SINGAPORE SCIENCE AND ENGINEERING FAIR

2011

Research Paper

Title of project: **Study of phylogenetic subclade 2.2.1 H5N1 viruses in Egypt**

Project Code: **MH12**

Name of student: **Chua Sin Ying**

School: **Hwa Chong Institution (College Section)**

Summary

H5N1 viruses of clade 2.2.1 first appeared in Egypt in 2006, resulting in an endemic. They are distinguished into two lineages, 2.2.1 proper (2.2.1p) and 2.2.1 variant (2.2.1v). Differences between 2.2.1p and 2.2.1 v viruses result in ineffectiveness of current H5N1 control measures. This study aims to examine these genetic differences. It will employ bioinformatics methods to analyze viral protein sequences and identify characteristic mutations of 2.2.1v viruses. Results confirm the existence of distinct phylogenetic differences, which requires revision of control measures in Egypt, and characteristic mutations which can affect viral biology, highlighting the need for further research.

Study of phylogenetic subclade 2.2.1 H5N1 viruses in Egypt

Abstract -

The highly pathogenic avian influenza (HPAI) H5N1 virus is a circulating Influenza strain capable of infecting the avian species as well as humans. H5N1 virus of phylogenetic clade 2.2.1 first appeared in Egypt in 2006 and has since been declared an endemic. H5N1 viruses in Egypt are distinguished into two lineages 2.2.1 proper (2.2.1p) and 2.2.1 variant (2.2.1v). Definite phenotypic differences exist between viruses of these two lineages, affecting the effectiveness of current H5N1 control measures in Egypt. This study aims to examine the genetic differences between 2.2.1p and 2.2.1v. It will study the phylogenetic differences between viruses of both lineages and aim to identify the characteristic mutations of 2.2.1v viruses. As such, it will also explore the structural positions and corresponding effects of the characteristic mutations. This research employs bioinformatics methods to analyze hemagglutinin (HA) sequences of the H5N1 virus obtained from both humans and avian species. Sequences were obtained from NCBI Influenza Virus Resource. Perl, a programming language, was used to compare protein sequences of both sub-clades and locate marker mutations and other programmes, such as MEGA 5, was used to further analyze the sequences. This study reports obvious phylogenetic differences between 2.2.1p and 2.2.1v viruses, and the gradual change of dominating strains from 2.2.1p to 2.2.1v requires a change in current vaccine strain selection. I170T, D62N and S139N are the 3 characteristic mutations determined. Deeper analysis has hinted that they can possibly affect viral biology, which warrants further research into the area.

1 INTRODUCTION

Since the end of 2003, a total of 63 countries reported cases of H5N1 in poultry and wild birds and 500 laboratory-confirmed human cases, out of which 296 died. The spread of H5N1 virus is a serious cause for concern due to its high case fatality rate of 59%, a H5N1 pandemic would certainly result in alarming mortality rates and economic loss [1]. In 2006, H5N1 viruses belonging to phylogenetic clade 2.2, subclade 2.2.1, were found in domestic poultry in Egypt. Subsequently, the number of cases escalated and an endemic status was declared by the Egyptian government in 2008. Human infection cases peaked from 2009 to 2010 and as of 4th March 2010, there were 104 human cases, out of which 30 were fatal [2]. H5N1 viruses in Egypt belong to 2 lineages: 2.2.1 proper (2.2.1p) and 2.2.1 variant (2.2.1v).

This study investigates the genetic differences between 2.2.1p and 2.2.1v viruses. This is crucial as 2.2.1v viruses are genetically different from 2.2.1p viruses and hence current vaccines, which are formulated for 2.2.1p viruses, would not be effective towards the variant lineage.

According to Abdelwhab *et. al*, 2.2.1v viruses are able to circulate in vaccinated chickens and escape identification by reverse transcriptase polymerase chain reaction (RT-qPCR) [3]. Research by Kim *et. al* also reported failure of commercially available H5 vaccines towards 2.2.1v viruses [4]. In its update on September 2010, the WHO expressed that current H5N1 vaccines made from vaccine strains A/Egypt/2321-NAMRU3/2007 and A/Egypt/3300-NAMRU3/2007 were ineffective towards a variant lineage of H5N1 viruses present in Egypt. Rather, the A/Egypt/1394-NAMRU3/2007 strain was more similar antigenically to the 2.2.1v viruses and hence proposed to be the next vaccine strain [5].

Despite its potential severity, current research has yet to explore 2.2.1v viruses extensively. Recent studies solely focus on the vaccine susceptibility of 2.2.1v viruses which examines a limited number of viral samples. This study hopes to address the gaps in current understanding by considering a large amount of H5N1 viral samples obtained from Egypt from 2006 to 2010.

2 AIMS / OBJECTIVES

This report aims to examine the genetic differences between 2.2.1p and 2.2.1v H5N1 viruses. Through this, a greater understanding towards 2.2.1v viruses can be established, thus contributing towards formulating effective control measures against this variant lineage. Furthermore, by understanding the genetic variance between the two lineages, it is also possible to assess the effectiveness of current vaccine strains as well as the proposed vaccine strain. Both contributions would be able to aid in controlling the spread of H5N1 in Egypt.

3 METHODOLOGY / MATERIALS

This research studies hemagglutinin (HA) protein sequences of H5N1 viruses collected in Egypt. These sequences were obtained from an online database, NCBI[1] Influenza Virus Resource [6], and were collected from January 2006 to December 2010. This time period would include the most recent viral sequences available as of 22nd December 2010. The sequences would be considered in time periods of 1 year[2]. A total of 289 HA sequences were used in this assay. As the HA is an essential glycoprotein which is involved in the tropism and pathogenesis of Influenza viruses, evaluating HA sequences would be most appropriate [7]. This study is inevitably limited by the availability of sequences in the database. As not all confirmed H5N1 cases in Egypt would be sequenced, the data is by no means complete. However, as these HA sequences were submitted by Egyptian health authorities, they are a substantial representation of H5N1 status in the country.

