



## Article

# Conserved Sequence Analysis of Influenza A Virus HA Segment and Its Application in Rapid Typing

Qianyu Lin <sup>1</sup>, Xiang Ji <sup>2</sup>, Feng Wu <sup>2</sup>  and Lan Ma <sup>1,2,3,\*</sup> 

<sup>1</sup> Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen 518055, China; lingy18@mails.tsinghua.edu.cn

<sup>2</sup> Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; hijiri\_byakuren@126.com (X.J.); wu8721@126.com (F.W.)

<sup>3</sup> Shenzhen Bay Laboratory, Shenzhen 518038, China

\* Correspondence: malan@sz.tsinghua.edu.cn; Tel.: +86-180-2872-1478

**Abstract:** The high mutation rate of the influenza A virus hemagglutinin segment poses great challenges to its long-term effective testing and subtyping. Our conserved sequence searching method achieves high-specificity conserved sequences on H1–H9 subtypes. In addition, PCR experiments show that primers based on conserved sequences can be used in influenza A virus HA subtyping. Conserved sequence-based primers are expected to be long-term, effective subtyping tools for influenza A virus HA.

**Keywords:** influenza 1; multiple sequence alignment 2; conserved sequence 3; rapid detection 4



**Citation:** Lin, Q.; Ji, X.; Wu, F.; Ma, L. Conserved Sequence Analysis of Influenza A Virus HA Segment and Its Application in Rapid Typing. *Diagnostics* **2021**, *11*, 1328. <https://doi.org/10.3390/diagnostics11081328>

Academic Editor: Ilka Engelmann

Received: 30 June 2021

Accepted: 22 July 2021

Published: 23 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Influenza is a common widespread infectious disease that has caused a considerable number of human deaths in the past century [1]. The annual influenza epidemic results in 3 to 5 million cases of severe illness and causes 0.29 to 0.65 million respiratory deaths [2]. Influenza A (referred to as “influenza” in the following text) virus is an RNA virus. The surface antigens of influenza are hemagglutinin (HA) and neuraminidase (NA); 18 different HAs and 11 different NAs are known to date. Influenza’s subtype is defined according to HA and NA subtypes. The influenza virus invades and infects cells via HA’s binding to cells’ sialic acid-containing receptors [3]. Specific binding with hemagglutinin to reduce the infectious capacity of the influenza virus is the main idea of existing treatment and prevention of influenza diseases. Thus, fast and accurate identification of HA subtypes is important for the diagnosis and treatment of influenza.

However, the high mutation rate of influenza virus results in a large number of single-nucleotide polymorphisms (SNPs) and antigenic drift (accumulation) [4,5]. The antigenic drifts result in the antigen’s protein sequence and structure changing, which may reduce the antibody’s specific binding ability with the antigen, finally improving the effectiveness of virus spread among the population vaccinated or immunized to the old strains. Unlike antigenic drift, SNPs do not influence the antibody function, but they can reduce the nucleic acid amplification testing’s specificity and sensitivity for the reason of mutation base mismatch to primers.

Taking conserved sequences as references is also necessary in antibody development for influenza disease treatment. The HA protein, which is one of the main antigens [6] on the surface of the influenza virus, is the key to the virus’s successful binding and its entering host cells. Current antibody development for influenza disease treatment generally focuses on the HA protein; to date, the antibodies targeting the HA stalk [7] and head [8] have allowed considerable 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 causes these vaccines and antibodies to have bad performance in

terms of the long-term effect. Of course, the long-term effect of influenza products needs to be robust against mutations, and a conserved sequence is a good choice for both testing primers for nucleic acid amplification and vaccine design [10,11].

