster RG, Webby RJ. Influenza Virus: Dealing with a Drifting and Shifting Pathogen. Viral immunology. 2018;31(2):174-83.

6. Mayr J, Lau K, Lai JCC, Gagarinov IA, Shi Y, McAtamney S, et al. Unravelling the Role of O-glycans in Influenza A Virus Infection. Scientific reports. 2018;8(1):16382.

7. Khanna M, Sharma S, Kumar B, Rajput R. Protective immunity based on the conserved hemagglutinin stalk domain and its prospects for universal influenza vaccine development. Biomed Res Int. 2014;2014:546274-.

8. Giles BM, Ross TM. A computationally optimized broadly reactive antigen (COBRA) based H5N1 VLP vaccine elicits broadly reactive antibodies in mice and ferrets. Vaccine. 2011;29(16):3043-54.

9. Webster RG, Govorkova EA. Continuing challenges in influenza. Annals of the New York Academy of Sciences. 2014;1323(1):115-39.

10. Sisteré-Oró M, López-Serrano S, Veljkovic V, Pina-Pedrero S, Vergara-Alert J, Córdoba L, et al. DNA vaccine based on conserved HA-peptides induces strong immune response and rapidly clears influenza virus infection from vaccinated pigs. PloS one. 2019;14(9):e0222201.

11. Estrada LD, Schultz-Cherry S. Development of a Universal Influenza Vaccine. J Immunol. 2019;202(2):392-8.

12. Needleman SB, Wunsch CD. A GENERAL METHOD APPLICABLE TO SEARCH FOR SIMILARITIES IN AMINO ACID SEQUENCE OF 2 PROTEINS. Journal of Molecular Biology. 1970;48(3):443-+.

13. Smith TF, Waterman MS. IDENTIFICATION OF COMMON MOLECULAR SUBSEQUENCES. Journal of Molecular Biology. 1981;147(1):195-7.

14. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8.

15. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang J-M, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 2011;39(Web Server issue):W13-W7.

16. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113-.

17. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, et al. The influenza virus resource at the National Center for Biotechnology Information. J Virol. 2008;82(2):596-601.

18. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges. 2017;1(1):33-46.

19. Knuth D, Jr J, Pratt V. Fast Pattern Matching in Strings. SIAM J Comput. 1977;6:323-50.

20. Wang Chi h, ZENG Y. Comparative study of different nucleic acid extraction methods for influenza a H1N1 virus. Journal of Clinical and Experimental Medicine. 2017;16:230-2.

21. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, et al. The Structure and Receptor Binding Properties of the 1918 Influenza Hemagglutinin. Science. 2004;303(5665):1838.

22. Liu J, Stevens DJ, Haire LF, Walker PA, Coombs PJ, Russell RJ, et al. Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. Proc Natl Acad Sci U S A. 2009;106(40):17175-80.

23. Collins PJ, Vachieri SG, Haire LF, Ogrodowicz RW, Martin SR, Walker PA, et al. Recent evolution of equine influenza and the origin of canine influenza. Proc Natl Acad Sci U S A. 2014;111(30):11175-80.

24. Song H, Qi J, Xiao H, Bi Y, Zhang W, Xu Y, et al. Avian-to-Human Receptor-Binding Adaptation by Influenza A Virus Hemagglutinin H4. Cell Reports. 2017;20(5):1201-14.

25. Ha Y, Stevens DJ, Skehel JJ, Wiley DC. H5 avian and H9 swine influenza virus haemagglutinin structures: possible origin of influenza subtypes. EMBO J. 2002;21(5):865-75.

26. de Vries RP, Tzarum N, Peng W, Thompson AJ, Ambepitiya Wickramasinghe IN, de la Pena ATT, et al. A single mutation in Taiwanese H6N1 influenza hemagglutinin switches binding to human-type receptors. EMBO Mol Med. 2017;9(9):1314-25.

27. Russell RJ, Gamblin SJ, Haire LF, Stevens DJ, Xiao B, Ha Y, et al. H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. Virology. 2004;325(2):287-96.

28. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and

education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res. 2018;47(D1):D464-D74.

29. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-W303.

30. Pitschi F, Devauchelle C, Corel E. Automatic detection of anchor points for multiple sequence alignment. BMC Bioinformatics. 2010;11:445-.

31. Duh D, Blažič B. Single mutation in the matrix gene of seasonal influenza A viruses critically affects the performance of diagnostic molecular assay. Journal of virological methods. 2018;251:43-5.

32. Krammer F. Novel universal influenza virus vaccine approaches. Curr Opin Virol. 2016;17:95-103.

33. Freyn AW, Ramos da Silva J, Rosado VC, Bliss CM, Pine M, Mui BL, et al. A Multi-Targeting, Nucleoside-Modified mRNA Influenza Virus Vaccine Provides Broad Protection in Mice. Molecular therapy : the journal of the American Society of Gene Therapy. 2020;28(7):1569-84.

34. Pardi N, Parkhouse K, Kirkpatrick E, McMahon M, Zost SJ, Mui BL, et al. Nucleoside-modified mRNA immunization elicits influenza virus hemagglutinin stalk-specific antibodies. Nature communications. 2018;9(1):3361.

35. Herrera-Rodriguez J, Meijerhof T, Niesters HG, Stjernholm G, Hovden A-O, Sørensen B, et al. A novel peptide-based vaccine candidate with protective efficacy against influenza A in a mouse model. Virology. 2018;515:21-8.

36. Khrustalev VV, Khrustaleva TA, Kordyukova LV. Selection and structural analysis of the NY25 peptide – A vaccine candidate from hemagglutinin of swine-origin Influenza H1N1. Microbial Pathogenesis. 2018;125:72-83.