[1] National Center for Biotechnology Information

[2] January 2006 – December 2006, January 2007 – December 2007, January 2008 – December 2008, January 2009 – December 2009, January 2010 – December 2010

Subsequently, MAFFT[3] [8], a multiple sequence alignment program, was employed to compare the sequences and align them to an optimal average length. The FFT-NS-i option was selected, which is the most accurate strategy available in MAFFT. This would ensure that the sequences are of similar lengths and hence allow for inter-sequence comparison to locate substitutions at specific positions.

A phylogenetic tree was then generated using MEGA[4] 5 [9] based on Maximum Likelihood (ML). ML, the statistical method used in this study, is an intensive and established approach that estimates evolutionary distances between sequences. The Tamura-Nei model was also used to estimate DNA sequence evolution [10]. The phylogenetic tree was constructed by inferring phylogenetic relationships between sequences, thus genetically similar sequences would be placed together. Sequences were then labelled based on the year when they were collected. It is hence possible to observe how H5N1 viruses have changed genetically in Egypt since 2006 and the degree of similarity between viruses.

To determine the exact genetic differences between 2.2.1p and 2.2.1v viruses, a Perl script [11], (Appendix A), was written to compare each sequence with the vaccine strain, A/Egypt/2321-NAMRU3/2007, to locate mutations. This strain was used for inter-sequence comparison as it is one of the earliest vaccine strains used in Egypt and it belongs to the 2.2.1p lineage. Thus, this comparison would be able to detect genetic differences between later H5N1 viruses and earlier ones. Mutations were identified based on amino acids substituted in the sequence, and those with frequencies of occurrence of above 10% were termed as marker mutations (mutations which are statistically significant). A percentage-time graph was generated to portray the frequencies of occurrence of marker mutations. From this graph, it would be possible to identify characteristic marker mutations that differentiate 2.2.1v viruses from 2.2.1p viruses.

3D structures of these characteristic marker mutations were constructed using YASARA[5] [12]. Their positions were marked out in the structures to enable further analysis in terms of phylogenetic effects. Interactions between ligands and residues were examined using Ligand Explorer [13], a program integrated into the RCSB Protein Data Bank [14].

[3] Multiple Alignment using Fast Fourier Transform
[4] Molecular Evolutionary Genetics Analysis
[5] Yet Another Scientific Artificial Reality Application

4 RESULTS / DISCUSSION - Phylogenetic analysis

The phylogenetic tree (Appendix B) displays two distinct phylogenetic clades, the 2.2.1p clade situated above and the 2.2.1v clade below. Identity of the clades was confirmed by referring to the location of the vaccine strains. It can be inferred that the dominant circulating strain has changed from 2.2.1p to 2.2.1v over the 5 years. A large majority of viruses obtained in 2009 and 2010 belonged to the 2.2.1v lineage, while most viruses obtained from 2006 to 2008 were 2.2.1p viruses. This change of dominant strains was gradual. From 2006 to 2007, the dominant strain was similar to the earlier 2.2.1p vaccine strain, A/Egypt/2321-NAMRU3/2007. In 2008, the dominant strain was situated on a different clade which contains later 2.2.1p vaccine strain, A/Egypt/3300-NAMRU3/2007. This clade is in fact genetically closer to the 2.2.1v clade. Finally, in 2009 and 2010, the dominant strain became of the 2.2.1v lineage. This would mean that 2.2.1v is now the dominant lineage circulating Egypt. This is a cause for concern as current H5N1 control measures in Egypt are ineffective and hence the endemic may worsen further.

Interestingly, it can be observed that human cases of H5N1 were concentrated in the 2.2.1v clade. This would mean that 2.2.1v viruses are more capable of infecting humans compared to previous 2.2.1p viruses. This observation is in good agreement with case statistics. From 2009 to 2010, the number of human infection cases peaked and these two years coincide with the period when 2.2.1v viruses began dominating. Thus, the rise of 2.2.1v viruses is a major concern as humans are more susceptible to infection by these viruses. Coupled with the H5N1 virus' high case fatality rate, the endemic in Egypt may become more widespread and mortality rate may increase.

The two vaccine strains for 2.2.1p differ genetically from current dominant 2.2.1v viruses. Thus, they are no longer applicable in Egypt. On the contrary, the new vaccine strain for 2.2.1v under consideration, A/Egypt/1394-NAMRU3/2007, is largely genetically similar to the 2.2.1v viruses circulating from 2009 to 2010. Furthermore, the vaccine strain is seen to be situated amongst viruses found in human infection cases, thus it would be expected to be especially effective in controlling the worsening of the H5N1 pandemic in Egypt.

In summary, the phylogenetic analysis reports that 2.2.1p and 2.2.1v viruses are highly different genetically as they belong to clades that are situated far away from each other. The dominant circulating H5N1 strain in Egypt has changed from 2.2.1p to 2.2.1v lineage in recent years and 2.2.1v viruses are also shown to be more capable of transmitting to humans. Thus attention has to be given to developing new control and prevention measures so as to curb the worsening of the endemic in Egypt.

4 RESULTS / DISCUSSION - Genetic differences

From the percentage-time graph, (Appendix C) 3 distinct categories of marker mutations can be seen: marker mutations due to vaccine strain selection, emerging marker mutations and random mutations. Only emerging marker mutations will be discussed due to their biological significance[6].

Emerging marker mutations are mutations that possess increasing trends of occurrence. These mutations were absent when the H5N1 virus first appeared in 2006 and remained insignificant until 2009 and 2010, when they increased to frequencies of occurrence of above 50%.