Sequence alignment, including pairwise sequence alignment [12,13] and multiple sequence alignment (MSA), is the main method in conserved sequence searching, and it plays an important role in bioinformatics. Different MSA algorithms have different performances in different sequence datasets. Clustal [14] is the most widely used tool for MSA. It calculates the distance matrix by pairwise alignment, builds guide trees, and makes progressive alignment following the guide tree. T-coffee [15] has a higher accuracy result than that of other methods, and it is applicable to small sample sizes and short-length sequence data, but the alignment speed is too slow when used on large-scale datasets. Muscle [16] has the fastest speed among the methods, but it has a high memory requirement and is not suitable for long-length sequences dataset. Therefore, current MSA methods are not applicable to some situations, such as calculating the conserved sequences from long-length sequences and large-scale datasets. With the aim of application to such datasets, in this work, we introduce a new method for conserved sequences based on a breadth-first search. Focusing on conserved regions instead of processing full-length alignment can greatly reduce the computational resource. Moreover, we applied our method to an influenza A virus HA dataset. Based on the conserved sequence, we designed the primers with high specificity and sensitivity for influenza A virus HA subtyping that were successfully tested via PCR experiments.

## 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 the NCBI influenza database [17]. We selected sequence types “Protein” and “Nucleotide”; defined the search set as type “A”, the host as “any”, the country/region as “any”, and the protein as “HA”; selected subtype from H1 to H9 with N in “any”; and set the minimum sequence length to 560/1680 for the protein/nucleotide sequence and the collection date from 1918 to 2018. The sample numbers of each subtype are show in Table 1.

**Table 1.** Data size of the dataset.

Subtype	Nucleotide	Protein
H1	23,543	19,916
H2	613	624
H3	19,358	21,743
H4	1868	1896
H5	5658	6325
H6	1745	1788
H7	2090	2203
H8	139	141
H9	3623	3718

### 2.2. Conserved Sequence Searching

The software used to implement the algorithm is MATLAB (R2020a for windows, MathWorks, Inc., Natick, MA, USA).

In this research, we took amino acid sequences in place of nucleotide conserved sequences to process conserved sequence searching, as 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 sequence searching is based on a breadth-first search, adding a new amino acid at the end of the current conserved sequences to generate

new candidate conserved sequences and selecting these sequences by recalculating their conserved probability in the global dataset.

### 2.2.1. Protein Conserved Sequence

The protein sequence dataset was downloaded directly from the NCBI database (<https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database> (accessed on 29 June 2021)). In other situations where only the nucleotide sequence dataset is available in the database, such as in the case of SARS-CoV-2 [18], amino acid sequences can be translated from the nucleotide sequence. Twenty amino acids were defined as length 1 conserved sequence string seeds. These seeds were the first 20 strings of queue  $q$  and their conservative rate is 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  $d_j$  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 used Equation (1) to calculate  $q_i$ 's conservative probability.

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

When  $p_i > 99\%$  (the threshold set in this study is 99%), we consider  $q_i$  as a conserved sequence of the target subtype; a new amino acid character is added to the end of the string to extend  $q_i$  and new strings are added to the end of the queue.

### 2.2.2. Nucleotide Conserved Sequence

We obtained the nucleotide conserved sequence candidates by locating the protein conserved sequences in the corresponding positions of nucleotide sequences. We used the "multialign" function in MATLAB to calculate the SNPs. The sequences with too many SNPs were not considered as nucleotide conserved sequences.

## 2.3. Polymerase Chain Reaction

PCR experiments were used to test our conserved sequence-based primer functions of clinical HA testing and subtyping. Influenza viruses are RNA viruses. In this study, we used their cDNA as substitute templates.

### 2.3.1. Influenza Virus Template Plasmid

The template concentration was diluted to 100 ug/mL [19]. 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 selected the primer pairs with similar annealing temperatures ( $\Delta t_m < 1$  °C). SNPs are unavoidable; therefore, we used 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')TGTAGCRTADATDGCAAGDATTGRTA	
H3	ITPNGSI	(5'–3')ATYACTCCAAATGGAAGCATY	532
	AEDMGN	(3'–5')ATKCCCATRRCYTCAGC	
H4	CYPFDV	(5'–3')TGYTAYCCATTTGATGTG	1243
	QGYKDI	(3'–5')RATGTCYTTGTATCCYTG	
H5	VTVTHA	(5'–3')GTBACKGTYACACAYGCY	1219
	LMENERTLD	(3'–5')RTCYAGAGTICTYTCATTTTCCATGAG	
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')CATGGCCCAGAAAYARGAAGGC	

