

Article

MADE: A Computational Tool for Predicting Vaccine Effectiveness for the Influenza A(H3N2) Virus Adapted to Embryonated Eggs

Hui Chen ^{1,†} , Junqiu Wang ^{2,3,†} , Yunsong Liu ^{2,4,†} , Ivy Quek Ee Ling ⁵, Chih Chuan Shih ⁵, Dafei Wu ², Zhiyan Fu ⁶, Raphael Tze Chuen Lee ⁷ , Miao Xu ⁸, Vincent T. Chow ⁹ , Sebastian Maurer-Stroh ^{7,10,11,12} , Da Zhou ^{3,*}, Jianjun Liu ^{1,13,*} and Weiwei Zhai ^{1,2,14,*}

- ¹ Human Genomics, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore; chen1@gis.a-star.edu.sg
 - ² Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; qiuqiu.wang@warwick.ac.uk (J.W.); liuyunsong@ioz.ac.cn (Y.L.); wudf@ioz.ac.cn (D.W.)
 - ³ School of Mathematical Science, Xiamen University, Xiamen 361005, China
 - ⁴ University of the Chinese Academy of Sciences, Beijing 100049, China
 - ⁵ Bioinformatics Core, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore; quekel@gis.a-star.edu.sg (I.Q.E.L.); shihcc@gis.a-star.edu.sg (C.C.S.)
 - ⁶ IHiS—Integrated Health Information Systems, Singapore 554910, Singapore; zhiyan.fu@ihis.com.sg
 - ⁷ Bioinformatics Institute, Agency for Science, Technology and Research, Singapore 138671, Singapore; leetc@bii.a-star.edu.sg (R.T.C.L.); sebastianms@bii.a-star.edu.sg (S.M.-S.)
 - ⁸ State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-Sen University Cancer Center, Guangzhou 510060, China; xumiao@susucc.org.cn
 - ⁹ NUHS Infectious Diseases Translational Research Program, Department of Microbiology & Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117545, Singapore; micctk@nus.edu.sg
 - ¹⁰ School of Biological Sciences (SBS), Nanyang Technological University (NTU), Singapore 637551, Singapore
 - ¹¹ National Public Health Laboratory (NPHL), Ministry of Health (MOH), Singapore 308442, Singapore
 - ¹² Department of Biological Sciences, National University of Singapore (NUS), Singapore 117543, Singapore
 - ¹³ Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore
 - ¹⁴ Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China
- * Correspondence: zhouda@xmu.edu.cn (D.Z.); liuj3@gis.a-star.edu.sg (J.L.); weiweizhai@ioz.ac.cn (W.Z.)
† These authors contributed equally to this work.



Citation: Chen, H.; Wang, J.; Liu, Y.; Ling, I.Q.E.; Shih, C.C.; Wu, D.; Fu, Z.; Lee, R.T.C.; Xu, M.; Chow, V.T.; et al. MADE: A Computational Tool for Predicting Vaccine Effectiveness for the Influenza A(H3N2) Virus Adapted to Embryonated Eggs. *Vaccines* **2022**, *10*, 907. <https://doi.org/10.3390/vaccines10060907>

Academic Editor: Elena A. Govorkova

Received: 27 April 2022

Accepted: 31 May 2022

Published: 6 June 2022

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



Copyright: © 2022 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/>).

Abstract: Seasonal Influenza H3N2 virus poses a great threat to public health, but its vaccine efficacy remains suboptimal. One critical step in influenza vaccine production is the viral passage in embryonated eggs. Recently, the strength of egg passage adaptation was found to be rapidly increasing with time driven by convergent evolution at a set of functionally important codons in the hemagglutinin (HA1). In this study, we aim to take advantage of the negative correlation between egg passage adaptation and vaccine effectiveness (VE) and develop a computational tool for selecting the best candidate vaccine virus (CVV) for vaccine production. Using a probabilistic approach known as mutational mapping, we characterized the pattern of sequence evolution driven by egg passage adaptation and developed a new metric known as the adaptive distance (AD) which measures the overall strength of egg passage adaptation. We found that AD is negatively correlated with the influenza H3N2 vaccine effectiveness (VE) and ~75% of the variability in VE can be explained by AD. Based on these findings, we developed a computational package that can Measure the Adaptive Distance and predict vaccine Effectiveness (MADE). MADE provides a powerful tool for the community to calibrate the effect of egg passage adaptation and select more reliable strains with minimum egg-passaged changes as the seasonal A/H3N2 influenza vaccine.

Keywords: egg passage adaptation; vaccine effectiveness; influenza H3N2 virus; adaptive evolution; vaccine production

1. Introduction

As an RNA virus, influenza evolves rapidly and causes annual epidemics resulting in 3 to 5 million cases of severe illness, and 290,000 to 650,000 deaths [1]. Due to its rapid antigenic drift [2,3], the World Health Organization (WHO) organizes consultation meetings twice a year to recommend the best candidate vaccine viruses (CVVs) for the world. Despite many years of efforts, vaccine efficacy against influenza viruses, especially the H3N2 subtype, remains suboptimal [4–6].