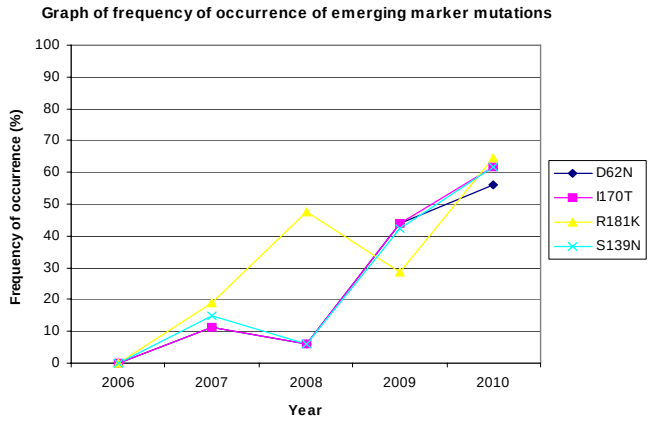


Fig. 1: Frequency of occurrence of emerging marker mutations in H5N1 viruses in Egypt

The 4 emerging marker mutations, D62N, I170T, R181K and S139N are presented separately in Fig. 1. These emerging mutations are likely characteristic mutations which differentiate 2.2.1p viruses from 2.2.1v viruses. To confirm this, the 2.2.1p vaccine strain was compared to the 2.2.1v vaccine strain. As the vaccine strains are representative of viruses in that clade, this assay would be able to identify general marker mutations between 2.2.1p and 2.2.1v viruses. The mutations found between vaccine strains were then categorized based on their trends of occurrence in the percentage-time graph (table found in Appendix E).

Mutations I170T, D62N and S139N were again detected in this assay. Although R181K was previously identified from the percentage-time graph, it was not detected when comparing vaccine strains. Furthermore, its trend of occurrence is unlike the usual rise of 2.2.1v viruses, as it peaks in 2008. Hence it would be excluded as a characteristic mutation.

The 3 characteristic mutations I170T, D62N and S139N are displayed in the 3D structure of the H5N1 HA protein (Appendix F).

I170T - The mutation I170T is a substitution of isoleucine by threonine at position 170 in the HA protein. Position 170 (corresponding to position 155 in the literature) is found in the sialic acid binding site, where the virus interacts with host cell receptors which contain sialic acid residues.

[6] Explanations on marker mutations due to vaccine strain selection and random mutations can be found in Appendix D

This allows the virus to bind to the host cell and gain entry into the cell [15] The residue at 170 is located in close proximity of the bound sialic acid, as shown in Fig. 2. The amino acid at position 170 has been shown to form strong hydrophobic interactions with sialic acids. Shown in Fig. 3, the distance between the sialic acid and residue at position 170 is only 2.73Å. As weak chemical bonds are normally formed within the distance of 5Å, the presence of hydrophobic interactions between the residue at position 170 and sialic acid is highly probable [16].

This highlights the importance of the residue at position 170 with regards to host cell recognition. The virus' ability to recognise and bind to host cell receptors may be altered by the mutation. The substitution of isoleucine by threonine resulted in the deletion of a hydrophobic methyl group and an addition of a hydrophilic hydroxyl group. Thus, hydrophobic interactions may likely diminish and instead replaced by hydrogen bonds, as the hydroxyl group is electronegative and able to participate in hydrogen bonding. As I170T has been identified as a characteristic mutation and it plays a role in the virus' virulence and receptor binding, it may be a contributing factor towards the variant virus' greater ability to infect humans. Changes in this position may have resulted in alterations of the way the virus associates with human cell receptors. This would need to be confirmed by specific experiments and further research which can observe the mutations under controlled and constant circumstances.

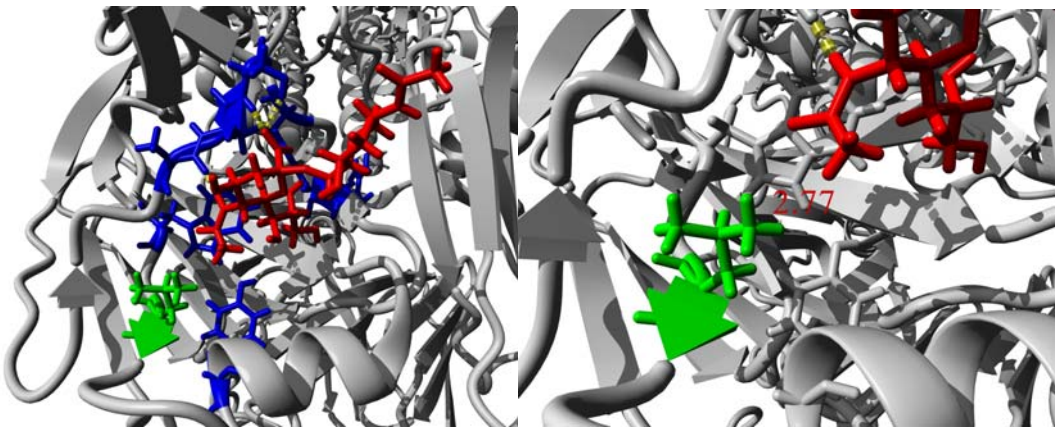
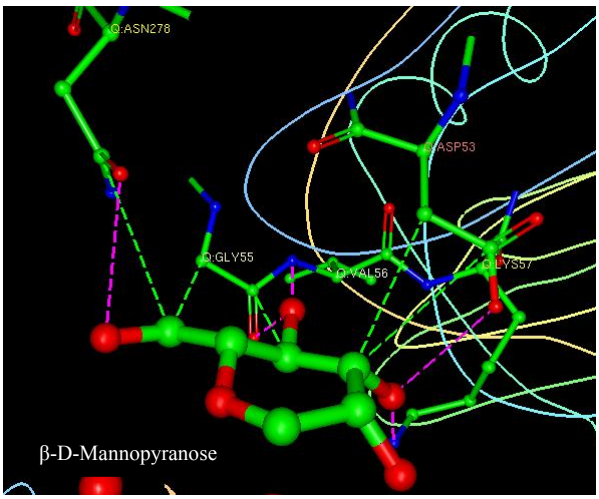


Figure 2 (left): 3D structure of HA sialic acid binding site and bound sialic acid. The binding site is highlighted in blue, with I170 marked out in green. Sialic acid is labelled red.
Figure 3 (right): 3D structure of I170 with bound sialic acid. The residue at position 170 is highlighted in green while the sialic acid is coloured red. The distance between both is 2.73Å. While hydrogen bonds are not detected, hydrophobic interactions are likely.