### 2.3.3. PCR Experiment

We used *E. coli* DH5 $\alpha$  Competent Cells (Takara, Kusatsu, Japan, 9057) for the plasmid resuspension to culture bacteria and SanPrep Column Plasmid Mini-Preps Kit (Sangon, Shanghai, China, B518191) to extract plasmid. The cDNA of HA segments was separated by restriction enzyme (Takara, Kusatsu, Japan) digestion. Finally, we used a NANODROP 2000 (Thermo Scientific, Waltham, MA, USA) to measure the cDNA concentration and diluted it to 50 ng/ $\mu$ L. The PCR was performed in a 50-microliter system with 5  $\mu$ L of 10X PCR buffer (Takara, Kusatsu, Japan, R001B), 0.25  $\mu$ L of TaKaRa Taq (Takara, R001B), 1  $\mu$ L of dNTP mixture (Takara, Kusatsu, Japan, R001B), 0.2  $\mu$ L of forward primer (Sangon, Shanghai, China) and 0.2  $\mu$ L of reverse primer (Sangon, Shanghai, China), 0.1  $\mu$ L of cDNA, and 43  $\mu$ L of nuclease-free water. Each primer pair tested 9 subtypes in the same condition and same batch. The PCR reaction conditions are shown in Table 4. Agarose gel electrophoresis images were collected using a gel imager (LIUYI, WD-9413B).

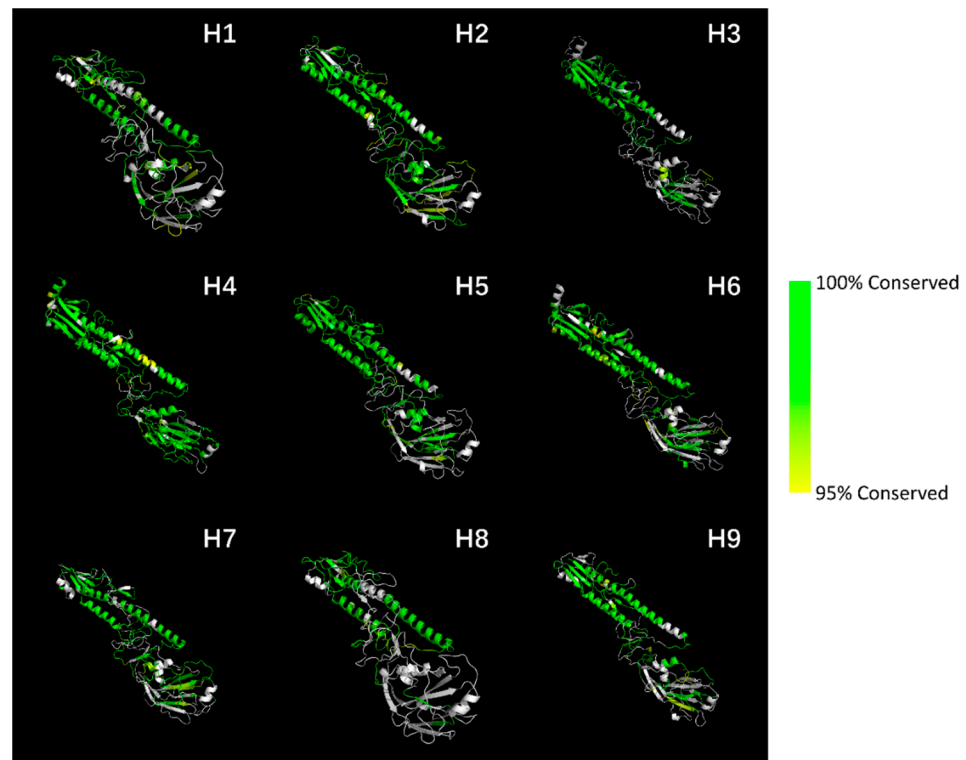
Table 4. PCR reaction conditions.

Subtype	Step 1	Step 2			Step 3	Cycles
		S1	S2	S3		
H1			56 °C/2 min	72 °C/90 s		
H2			58.5 °C/1 min	72 °C/70 s		
H3			59 °C/1 min	72 °C/30 s		
H4			58.5 °C/2 min	72 °C/70 s		
H5	94 °C	94 °C	59 °C/2 min	72 °C/70 s	72 °C	30
H6	5 min	30 s	58 °C/30 s	72 °C/30 s	7 min	
H7			50 °C/1 min	72 °C/30 s		
H8			59 °C/30 s	72 °C/70 s		
H9			65 °C/1 min	72 °C/30 s		

### 3. Results

#### 3.1. Conserved Sequences

The protein conserved sequences (length  $\geq 5$ ) were mapped to the spatial structure of the influenza virus HA protein, and the visualization results are shown in Figure 1. Conserved sequences with conservative rates are listed in Table S1.