There are many factors that can influence influenza vaccine effectiveness (VE) [7,8]. In addition to antigenic drift [2,3], several other factors ranging from glycosylation of the hemagglutinin [9], egg passage adaptation [10], repeated vaccination [11], imprinting and cohort effect [12] as well as the waning effect [13] can contribute to the variability of vaccine efficacies [4]. Among these factors, egg passage adaptation during vaccine production has been implicated in low vaccine efficacy across many years [14–16], but its link to vaccine effectiveness has been constrained to individual mutations occurred in different years, lacking a systematic metric integrating substitutions across many sites. Using a probabilistic inference procedure known as mutational mapping [17], our recent study found that egg passage adaptation was driven by substitutions in several key codons in HA1 across years [18] and the strength of egg passage adaptation has become progressively stronger in the recent past. When the H3N2 influenza had just crossed species boundary from birds to humans in 1968, passaging them in an avian environment led to weak positive selection across many codon positions. As influenza viruses became well-adapted to the human host, passaging them in embryonated eggs (an avian environment) led to much elevated passage adaptation at a set of functionally important codons (e.g., codon 186 and 194) [18]. The intensity of natural selection calculated as the rate of nonsynonymous to synonymous ratio (i.e., d_N/d_S) is often infinity due to the extremely high rate of nonsynonymous changes [18]. In order to measure the overall strength of egg passage adaptation, a new metric known as the adaptive distance (AD) was developed to quantify the intensity of egg passage adaptation in a given candidate vaccine virus (CVV) [10]. For the first time, the overall level of egg passage adaptation can be combined into a single metric (AD) integrating multiple substitutions distributed across many codons. Interestingly, a strong negative correlation was found between AD and VE for vaccine strains in recent years, and as much as 75% of the variation in VE can be explained by AD [10]. Thus, increasingly stronger egg passage adaptation has led to influenza vaccines with more substitutions and poor vaccine efficacy.

In this study, we extended the findings from our previous work [10,18] and have developed a new computational tool designated as MADE: Measuring Adaptive Distance and predicting vaccine Effectiveness using allelic barcodes (MADE). Based on allelic status (i.e., allelic barcodes) at a given set of positively selected codons in the HA1 gene for egg passage adaptation, MADE can: (a) calibrate the level of egg passage adaptation by calculating AD for any given candidate vaccine virus (CVV) and predict its potential vaccine effectiveness; (b) since there are a large number of sequences in the database with unknown passage history, and egg passage adaptation can confound many evolutionary analyses including “contaminating” the signal of adaptation in humans [18,19], MADE can infer whether a given isolate with unknown passage history has been grown in embryonated eggs using a machine learning approach known as XGBoost [20]. If the inferred passage history is not embryonated eggs, MADE can predict its passage history (e.g., MDCK or other growth medium) for the input sequence. In general, we aim to develop a computational tool that can select reliable strains with minimum egg-passages changes for the seasonal A/H3N2 influenza vaccine.

2. Materials and Methods

2.1. Data Curation and Computational Analysis

The steps retrieving the public data and performing the computational analysis are similar to our recently published work [10] and are explained in greater detail in the Supplementary Materials. From the Global Initiative on Sharing All Influenza Data (GISAID) [21], we retrieved 76,489 influenza A/H3N2 HA1 sequences and their associated passage histories (Table S1). After quality control and further annotating passage histories of all sequences (Table S2), 69,362 sequences were retained for subsequent analysis and the numbering of different codon positions is based on the HA1 sequence with 329 amino acids. The subsequent analysis follows the below steps: (a) sequence alignment and phylogenetic reconstruction [22] (Figure S1); (b) using a probabilistic approach known as mutational mapping to sample possible evolutionary histories of the sequences [17]. The mutational mapping method is constructed based on theories from the continuous time Markov chain and has been used widely in inferring the history of mutations [17]. (c) Given the sampled evolutionary histories of the input sequences, two statistical tests (i.e., the enrichment and convergent test) were used to identify codons driving the egg passage adaptation [10] (Figure 1A). In the enrichment test, we asked the question whether substitutions happened in a given codon are more enriched in egg terminal branches, while in the convergent test, we asked the question whether substitutions happened in a given codon along the egg terminal branches are more likely to be convergent substitutions (Supplementary Materials). Using theories from the continuous time Markov chain, we can explicitly test for the significance of these patterns across all codons and identify codons responsible for egg passage adaptation. (d) Calculating the enrichment score (ES) for all the alleles (amino acids) at the codons responsible for egg passage adaptation (Figure 1B). The ES is calculated as the ratio of frequencies for a given allele in egg-passaged strains versus all the strains (see Results (Section 3)). Alleles with high ES score are those alleles occur specifically in egg-passaged strains. For each sequence, we can extract a high-dimensional vector of ES scores at the codons responsible for egg passage adaptation. This high-dimensional vector serves as an allelic barcode for a given input sequence. (e) Given the high dimensional ES scores for all the sequences, we can perform principal component analysis across all strains. The adaptive distance (AD) is defined as the distance between the input strain (e.g., CVV) and the centroid of the major cluster for most of the non-egg sequences (Figure 1). Adaptive distance integrates patterns of adaptive evolution across multiple codons and captures the intensity of overall level of egg passage adaptation (see Results (Section 3)). (f) Performing the linear regression between AD and VE for the historical vaccine strains, and predicting the potential VE of the input CVV based on its AD value. Since the predicted VE is not a direct measurement of VE in humans, but rather is predicted based on the level of egg passage adaptation captured by AD, we denoted it as VE_{ad} .