D62N - The mutation D62N is an amino acid substitution of aspartic acid by asparagine at position 62. Position 62 (corresponding to position 53 in the literature) is found to be involved in the binding of β-D-Mannopyranose, a polysaccharide usually found on the surface of pathogens [17]. The presence of mannose sugars on the surface of the HA protein is essential for the antiviral effects of

mannose binding lectin (MBL). MBL is an innate immune protein that recognizes mannose sugars on the glycoproteins of Influenza viruses, thus initiating a series of immune responses such as viral aggregation, opsonization and neutralization [18]. It has been reported that a deficiency in MBL responses will result in failure of the immune system to inhibit Influenza viruses [19].

As shown in Appendix F, position 62 is located at the surface of the HA protein, hence it is likely to participate in various interactions with other molecules [20]. Upon detailed examination, the residue at site 62 is found to be able to bind with β-D-Mannopyranose via hydrophobic interactions, as displayed in Fig. 4. The distance between the residue at position 62 and β-D-Mannopyranose is shown to be 2.94Å in Fig. 5. Therefore, as the residue at position 62 is found to be involved in the binding of one of the oligosaccharides recognized by MBL, a mutation at this position would likely affect its ability to bind to the virus. This may affect the immune system's ability to inhibit these mutant viruses. As the substitution of aspartic acid by asparagine would result in the removal of a negative charge, hydrophobic interactions between the residue and β-D-Mannopyranose would be expected to be affected. However, exact effects of the D62N mutation are still unknown as no specific laboratory examinations have been conducted on it.



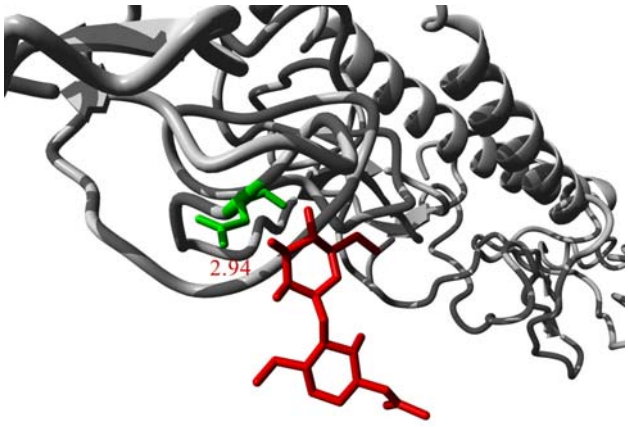


Figure 4 (left): Interactions between β-D-Mannopyranose and HA protein. Hydrophobic interactions are labelled in green and hydrogen bonds are labelled in magenta. The aspartic acid at position 53 (corresponding to 62 in this study) is found to possess hydrophobic interactions with the polysaccharide.
Figure 5 (right): The distance between the residue at position 62, highlighted in green, and the β-D-Mannopyranose, highlighted in red, is shown to be 2.94Å.

S139N - The mutation S139N is an amino acid substitution of serine by asparagine at position 139 (corresponding to position 124 in the literature). This position has been found to be located in a known antigenic site of the HA protein [21]. The antigenic site is essential for antibodies to recognize and bind to the viruses, thus activating various antiviral reactions. Through research which studies the binding of monoclonal antibodies (mAbs) to the virus, it is discovered that the

residue at position 139 plays a crucial role in the binding of the antibody [22]. It is also one of the targets for neutralizing antibodies as well as a carbohydrate recognition site for host cell receptors [23]. Therefore, mutations at the position are expected to alter the ability of antibodies to bind to the HA of the virus. Li *et. al* reported that mutations at position 139 are expected to improve the immunogenicity of the viruses [24].

Apart from its location in the antigenic site, the residue at position 139 is also shown to be under positive selection pressure. Mutations occur often at sites which are under positive selection, because they are likely to confer favourable characteristics to the viruses [25]. This supports the conclusion that S139N is a characteristic mutation of 2.2.1v viruses. Although the residue at position 139 is not shown to be directly involved in the binding of antibodies, it is situated inside the antigenic binding site. Furthermore, research has shown that mutations in this area can affect the virus' ability to escape the antiviral activities of the immune system. Therefore, the residue at position 139 should be researched in greater details in order to determine its actual role in the binding of antibodies and hence understand the effects of the S139N mutation.

In summary, all 3 mutations are found to be located at surface positions involved in the binding of various ligands. Thus it is probable that the additive effect of these mutations have resulted in 2.2.1v viruses being more capable of infecting humans, through affecting the interactions between human cell receptors or antibodies. However, definite conclusions can only be made after observing these mutations in controlled environments.

6 CONCLUSIONS

This study provides new insights into the 2.2.1v H5N1 viruses of Egypt. The phylogenetic analysis reports that current H5N1 control measures should be revised as 2.2.1p and 2.2.1v viruses are phylogenetically different. I170T, D62N and S139N were the 3 most probable characteristic mutations identified. They are located in essential sites in the HA protein, hence they may confer the virus with phenotypic changes. This study is still limited by the availability of sequences in the databases, which may affect the validity of the results. Future improvements are to increase the number of sequences by considering more sources and types of viral data. For example, HI assays and PCR tests can be carried out on patients to provide clinical statistics of 2.2.1v viruses. Additional databases such as GISAID EpiFlu [26] can be explored to obtain more sequences. Moreover, to effectively conclude the validity of results, laboratory tests have to be carried out under controlled and specific conditions. However, this study provides an edge in identifying genetic differences between 2.2.1p and 2.2.1v viruses as many viral samples were examined in the process, hence obtaining an accurate and comprehensive coverage.

ACKNOWLEDGEMENTS

The author would like to thank A*STAR mentor Dr. Sebastian Maurer-Stroh, Hwa Chong Institution teacher mentors Mr. Li Jiexun and Mr. Chia Keng Hian for their invaluable guidance, Mr. Raphael Lee and staff of Bioinformatics Institute for their help and support and Bioinformatics Institute, A*STAR for the use of equipment.