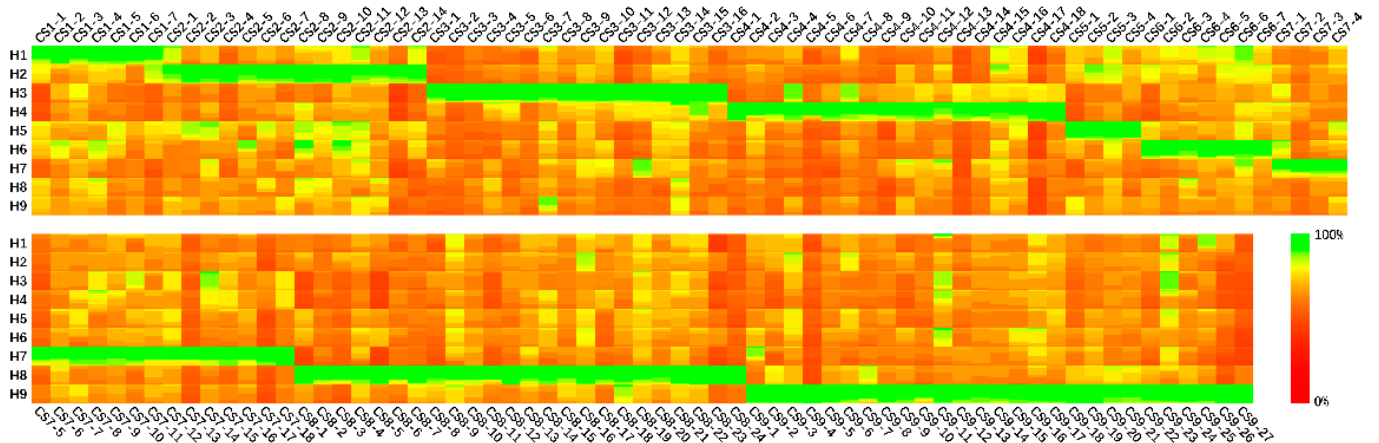


**Figure 1.** Visualization of HA protein conserved sequence mapping on spatial structure. The influenza virus HA protein is a homotrimer. We removed duplicate parts for enhanced presentation. The structures of H1 (1ruz) [20], H2 (2wr0) [21], H3 (4uo0) [22], H4 (5xl1) [23], H5 (1jsm) [24], H6 (5t08) [25], H7 (1ti8) [26], and H9 (1jsd) [24] are from the Protein Data Bank database [27]. The structure of the H8 subtype is from H1 structure's homologous modelling [28]. The colors from green (RGB:010) to yellow (RGB:110) represent different conservative rates from 100 to 95%; conservative rates lower than 95% are colored in white.

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

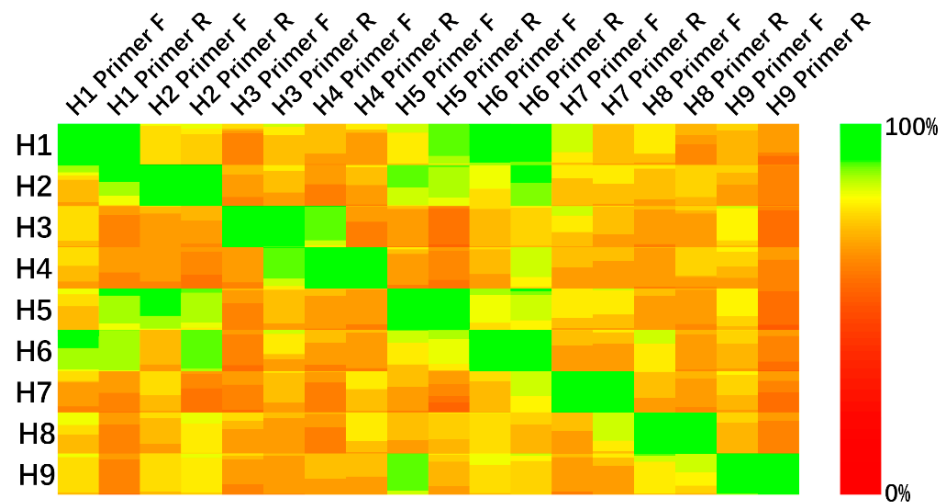
The protein conserved sequences were separately calculated from each subtype of the HA protein datasets. Focusing on subtype identification, only the conserved sequences with unique specificity for a single subtype were considered in the primer design. Figure 2 shows the most matching bases of nucleotide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