2.2. Classify Input Strains with Unknown Passage History

Since MADE can effectively measure the strength of egg passage adaptation in a given sequence, it also provides an additional feature distinguishing egg strains and non-egg-passaged strains. In order to achieve this goal, we first transformed the sequence dataset into the one-hot encoded dataset, based on which Random Forest and XGBoost methods were applied to predict the passage histories (see Supplementary Materials, Figures S4 and S5, Tables S3 and S4 for details). The performance of the algorithm was evaluated using the precision score (F1-score, defined as $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$), which measures the accuracy of a test based on precision and recall. Here, precision is the ratio of true positives to all predicted positives (i.e., $TP / (TP + FP)$), while recall is the ratio of true positives to all actual positives (i.e., $TP / (TP + FN)$).

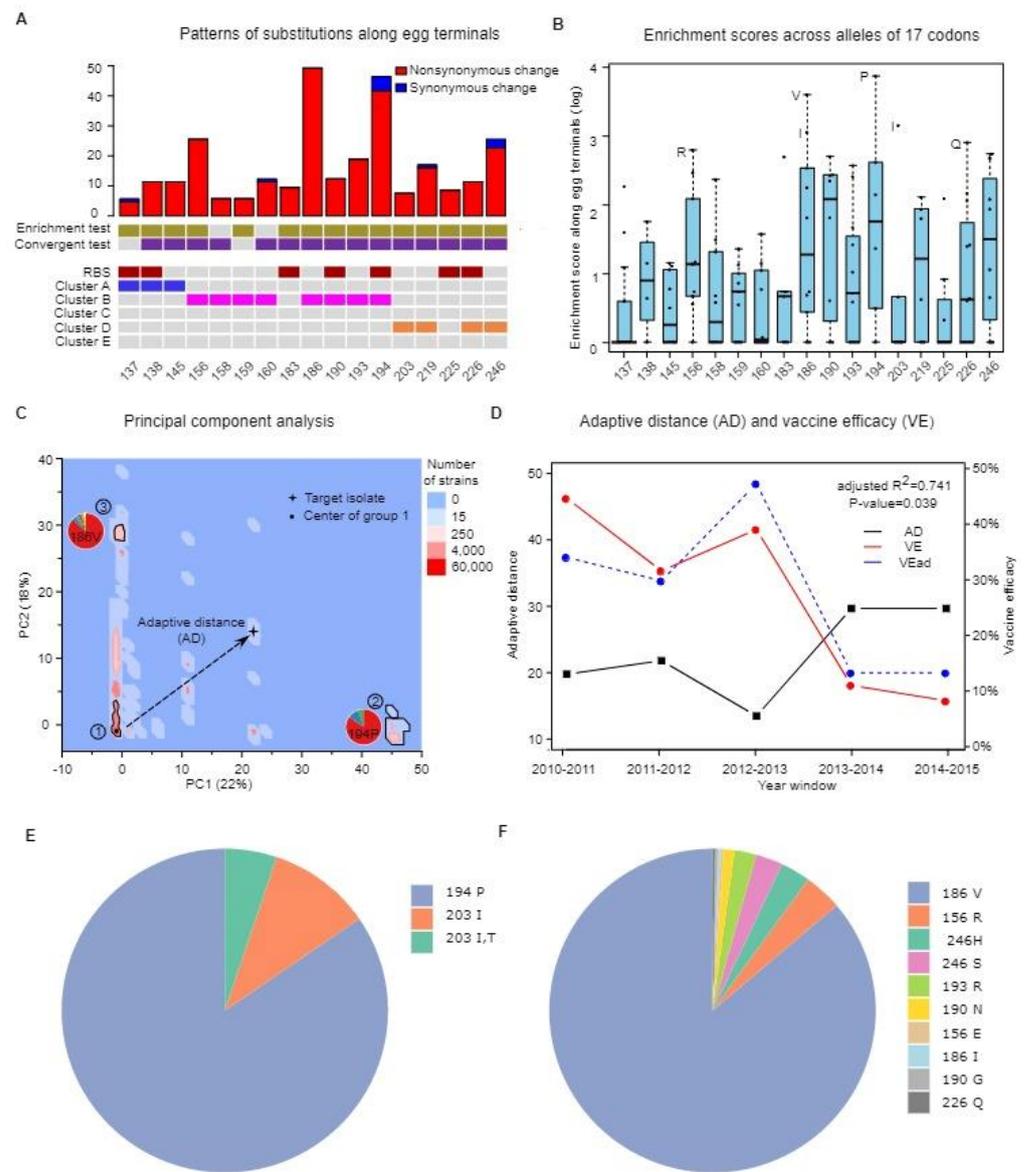


Figure 1. Statistical approaches for characterizing egg passage adaptation. **(A)** Numbers of nonsynonymous and synonymous changes at HA1 codons responsible for egg passage adaptation. The codons showing statistical significance from the enrichment and convergent tests (q -value < 0.05) are labelled in yellow and purple, respectively. Codons located in functional domains such as receptor binding sites (RBS), antigenic epitope A, B and D will be labeled in red, blue, violet and orange color respectively. **(B)** Enrichment scores across codons responsible for the egg passage adaptation. Alleles with enrichment scores higher than 20 are labeled. **(C)** Principal component analysis of the 17-dimensional space of ES scores for all the sequences. Discrete subgroups of sequences carrying different passage-related alleles distribute in clusters away from the major cluster. Pie charts display major adaptive alleles in different groups. For example, 194P and 186V are the dominant adaptive alleles observed in cluster 2 and 3, respectively. Adaptive distance (AD) is defined as the distance between the target strain (e.g., CVV) and the centroid of the major cluster for most of the non-egg sequences (i.e., group 1). **(D)** Correlation between the adaptive distance (AD) and vaccine efficacy (VE) between influenza seasons from 2010 to 2015. The predicted VE_{ad} is drawn as the dashed line. **(E)** Proportion of different alleles with enrichment score >10 in the cluster 2 of the PCA map (panel C). **(F)** Proportion of different alleles with enrichment score >10 in the cluster 3 of the PCA map (panel C).