REFERENCE

[1] A. Tarantola, P. Barboza, V. Gauthier, S. Ioos, N. El Omeiri, and M. Gastellu-Etchegorry, "The influenza A(H5N1) epidemic at six and a half years: 500 notified human cases and more to come," *Euro Surveillance: Bulletin Européen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, vol. 15, 2010.

[2] S.H. Taha, M.A. Mehrez, M.Z. Sitohy, A.G.I. Abou Dawood, M.M. Abd-El Hamid, and W.H. Kilany, "Effectiveness of esterified whey proteins fractions against Egyptian Lethal Avian Influenza A (H5N1)," *Virology Journal*, vol. 7, 2010, p. 330.

[3] E.M. Abdelwhab, A.M. Erfan, C. Grund, M. Ziller, A. Arafa, M. Beer, M.M. Aly, H.M. Hafez, and T.C. Harder, "Simultaneous detection and differentiation by multiplex real time RT-PCR of highly pathogenic avian influenza subtype H5N1 classic (clade 2.2.1 proper) and escape mutant (clade 2.2.1 variant) lineages in Egypt," *Virology Journal*, vol. 7, 2010, p. 260.

[4] J. Kim, G. Kayali, D. Walker, H.L. Forrest, A.H. Ellebedy, Y.S. Griffin, A. Rubrum, M.M. Bahgat, M.A. Kutkat, M.A.A. Ali, J.R. Aldridge, N.J. Negovetich, S. Krauss, R.J. Webby, and R.G. Webster, "Puzzling inefficiency of H5N1 influenza vaccines in Egyptian poultry," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, Jun. 2010, pp. 11044-11049.

[5] "Antigenic and genetic characteristics of influenza A(H5N1)and influenza A(H9N2) viruses and candidate vaccine viruses developed for potential use in human vaccines," *Relevé Épidémiologique Hebdomadaire / Section D'hygiène Du Secrétariat De La Société Des Nations = Weekly Epidemiological Record / Health Section of the Secretariat of the League of Nations*, vol. 85, Oct. 2010, pp. 418-424.

[6] Y. Bao, P. Bolotov, D. Dernovoy, B. Kiryutin, L. Zaslavsky, T. Tatusova, J. Ostell, and D. Lipman, "The influenza virus resource at the National Center for Biotechnology Information," *Journal of Virology*, vol. 82, Jan. 2008, pp. 596-601.

[7] R.P. de Vries, E. de Vries, K.S. Moore, A. Rigter, P.J.M. Rottier, and C.A.M. de Haan, "Only two residues are responsible for the dramatic difference in receptor binding between swine and new pandemic H1 hemagglutinin," *The Journal of Biological Chemistry*, Dec. 2010.

[8] K. Katoh and H. Toh, "Parallelization of the MAFFT multiple sequence alignment program," *Bioinformatics (Oxford, England)*, vol. 26, Aug. 2010, pp. 1899-1900.

[9] S. Kumar, M. Nei, J. Dudley, and K. Tamura, "MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences," *Briefings in Bioinformatics*, vol. 9, Jul. 2008, pp. 299-306.

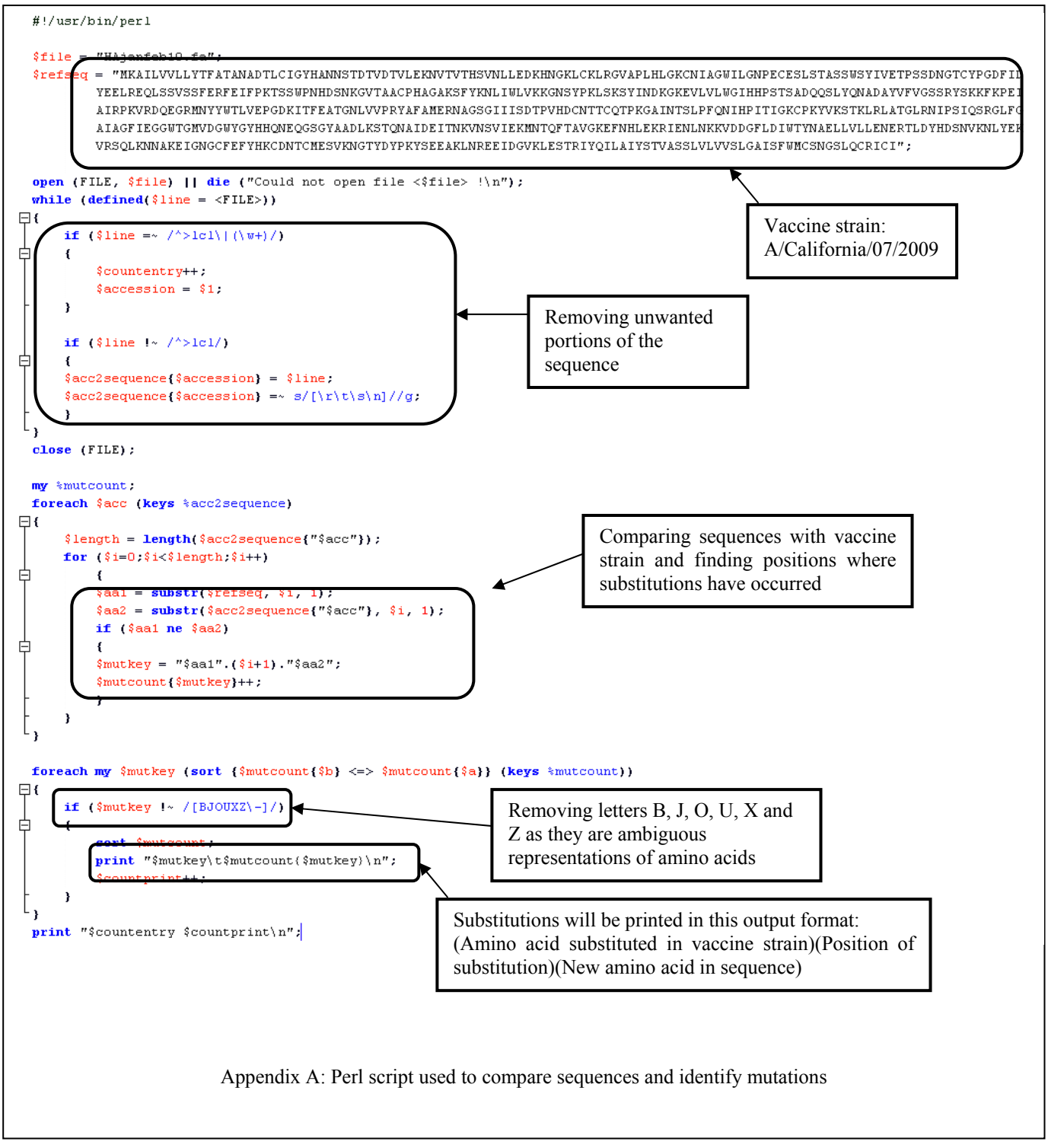
[10] A. Hobolth and R. Yoshida, "Maximum likelihood estimation of phylogenetic tree and substitution rates via generalized neighbor-joining and the EM algorithm," *Algebraic Biology 2005, Computer Algebra in Biology*, vol. 1, 2005, pp. 41-50.