As Figure 3 shows, the conserved sequence-based PCR primers are expected to have a good specificity and sensitivity performance in HA subtype identification and are robust in global influenza virus strains.



**Figure 2.** Nucleotide conserved sequence matching in H1–H9 subtypes. The matching of each nucleotide conserved sequence with each subtype is shown in  $100 \times 100$  pixel squares, and each pixel represents an equal proportion of samples. Base matching of 100% is expressed in green (RGB010) and 0% base matching in red (RGB100). Row names are labelled by HA subtype, and column names are labelled by conserved sequences name: CS (conserved sequence) + HA subtype + order.

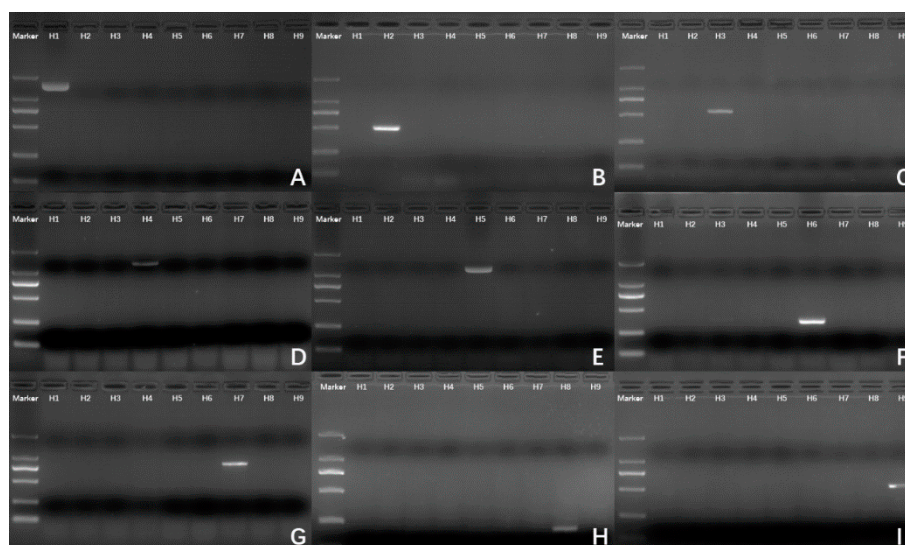
The conserved sequences are cut to a proper length to fit the PCR primer design. Figure 3 shows the designed primers’ matching in H1–H9 subtypes.



**Figure 3.** Matching of designed primers in H1–H9 subtypes. Each square contains  $100 \times 100$  pixels, and each pixel inside the squares represents an equal proportion of samples. The HA subtype is labelled on the left of the figure; 100% base matching is expressed in green (RGB010); and 0% base matching is expressed in red (RGB100). Row names are labelled by HA subtype and column names are labelled by primer names.

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 a calculated nucleotide conserved sequence set and were designed following the general primer design rules. To improve the sensitivity of primer pairs in matching the target subtype, we used degenerate bases to smooth the influence of SNPs (Figure 4).



**Figure 4.** Agarose gel electrophoresis images of PCR experiment. (A–I) Using H1–H9-specific primers for H1–H9 cDNA (lanes marked as labelled). The marker used to the left side of each gel is a DNA ladder with the bands of 100, 250, 500, 750, 1000, and 2000 bp; the sample lanes are labelled from H1 to H9 in each gel. The original figures can be found in the Supplementary Materials.