3. Results

From the GISAID database, we extracted 69,362 HA1 sequences from the H3N2 influenza spanning 1968 to 2018 (Methods (Section 2), Table S1). Among all the sequences, 898 sequences were passaged in embryonated eggs and majority of the sequences were grown in mammalian cells (e.g., MDCK cells, Table S2). Using a maximum likelihood approach [22], we inferred the phylogenetic relationships and the mutational parameters (e.g., substitution matrix) of all the sequences (Figure S1). Conditioning on maximum likelihood parameters, we performed mutational mapping, a probabilistic approach to infer the history of mutations along each branch for all the codon positions [17,23]. In order to identify codons responsible for egg passage adaptation, we implemented two statistical tests based on patterns of sequence evolution [10]. In the enrichment test, a statistical procedure is implemented to examine whether changes along terminal branches were more enriched in egg-passaged sequences (i.e., egg terminal branches) than expected by chance. The second test (i.e., the convergent test) identifies codons with repeatedly the same change (i.e., convergent substitutions) along egg terminal branches (Methods (Section 2) and Supplementary Materials). Applying these two statistical tests, we identified 17 positively selected codon positions in the HA1 gene at a false discovery rate of 5% (Figure 1A). These codons strongly enriched for antigenic epitopes B and D as well as the receptor-binding sites (RBS) (Figure 1A). For example, 8 out of 17 codons are from epitope B sites, which had 180 nonsynonymous substitutions and only 6 synonymous substitutions, indicating strong adaptive evolution driving the evolution of these codons in these functional domains (Figure S2). When we inferred substitutions along the egg terminal branches for these 17 codons, we observed an average of 289 nonsynonymous changes and 11 synonymous changes, which contributed to 63.5% of all the nonsynonymous changes, but only 5.2% of the synonymous substitutions for all codons along the terminal branches. This high level of nonsynonymous to synonymous ratio has hitherto not been observed in any naturally evolving system [24], indicating extremely strong adaptive evolution in embryonated eggs (Figure 1A).

When an allele (amino acid) is preferentially selected in egg-passaged sequences, it will be in relatively high frequency in the egg sequences relative to the background frequency. In order to systematically measure frequency differences in egg-passaged strains versus all the sequences, we previously developed an enrichment score (ES) as $p_{\text{egg}}/p_{\text{all}}$ for each allele observed over all 329 HA1 codons [10]. Here, p_{egg} denotes the proportion of strains carrying the specific allele in the egg-passaged strains, while p_{all} represents the proportion of strains carrying the specific allele among all sequences. Interestingly, a few alleles from several codons responsible for egg passage adaptation exhibit extremely high ESs (i.e., high frequency in egg strains, low frequency in the total set), such as 186V (ES = 35.13) and 194P (ES = 46.6) (Figure 1B). Alleles with high enrichment scores also tend to enrich in important antigenic epitopes such as epitope B as well as RBS (Figure S2B). In summary, we defined an important statistic known as the ES score which could measure the preference of individual alleles at codons responsible for egg passage adaptation.

Given the extremely high level of nonsynonymous to synonymous change (d_N/d_S is often infinity), we previously developed an important metric known as the adaptive distance (AD) to measure the level of egg passage adaptation that occurred in the sequence [10]. For a given sequence, enrichment scores for alleles at the 17 HA1 positions provided a 17-dimensional metric defining the unique sequence feature of that input sequence (i.e., allelic barcode). Sequences bearing preferred alleles (i.e., amino acid) at selected codons will have very high ES scores across the 17 dimensions. Using principal component analysis (PCA), we projected the 17-dimensional space into the first two leading principal components (PCs). Interestingly, most of the sequences not passaged in embryonated eggs cluster as a major group (group 1, Figure 1C), whereas egg-passaged sequences reside within various clusters away from the major group. Inspecting allelic configurations in each dispersed cluster, specific egg-passage related alleles are enriched in the clusters similar to the antigenic map [3] (Figure 1C,E,F). We thus defined the adaptive distance (AD) as

the distance from the input strain to the centroid of the major group representing viruses without egg passage adaptation (i.e., group 1, Figure 1C, Methods (Section 2)). When we performed a linear regression between AD and VE curated from a recent meta-analysis combining many studies [5], we found a strong negative relationship with a correlation coefficient of $R^2 = 0.741$ (p -value = 0.039) (Figure 1D). This high linear correlation allowed us to predict VE of any input CVV with a regression line as $VE_{ad} = -0.022 \times AD + 0.78$ (Figure 1D).

When we investigated the historical records for these vaccine strains, the egg passage adaptation correlated very well with the VE data. For example, A/Victoria/361/2011 was recommended as the vaccine strain for 2012–2013 by the WHO [14]. When we calculated AD for multiple A/Victoria/361/2011 strains grown in MDCK cells, the MDCK-passaged sequences yielded a mean AD of 0.384 (located within cluster 1, Figure 1C), whereas the egg-passaged sequences had a mean AD of 29.528 (located in cluster 3 in Figure 1C), resulting in a low predicted VE_{ad} (~12.1%). Looking at the substitutions in the antigenic epitope B (H156Q, ES = 2.139, G186V, ES = 35.130) as well as epitope D (S219Y, ES = 5.186), consistent with the low vaccine performance in the 2012–2013 season [14]. These observations suggest that egg passage adaptation can often affect the antigenicity of the virus and subsequently lead to poor vaccine efficacy (see Discussions (Section 4)).