[11] B.K.H. Chan, "Perl 5 Tutorial," 2003.

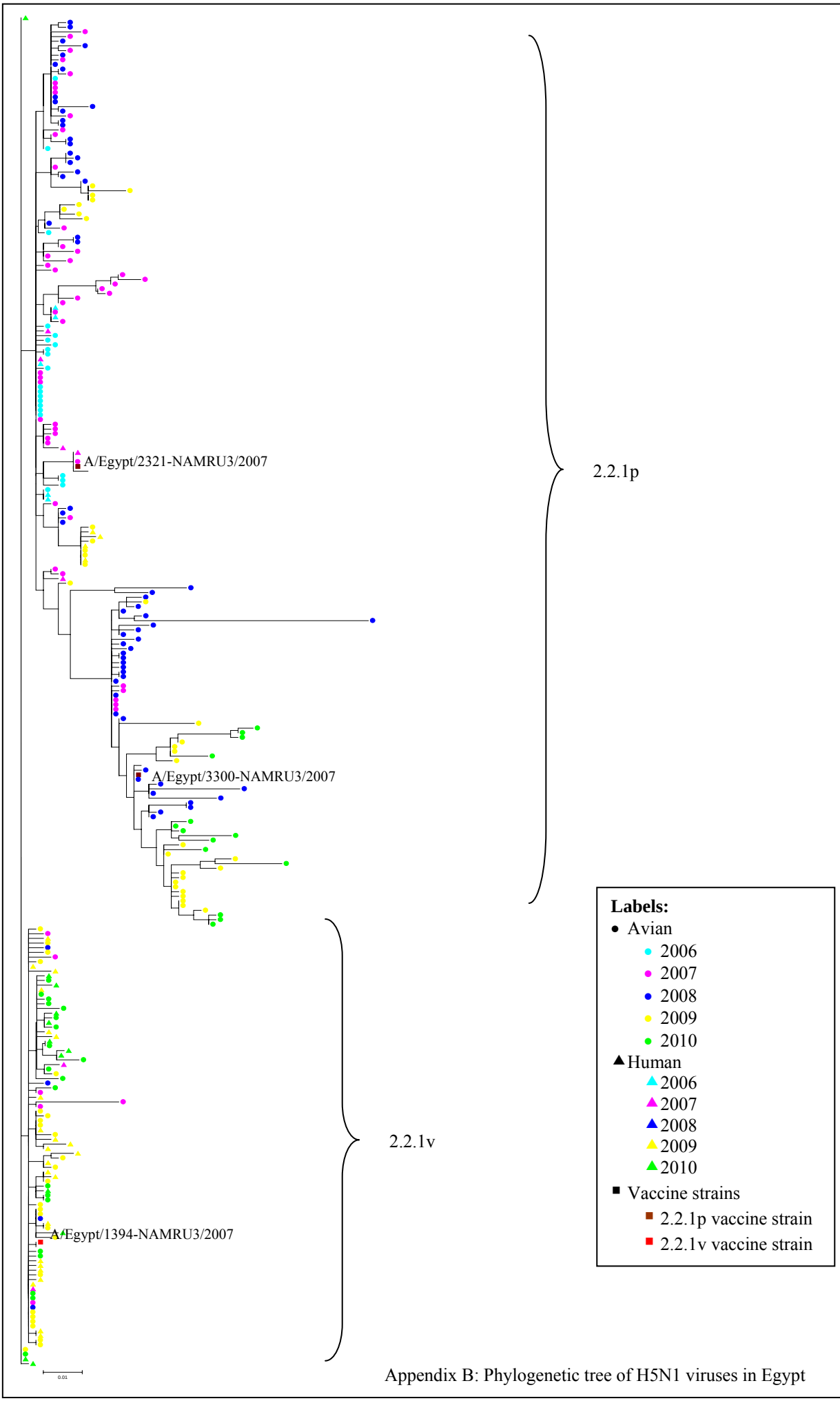
[12] E. Krieger, G. Koraimann, and G. Vriend, "Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field," *Proteins*, vol. 47, May. 2002, pp.