The results regarding the length of the PCR amplification bands met our expectations. The nine primer pairs all have good performance in H1–H9 subtype identification with no PCR amplification bands on non-target subtypes lanes.

#### 4. Discussion

The current MSA methods have bad performance in aligning datasets with long sequences and large sample sizes. Full-length MSA is unnecessary in many situations, such as conserved sequence analysis and primer design. In this study, based on the breadth-first search strategy, the time cost of our conserved sequence searching method depends on the conservative probability threshold setting and dataset sequence similarity. Taking the dataset containing  $m$  conserved protein sequences with  $L$  amino acids as an example, the time cost  $T \propto (m \times L \times N)$ . In addition, our conserved sequence results can be taken as anchors [29] to separate the long sequences into several shorter segments. Traditional MSA methods can be effectively optimized via the conserved sequence searching algorithm.

In clinical practicing, seasonal influenza testing relies on several early sample sequences. Focusing on the new mutations of the current strain, the designed seasonal influenza testing primers are unavailable in long-term testing of multi-strains [30]. It is also undeniable that focusing on the long-term effectiveness of testing primers negatively affects timeliness. The influenza primers designed based on conserved sequences cannot distinguish between epidemic and seasonal influenza strains.

Unlike testing primers, which are limited to a single strain, the ideal vaccines are expected to help vaccinated people to generate immunity against various seasonal strains. Using the conserved domain to induce immunity is a good choice. Recent research reported on a universal vaccine [31] replacing the HA head with the HA stem as the target domain to reduce the negative effect of antigenic drift on the long-term effect of the vaccine. However, as our results show, many mutations exist on the HA stem, and missense mutations can even change the protein structure, causing the antibody's specific binding to fail. Selecting highly conserved sequences as a target domain is significant in designing a vaccine with long-term effectiveness.

Considering the length of the high conserved sequences, mRNA vaccines [32,33] and peptide vaccines [34,35] have potential in universal influenza vaccine research. Existing mRNA vaccine research [32] uses conserved sequences from multiple segments, including the HA stem, NA, M2, and NP, to strengthen the vaccine's effect on the influenza virus

antigen. Similarly, it is possible to design an mRNA vaccine using conserved sequences from different HA subtypes to provide broad cross-protection. Including the sequences from the HA stem [35], multitargeting is also feasible in peptide vaccine [34] design.

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

## 5. Conclusions

This study applied a conserved sequence searching algorithm based on a breadth-first search to an influenza A HA segment dataset and provided candidate sequences for long-term, effective testing primers and vaccine design for different HA subtypes. Via the PCR experiment, we proved the feasibility of conserved sequence-based primers in long-term influenza A virus HA testing and subtyping.

**Supplementary Materials:** Available online: <https://www.mdpi.com/article/10.3390/diagnostics11081328/s1>. Table S1: Conserved sequences information.

**Author Contributions:** Conceptualization, Q.L., X.J. and L.M.; methodology, Q.L.; software, Q.L.; validation, Q.L.; formal analysis, Q.L., X.J. and F.W.; investigation, Q.L., X.J. and F.W.; resources, L.M.; data curation, Q.L.; writing—original draft preparation, Q.L. and F.W.; writing—review and editing, Q.L. and F.W.; visualization, Q.L.; supervision, L.M.; project administration, L.M.; funding acquisition, L.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Key R & D Plan in China (2016YFD0501103), Shenzhen strategic emerging industry development special funds (JCYJ20170816143646446), Shenzhen Science and Technology research and development funds (JCYJ20200109143018683).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All sequence data are from the NCBI influenza database: <https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database> (accessed on 29 June 2021).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Short, K.R.; Kedzierska, K.; van de Sandt, C.E. Back to the Future: Lessons Learned from the 1918 Influenza Pandemic. *Front. Cell. Infect. Microbiol.* **2018**, *8*, 343. [[CrossRef](#)] [[PubMed](#)]
2. Ask the Expert: Influenza Q&A. 2018. Available online: [https://www.who.int/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)) (accessed on 6 November 2018).
3. Chiu, W.; Burnett, R.; Garcea, R. Attachment and entry of influenza virus into host cells. Pivotal roles of hemagglutinin. In *Structural Biology of Viruses*; Chiu, W., Burnett, R.M., Garcea, R.L., Eds.; Oxford University Press: New York, NY, USA, 1997; pp. 80–104.
4. Ketklo, S.; Ketklo, 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**, *165*, 2503–2512. [[CrossRef](#)]
5. Kim, H.; Webster, R.G.; Webby, R.J. Influenza Virus: Dealing with a Drifting and Shifting Pathogen. *Viral Immunol.* **2018**, *31*, 174–183. [[CrossRef](#)]
6. Mayr, J.; Lau, K.; Lai, J.C.C.; Gagarinov, I.A.; Shi, Y.; McAtamney, S.; Chan, R.W.Y.; Nicholls, J.; von Itzstein, M.; Haselhorst, T. Unravelling the Role of O-glycans in Influenza A Virus Infection. *Sci. Rep.* **2018**, *8*, 16382. [[CrossRef](#)]
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. [[CrossRef](#)] [[PubMed](#)]
8. Giles, B.M.; Ross, T.M. A computationally optimized broadly reactive antigen (COBRA) based H5N1 VLP vaccine elicits broadly reactive antibodies in mice and ferrets. *Vaccine* **2011**, *29*, 3043–3054. [[CrossRef](#)] [[PubMed](#)]
9. Webster, R.G.; Govorkova, E.A. Continuing Challenges in Influenza. *Ann. N. Y. Acad. Sci.* **2014**, *1323*, 115–139. [[CrossRef](#)]