To make these methods available to the research community, we developed MADE (Figure 2), a software that can perform the above-mentioned functions for any given candidate CVV including (1) calculating ESs for alleles at the 17 positively selected codons in HA1 gene, (2) performing principal component analysis and calculating the AD for the CVV, and (3) predicting VE_{ad} of the input CVV based on the signal of adaptive evolution measured in AD. As MADE is assuming appreciable levels of egg passage adaptation, the prediction will not be performed for strains with very little signal of egg passage adaptation (Figure 2). Since many egg-passaged sequences bear specific alleles, we developed machine learning methods to predict whether the input sequence has truly been passaged in embryonated eggs. Using one-hot encoding along with machine learning methods such as Random Forest [25] and XGboost [20], we can predict the passage history of the input sequence very accurately (F1-score of 0.8624 under Random Forest and 0.864 under XGboost) based on the allelic configuration at these 17 codons.

In addition to predict potential vaccine efficacy of an egg-passaged CVV, as 1/3 of the sequences in the GISAID database have unknown passage history and egg passage adaptation can often confound many sequence studies including evolutionary analysis of influenza evolution in humans [18,19], the machine learning model in the pre-screening step can also be further extended to distinguish egg-passaged strains from non-egg strains. We thus constructed a multi-class classification method which can further classify the unknown passage history into four passage types including “Cell”, “MDCK”, “SIAT” and “Other”. These multi-class machine learning models can achieve good performances with F1-score of 0.9566 under the Random Forest and 0.9521 under the XGboost method (see Methods (Section 2) and Supplementary Materials).

After all the analysis, MADE will output a complete report including: (a) the sequence information of the CVV; (b) allelic status and enrichment scores at the 17 codons responsible for the egg passage adaptation; (c) AD and predicted VE_{ad} ; (d) the likely passage history of the input sequence if the growth medium is not embryonated eggs. The backend engine for generating the report is “R Markdown”. MADE is freely available online in two different versions with an open-source form available at github (https://github.com/chenh1gis/MADE_docker_v1 (accessed on 1 June 2022)) and an interactive web interface available at <http://39.105.1.41/made> (accessed on 1 June 2022).

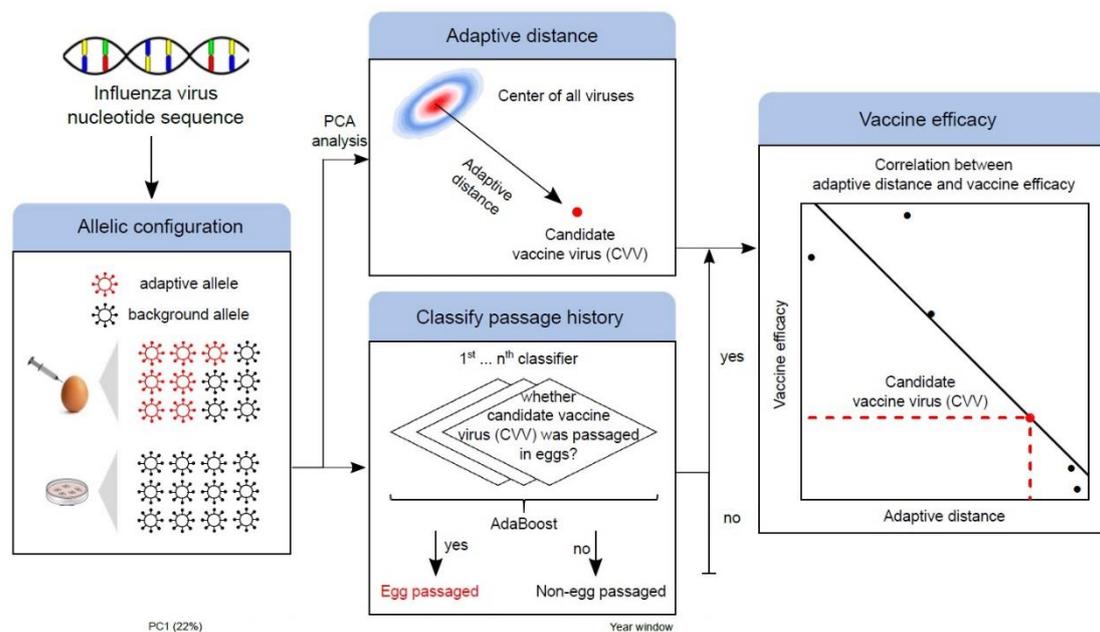


Figure 2. Schematic workflow of MADE. For any candidate vaccine virus (CVV), enrichment scores (ES) of all the alleles (amino acids) at the 17 positively selected HA1 codons were computed. These high dimensional vectors calculated at the 17 codons serve as allelic barcodes for each strain. Principal component analysis of the ES scores for all existing sequences was conducted, and the adaptive distance (AD) for the CVV was computed. Subsequently, MADE predicted the VE_{ad} of the input CVV based on the linear relationship between AD and vaccine effectiveness (VE). In addition, a machine learning algorithm can be applied to classify whether the input sequence with unknown passage history has been grown in embryonated eggs or other passage mediums (e.g., MDCK).