393-402.
[13] C. Zhang, *Ligand Explorer*, San Diego Supercomputer Center, 2004.
[14] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, Jan. 2000, pp. 235-242.
[15] M. Li and B. Wang, "Computational studies of H5N1 hemagglutinin binding with SA-alpha-2, 3-Gal and SA-alpha-2, 6-Gal," *Biochemical and Biophysical Research Communications*, vol. 347, Sep. 2006, pp. 662-668.
[16] G. Feverati and C. Lesieur, "Oligomeric interfaces under the lens: gemini," *PloS One*, vol. 5, 2010, p. e9897.
[17] E.C. van Asbeck, A.I.M. Hoepelman, J. Scharringa, B.L. Herpers, and J. Verhoef, "Mannose binding lectin plays a crucial role in innate immunity against yeast by enhanced complement activation and enhanced uptake of polymorphonuclear cells," *BMC Microbiology*, vol. 8, 2008, p. 229.
[18] P.C. Reading, J.L. Miller, and E.M. Anders, "Involvement of the mannose receptor in infection of macrophages by influenza virus," *Journal of Virology*, vol. 74, Jun. 2000, pp. 5190-5197.
[19] W. Chang, M.R. White, P. Moyo, S. McClear, S. Thiel, K.L. Hartshorn, and K. Takahashi, "Lack of the pattern recognition molecule mannose-binding lectin increases susceptibility to influenza A virus infection," *BMC Immunology*, vol. 11, Dec. 2010, p. 64.
[20] H.X. Zhou and Y. Shan, "Prediction of protein interaction sites from sequence profile and residue neighbor list," *Proteins*, vol. 44, Aug. 2001, pp. 336-343.
[21] J. Taylor, "New Strain of Avian Influenza Now Found to Infect Humans," *Medscape Medical News*, Mar. 2006.
[22] N.V. Kaverin, I.A. Rudneva, N.A. Ilyushina, N.L. Varich, A.S. Lipatov, Y.A. Smirnov, E.A. Govorkova, A.K. Gitelman, D.K. Lvov, and R.G. Webster, "Structure of antigenic sites on the haemagglutinin molecule of H5 avian influenza virus and phenotypic variation of escape mutants," *The Journal of General Virology*, vol. 83, Oct. 2002, pp. 2497-2505.
[23] N.A. Ilyushina, I.A. Rudneva, A.S. Gambaryan, A.B. Tuzikov, and N.V. Bovin, "Receptor specificity of H5 influenza virus escape mutants," *Virus Research*, vol. 100, Mar. 2004, pp. 237-241.
[24] J. Li, Y. Wang, Y. Liang, B. Ni, Y. Wan, Z. Liao, K. Chan, K. Chan, K. Yuen, X. Fu, X. Shang, S. Wang, D. Yi, B. Guo, B. Di, M. Wang, X. Che, and Y. Wu, "Fine antigenic variation within H5N1 influenza virus hemagglutinin's antigenic sites defined by yeast cell surface display," *European Journal of Immunology*, vol. 39, Dec. 2009, pp. 3498-3510.
[25] L. Campitelli, M. Ciccozzi, M. Salemi, F. Taglia, S. Boros, I. Donatelli, and G. Rezza, "H5N1 influenza virus evolution: a comparison of different epidemics in birds and humans (1997-2004)," *The Journal of General Virology*, vol. 87, Apr. 2006, pp. 955-960.
[26] The GISAID Foundation, "GISAID EpiFlu Database," *GISAID*, 2008.