10. Sisteré-Oró, M.; Lopez-Serrano, S.; Veljkovic, V.; Pina-Pedrero, S.; Vergara-Alert, J.; Cordoba, L.; Perez-Maillo, M.; Pleguezuelos, P.; Vidal, E.; Segales, J.; 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*, e0222201. [[CrossRef](#)]
11. Estrada, L.D.; Schultz-Cherry, S. Development of a Universal Influenza Vaccine. *J. Immunol.* **2019**, *202*, 392–398. [[CrossRef](#)]
12. Needleman, S.B.; Wunsch, C.D. A general method applicable to search for similarities in the amino acid sequence of 2 proteins. *J. Mol. Biol.* **1970**, *48*, 443. [[CrossRef](#)]
13. Smith, T.F.; Waterman, M.S. Identification of common molecular subsequences. *J. Mol. Biol.* **1981**, *147*, 195–197. [[CrossRef](#)]
14. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)]
15. Di Tommaso, P.; Moretti, S.; Xenarios, I.; Orobítg, M.; Montanyola, A.; Chang, J.-M.; Taly, J.-F.; Notredame, C. 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*, W13–W17. [[CrossRef](#)]
16. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 1–19. [[CrossRef](#)]
17. Bao, Y.; Bolotov, P.; Dernovoy, D.; Kiryutin, B.; Zaslavsky, L.; Tatusova, T.; Ostell, J.; Lipman, D. The influenza virus resource at the National Center for Biotechnology Information. *J. Virol.* **2008**, *82*, 596–601. [[CrossRef](#)] [[PubMed](#)]
18. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Glob. Chall.* **2017**, *1*, 33–46. [[CrossRef](#)]
19. Wang, C.H.; Zeng, Y. Comparative study of different nucleic acid extraction methods for influenza A H1N1 virus. *J. Clin. Exp. Med.* **2017**, *16*, 230–232.
20. Gamblin, S.J.; Haire, L.F.; Russell, R.J.; Stevens, D.J.; Xiao, B.; Ha, Y.; Vasisht, N.; Steinhauer, D.A.; Daniels, R.S.; Elliot, A.; et al. The Structure and Receptor Binding Properties of the 1918 Influenza Hemagglutinin. *Science* **2004**, *303*, 1838. [[CrossRef](#)]
21. Liu, J.; Stevens, D.J.; Haire, L.F.; Walker, P.A.; Coombs, P.J.; Russell, R.J.; Gamblin, S.J.; Skehel, J.J. Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17175–17180. [[CrossRef](#)]
22. Collins, P.J.; Vachieri, S.G.; Haire, L.F.; Ogorodowicz, R.W.; Martin, S.R.; Walker, P.A.; Xiong, X.; Gamblin, S.J.; Skehel, J.J. Recent evolution of equine influenza and the origin of canine influenza. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11175–11180. [[CrossRef](#)]
23. Song, H.; Qi, J.; Xiao, H.; Bi, Y.; Zhang, W.; Xu, Y.; Wang, F.; Shi, Y.; Gao, G.F. Avian-to-Human Receptor-Binding Adaptation by Influenza A Virus Hemagglutinin H4. *Cell Rep.* **2017**, *20*, 1201–1214. [[CrossRef](#)]
24. Ha, Y.; Stevens, D.J.; Skehel, J.J.; Wiley, D.C. H5 avian and H9 swine influenza virus haemagglutinin structures: Possible origin of influenza subtypes. *EMBO J.* **2002**, *21*, 865–875. [[CrossRef](#)]