4. Discussion

Using a powerful probabilistic approach known as mutational mapping, we have developed an efficient tool that can measure the extent of egg passage adaptation in a given CVV and predict its potential vaccine effectiveness (VE_{ad}). For the first time, the strength of egg passage adaptation can be integrated into a single metric (i.e., AD) alleviating the challenge studying different substitutions across years. The strong correlation between AD and VE provides an important connection linking egg passage adaptation to vaccine effectiveness. Following up statistical evidences provided by MADE, subsequent experimental methods (e.g., animal challenge experiments using ferrets) can be employed to test the immunogenicity and functional consequence of these passaged strains (e.g., analyzing elicited antibodies for their neutralizing abilities against the wild-type and passaged viruses) [26]. Moreover, the statistical approach developed here can potentially be further extended to other viral types including influenza H1N1. To facilitate its accessibility, we have integrated MADE with Flusurver and FluCluster-AI (<https://flusurver.bii.a-star.edu.sg/> (accessed on 1 June 2022)), a popular tool hosted at GISAID that can allow users to link candidate mutations with literature-reported and 3D structurally relevant phenotypes based on information submitted to GISAID. Taken together, MADE provides a swift and powerful tool for the research community to select the best candidate strains without strong signals of egg passage adaptation.

The analysis of adaptive distance provides a unique opportunity connecting adaptive evolution, immunogenicity of the viral strains and vaccine efficacy. For example, in addition to A/Victoria/361/2011 mentioned earlier, in the strain selected for both 2016–2017 and 2017–2018 seasons (i.e., A/Hong Kong/4801/2014), egg passage adaptation generated a series of adaptive substitutions including HA1 T160K, L194P and N96S [27], with the L194P mutation being one of the strongest selected alleles with an ES of 46.6. Human sera collected from individuals vaccinated with egg-based strains displayed much reduced

inhibition abilities against circulating strains, leading to one of the most severe influenza seasons since the A/H1N1 pandemic of 2009 in the United States [27]. In addition, even though the adaptive distance calculated based on ES is not a direct measurement of the antigenic property, it behaves very similarly to the antigenic space generated using the hemagglutination inhibition (HI) assay [3]. For example, the PCA map based on ES scores is similar to the antigenic space with discrete islands representing different passage-related substitutions across years. Substitutions with higher enrichment scores tend to be those with large adaptive distance as well as strong antigenic jumps. Thus, the evolutionary analysis of egg passage adaptation provides an important means connecting egg passage adaptation, immunogenicity of the viral strains in humans as well as vaccine effectiveness [28].

It is worth pointing out that in addition to egg adaptation, many other viral [9] and host factors [11–13] have also been implicated in the variable vaccine efficacy [4]. For example, using a recently developed method for predicting vaccine efficacy based on antigenic distance [29], we observed that a slightly different trend of VE was predicted across the years (Figure 1D and Figure S3), suggesting that different factors can influence VE through different mechanisms. Even though MADE provides an important computational tool integrating signal of egg passage adaptation, it is still an inference method based on historical data. As influenza viruses are constantly co-evolving with humans, we can imagine egg passage adaptation will also change accordingly through time and it will be important to update the inference procedure continuously. Moreover, further experimental studies will be needed to further confirm functional consequences of egg passage substitutions. Taken together, how to holistically integrate multiple factors and how we can combine computational and experimental evidence for vaccine efficacy will be a crucial question for the field.

Even though many efforts including a universal influenza vaccine [30], non-egg cell lines [31,32], genetically engineered viruses with egg-adapted neuraminidase (NA) [33,34] and DNA vaccines [4] have been explored to improve vaccine efficacy, the transition to newer technologies [35] will likely progress rather slowly due to economic and technological limitations. It is likely that the egg-based vaccine production system will continue for quite some time [4,7]. Given the steadily increasing intensity of egg passage adaptation of the H3N2 influenza strains [18], MADE provides a timely and powerful method for the research and public health community circumventing the impact of egg passage adaptation when selecting seasonal A/H3N2 influenza vaccine strains.

5. Conclusions

We have developed a computational package called MADE that can predict VE_{ad} for any candidate vaccine strain. We believe that MADE will help the community select more reliable A/H3N2 influenza strains with minimum egg-passaged adaptation, providing a unique opportunity for improving influenza vaccines.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/vaccines10060907/s1>, Table S1: The complete list of influenza H3N2 sequences used in this study and acknowledgements to the GISAID database. Table S2: Number of sequences from different passage histories. Table S3: Performance of the binary classification (Egg and non-egg). Table S4: Performance of the multi-class classification (Egg, Cell, MDCK and SIAT). Figure S1: The maximum likelihood tree with labeled passage histories. Figure S2: Patterns of egg passage adaptation. Figure S3: Correlation between the adaptive distance (AD) and vaccine efficacy (VE). Figure S4: Performance of the binary classification. Figure S5: Performance of the multi-class classification.