25. de Vries, R.P.; Tzarum, N.; Peng, W.; Thompson, A.J.; Ambepitiya Wickramasinghe, I.N.; de la Pena, A.T.T.; van Breemen, M.J.; Bouwman, K.M.; Zhu, X.; McBride, R.; et al. A single mutation in Taiwanese H6N1 influenza hemagglutinin switches binding to human-type receptors. *EMBO Mol. Med.* **2017**, *9*, 1314–1325. [[CrossRef](#)]
26. Russell, R.J.; Gamblin, S.J.; Haire, L.F.; Stevens, D.J.; Xiao, B.; Ha, Y.; Skehel, J.J. H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. *Virology* **2004**, *325*, 287–296. [[CrossRef](#)]
27. Burley, S.K.; Berman, H.M.; Bhikadiya, C.; Bi, C.; Chen, L.; Di Costanzo, L.; Christie, C.; Dalenberg, K.; Duarte, J.M.; Dutta, S.; 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*, D464–D474. [[CrossRef](#)]
28. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [[CrossRef](#)] [[PubMed](#)]
29. Pitschi, F.; Devauchelle, C.; Corel, E. Automatic detection of anchor points for multiple sequence alignment. *BMC Bioinform.* **2010**, *11*, 1–11. [[CrossRef](#)] [[PubMed](#)]
30. Duh, D.; Blažič, B. Single mutation in the matrix gene of seasonal influenza A viruses critically affects the performance of diagnostic molecular assay. *J. Virol. Methods* **2018**, *251*, 43–45. [[CrossRef](#)] [[PubMed](#)]
31. Krammer, F. Novel universal influenza virus vaccine approaches. *Curr. Opin. Virol.* **2016**, *17*, 95–103. [[CrossRef](#)]
32. Freyn, A.W.; da Silva, J.R.; Rosado, V.C.; Bliss, C.M.; Pine, M.; Mui, B.L.; Tam, Y.K.; Madden, T.D.; de Souza Ferreira, L.C.; Weissman, D.; et al. A Multi-Targeting, Nucleoside-Modified mRNA Influenza Virus Vaccine Provides Broad Protection in Mice. *Mol. Ther.* **2020**, *28*, 1569–1584. [[CrossRef](#)]
33. Pardi, N.; Parkhouse, K.; Kirkpatrick, E.; McMahon, M.; Zost, S.J.; Mui, B.L.; Tam, Y.K.; Kariko, K.; Barbosa, C.J.; Madden, T.D.; et al. Nucleoside-modified mRNA immunization elicits influenza virus hemagglutinin stalk-specific antibodies. *Nat. Commun.* **2018**, *9*, 3361. [[CrossRef](#)] [[PubMed](#)]
34. Herrera-Rodriguez, J.; Meijerhof, T.; Niesters, H.G.; Stjernholm, G.; Hovden, A.-O.; Sorensen, B.; Okvist, M.; Sommerfelt, M.A.; Huckriede, A. A novel peptide-based vaccine candidate with protective efficacy against influenza A in a mouse model. *Virology* **2018**, *515*, 21–28. [[CrossRef](#)] [[PubMed](#)]
35. Khrustalev, V.V.; Khrustaleva, T.A.; Kordyukova, L.V. Selection and structural analysis of the NY25 peptide—A vaccine candidate from hemagglutinin of swine-origin Influenza H1N1. *Microb. Pathog.* **2018**, *125*, 72–83. [[CrossRef](#)] [[PubMed](#)]