Author Contributions: W.Z. and H.C. conceived this work. J.L. helped supervise the work. H.C. performed the analysis and created the tool. D.Z., J.W. and Y.L. conceived and implemented the machine learning algorithms. D.W. helped with the data curation. I.Q.E.L., C.C.S. and Z.F. provided the technical support for MADE. R.T.C.L. and S.M.-S. integrated MADE into FluSurver and FluCluster-AI (<https://flusurver.bii.a-star.edu.sg/> (accessed on 1 June 2022)). S.M.-S., V.T.C. and M.X. participated in the discussions. H.C., J.W. and W.Z. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This study is funded by NMRC Open Fund Young Individual Research Grant (MOH-OFYIRG18nov-0013) and COVID-19 TOP UP GRANT (COVID19TUG21-0129) awarded to H.C. D.Z. is supported in part by National Natural Sciences Foundation of China (grant 11971405) and W.Z. is supported in part by Strategic Priority Research Program of the Chinese Academy of Sciences XDPB17 (XDPB17), National Science Foundation of China (grant 31970566), National Key R&D program of China (grant 2018YFC1406902 and 2018YFC0910400). Besides, R.T.C.L. was supported by NMRC grant MOH-OFYIRG19nov-0013/MOH-000565-00 to S.M.S.

Data Availability Statement: Two versions of MADE (Measuring Adaptive Distance and vaccine Effectiveness using allelic barcodes) are available at https://github.com/chenh1gis/MADE_docker_v1 (accessed on 31 May 2022) and <http://39.105.1.41/made> (accessed on 31 May 2022).

Acknowledgments: We would like to thank Rasmus Nielsen, Chung-I Wu, Edward A Belongia, Yanhua Qu for constructive comments on the manuscript. We also would like to thank Ai Shan Lee for helpful technical support and Jacob Josiah Santiago Alvarez for the help with the English presentation of the manuscript. We also acknowledge Richard J Sugrue for valuable feedback.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. 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).
2. Popova, A.V.; Safina, K.R.; Ptushenko, V.V.; Stolyarova, A.V.; Favorov, A.V.; Neverov, A.D.; Bazykin, G.A. Allele-specific nonstationarity in evolution of influenza A virus surface proteins. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 21104–21112. [[CrossRef](#)]
3. Smith, D.J.; Lapedes, A.S.; De Jong, J.C.; Bestebroer, T.M.; Rimmelzwaan, G.F.; Osterhaus, A.D.; Fouchier, R.A. Mapping the antigenic and genetic evolution of influenza virus. *Science* **2004**, *305*, 371–376. [[CrossRef](#)]
4. Belongia, E.A.; McLean, H.Q. Influenza Vaccine Effectiveness: Defining the H3N2 Problem. *Clin. Infect. Dis.* **2019**, *69*, 1817–1823. [[CrossRef](#)]
5. Belongia, E.A.; Simpson, M.D.; King, J.P.; Sundaram, M.E.; Kelley, N.S.; Osterholm, M.T.; McLean, H.Q. Variable influenza vaccine effectiveness by subtype: A systematic review and meta-analysis of test-negative design studies. *Lancet Infect. Dis.* **2016**, *16*, 942–951. [[CrossRef](#)]
6. Rizzo, C.; Gesualdo, F.; Loconsole, D.; Pandolfi, E.; Bella, A.; Orsi, A.; Tozzi, A.E. Moderate Vaccine Effectiveness against Severe Acute Respiratory Infection Caused by A(H1N1)pdm09 Influenza Virus and No Effectiveness against A(H3N2) Influenza Virus in the 2018/2019 Season in Italy. *Vaccines* **2020**, *8*, 427. [[CrossRef](#)]
7. McLean, H.Q.; Belongia, E.A. Influenza Vaccine Effectiveness: New Insights and Challenges. *Cold Spring Harb. Perspect. Med.* **2020**, *11*, a038315. [[CrossRef](#)]
8. Monto, A.S.; Petrie, J.G. Improving influenza vaccine effectiveness: Ways to begin solving the problem. *Clin. Infect. Dis.* **2019**, *69*, 1824–1826. [[CrossRef](#)]
9. Skowronski, D.M.; Chambers, C.; Sabaiduc, S.; De Serres, G.; Winter, A.L.; Dickinson, J.A.; Li, Y. A Perfect Storm: Impact of Genomic Variation and Serial Vaccination on Low Influenza Vaccine Effectiveness During the 2014–2015 Season. *Clin. Infect. Dis.* **2016**, *63*, 21–32. [[CrossRef](#)]
10. Chen, H.; Alvarez JJ, S.; Ng, S.H.; Nielsen, R.; Zhai, W. Passage Adaptation Correlates with the Reduced Efficacy of the Influenza Vaccine. *Clin. Infect. Dis.* **2019**, *69*, 1198–1204. [[CrossRef](#)]
11. Smith, D.J.; Forrest, S.; Ackley, D.H.; Perelson, A.S. Variable efficacy of repeated annual influenza vaccination. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14001–14006. [[CrossRef](#)]
12. Francis, T. On the doctrine of original antigenic sin. *Proc. Am. Philos. Soc.* **1960**, *104*, 572–578.
13. Petrie, J.G.; Ohmit, S.E.; Truscon, R.; Johnson, E.; Braun, T.M.; Levine, M.Z.; Monto, A.S. Modest Waning of Influenza Vaccine Efficacy and Antibody Titers During the 2007–2008 Influenza Season. *J. Infect. Dis.* **2016**, *214*, 1142–1149. [[CrossRef](#)]
14. Skowronski, D.M.; Janjua, N.Z.; De Serres, G.; Sabaiduc, S.; Eshaghi, A.; Dickinson, J.A.; Li, Y. Low 2012–13 influenza vaccine effectiveness associated with mutation in the egg-adapted H3N2 vaccine strain not antigenic drift in circulating viruses. *PLoS ONE* **2014**, *9*, e92153. [[CrossRef](#)]

15. Ortiz de Lejarazu-Leonardo, R.; Montomoli, E.; Wojcik, R.; Christopher, S.; Mosnier, A.; Pariani, E.; Trilla Garcia, A.; Fickenscher, H.; Gärtner, B.C.; Jandhyala, R.; et al. Estimation of Reduction in Influenza Vaccine Effectiveness Due to Egg-Adaptation Changes-Systematic Literature Review and Expert Consensus. *Vaccines* **2021**, *9*, 1255. [[CrossRef](#)]
16. Kang, M.; Zanin, M.; Wong, S.S. Subtype H3N2 Influenza A Viruses: An Unmet Challenge in the Western Pacific. *Vaccines* **2022**, *10*, 112. [[CrossRef](#)] [[PubMed](#)]
17. Nielsen, R. Mapping mutations on phylogenies. *Syst. Biol.* **2002**, *51*, 729–739. [[CrossRef](#)]
18. Chen, H.; Deng, Q.; Ng, S.H.; Lee, R.T.C.; Maurer-Stroh, S.; Zhai, W. Dynamic Convergent Evolution Drives the Passage Adaptation across 48 Years' History of H3N2 Influenza Evolution. *Mol. Biol. Evol.* **2016**, *33*, 3133–3143. [[CrossRef](#)]
19. McWhite, C.D.; Meyer, A.G.; Wilke, C.O. Sequence amplification via cell passaging creates spurious signals of positive adaptation in influenza virus H3N2 hemagglutinin. *Virus Evol.* **2016**, *2*, vew026. [[CrossRef](#)]
20. Chen, T.G.C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
21. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **2017**, *1*, 33–46. [[CrossRef](#)]
22. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [[CrossRef](#)] [[PubMed](#)]
23. Zhai, W.; Slatkin, M.; Nielsen, R. Exploring variation in the d(N)/d(S) ratio among sites and lineages using mutational mappings: Applications to the influenza virus. *J. Mol. Evol.* **2007**, *65*, 340–348. [[CrossRef](#)]
24. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)] [[PubMed](#)]
25. Breiman, L. Random forests. *Machine Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
26. Lin, Y.; Wharton, S.A.; Whittaker, L.; Dai, M.; Ermetal, B.; Lo, J.; McCauley, J.W. The characteristics and antigenic properties of recently emerged subclade 3C.3a and 3C.2a human influenza A(H3N2) viruses passaged in MDCK cells. *Influenza Other Respir Viruses* **2017**, *11*, 263–274. [[CrossRef](#)]
27. Barr, I.G.; Donis, R.O.; Katz, J.M.; McCauley, J.W.; Odagiri, T.; Trusheim, H.; Wentworth, D.E. Cell culture-derived influenza vaccines in the severe 2017–2018 epidemic season: A step towards improved influenza vaccine effectiveness. *NPJ Vaccines* **2018**, *3*, 44. [[CrossRef](#)] [[PubMed](#)]
28. Wu, N.C.; Zost, S.J.; Thompson, A.J.; Oyen, D.; Nycholat, C.M.; McBride, R.; Wilson, I.A. A structural explanation for the low effectiveness of the seasonal influenza H3N2 vaccine. *PLoS Pathog.* **2017**, *13*, e1006682. [[CrossRef](#)]
29. Bonomo, M.E.; Deem, M.W. Predicting Influenza H3N2 Vaccine Efficacy from Evolution of the Dominant Epitope. *Clin. Infect. Dis.* **2018**, *67*, 1129–1131. [[CrossRef](#)] [[PubMed](#)]
30. Paules, C.I.; Sullivan, S.G.; Subbarao, K.; Fauci, A.S. Chasing Seasonal Influenza—The Need for a Universal Influenza Vaccine. *N. Engl. J. Med.* **2018**, *378*, 7–9. [[CrossRef](#)]
31. Takada, K.; Kawakami, C.; Fan, S.; Chiba, S.; Zhong, G.; Gu, C.; Kawaoka, Y. A humanized MDCK cell line for the efficient isolation and propagation of human influenza viruses. *Nat. Microbiol.* **2019**, *4*, 1268–1273. [[CrossRef](#)]
32. Aldeán J, Á.; Salamanca, I.; Ocaña, D.; Barranco, J.L.; Walter, S. Effectiveness of cell culture-based influenza vaccines compared with egg-based vaccines: What does the literature say? *Rev. Esp. Quimioter* **2022**, *35*, 241–248. [[CrossRef](#)]
33. Kuwahara, T.; Takashita, E.; Fujisaki, S.; Shirakura, M.; Nakamura, K.; Kishida, N.; Odagiri, T. Isolation of an Egg-Adapted Influenza A(H3N2) Virus without Amino Acid Substitutions at the Antigenic Sites of Its Hemagglutinin. *Jpn. J. Infect. Dis.* **2018**, *71*, 234–238. [[CrossRef](#)] [[PubMed](#)]
34. Lee, P.S.; Ohshima, N.; Stanfield, R.L.; Yu, W.; Iba, Y.; Okuno, Y.; Wilson, I.A. Receptor mimicry by antibody F045-092 facilitates universal binding to the H3 subtype of influenza virus. *Nat. Commun.* **2014**, *5*, 3614. [[CrossRef](#)] [[PubMed](#)]
35. Nuwarda, R.F.; Alharbi, A.A.; Kayser, V. An Overview of Influenza Viruses and Vaccines. *Vaccines* **2021**, *9*, 1032. [[CrossRef](#)] [[PubMed](#)]