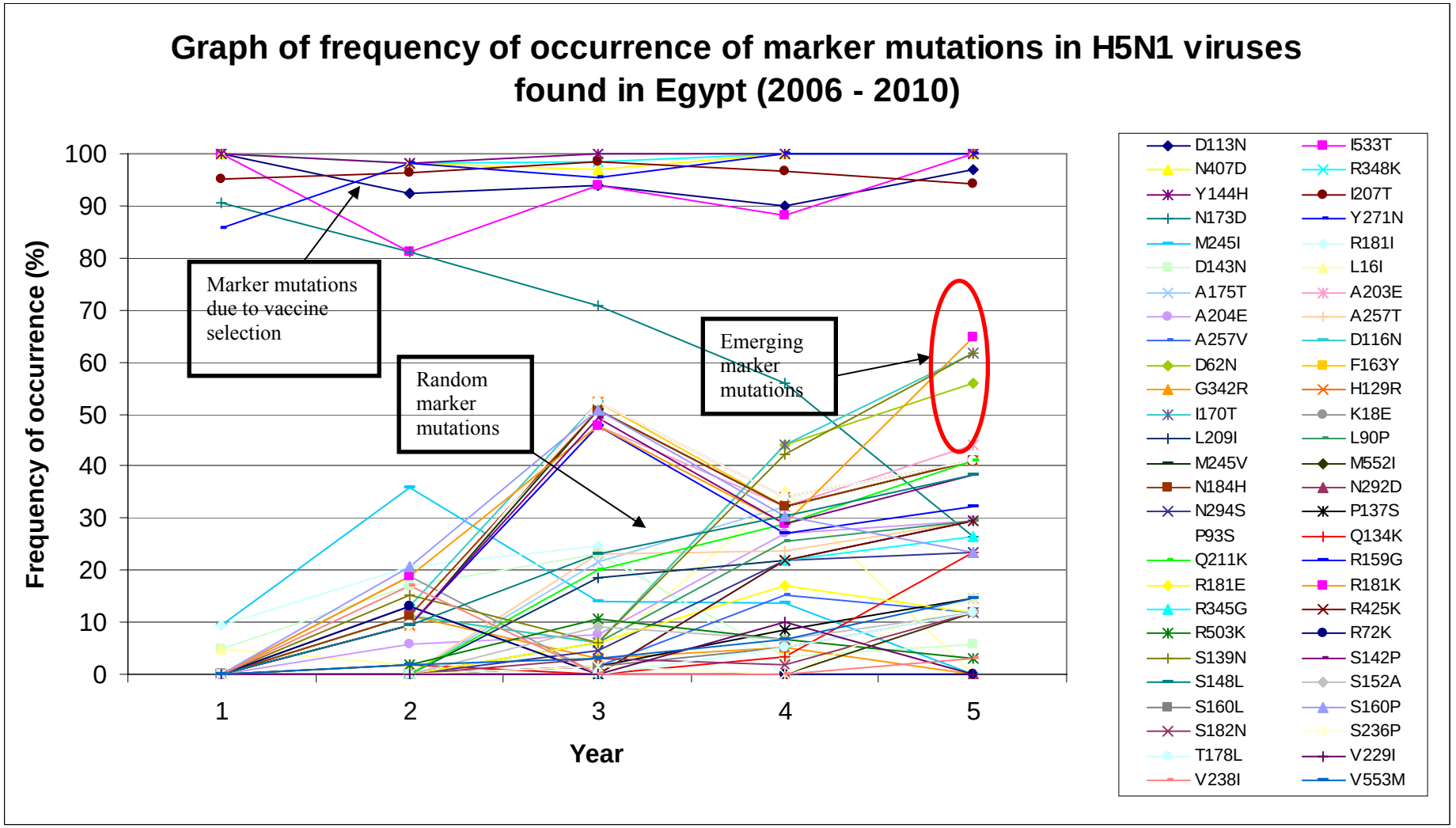
APPENDICES



Appendix A: Perl script used to compare sequences and identify mutations



Appendix B: Phylogenetic tree of H5N1 viruses in Egypt



Appendix C: Frequency of occurrence of marker mutation in H5N1 viruses in Egypt

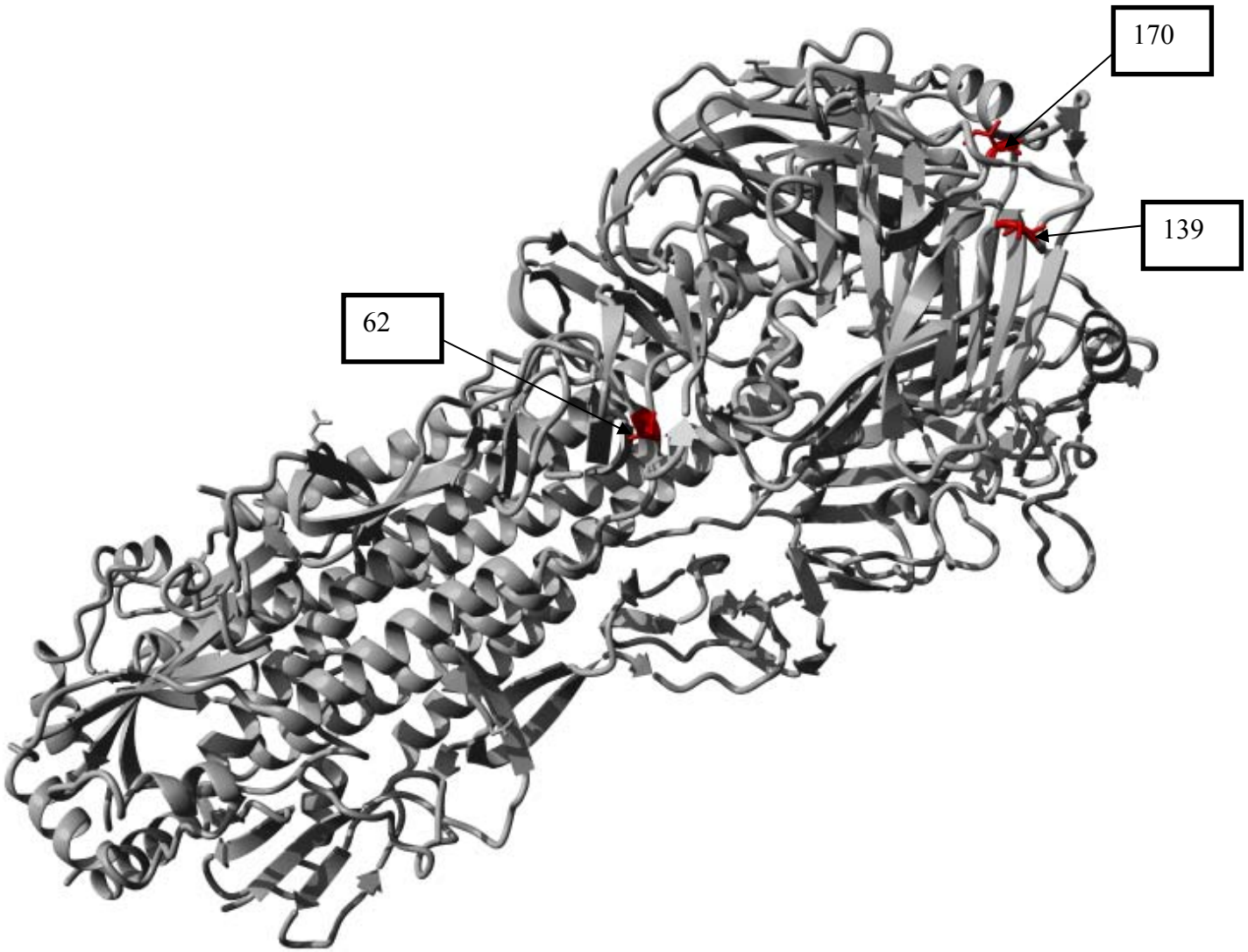
Appendix D

Some marker mutations presented in Appendix C maintained persistently high frequencies of occurrence throughout the 5 years. These mutations are more likely caused by consistent phylogenetic differences between the vaccine strain selected and the dominant circulating strains. As can be seen from the phylogenetic tree, genetic differences would always exist between different viruses. These consistently prevalent mutations are representative of the innate genetic differences between the vaccine strain and other viral sequences, thus they are not caused by the H5N1 viruses mutating but unavoidable discrepancies when selecting the vaccine strain.

Random mutations are mutations that have irregular and normally low frequencies of occurrence. These are random changes to the nucleotide sequence which likely do not possess phenotypic significance.

Mutation	Category
Y144H	Mutation due to vaccine selection
K172R	Not present in circulating strains
I170T	Emerging mutation
D113N	Mutation due to vaccine selection
N407D	Mutation due to vaccine selection
Y271N	Mutation due to vaccine selection
N173D	Diminishing mutation
R348K	Mutation due to vaccine selection
D62N	Emerging mutation
S139N	Emerging mutation
I533T	Mutation due to vaccine selection
F557S	Not present in circulating strains

Appendix E: Mutations present between A/Egypt/2321-NAMRU3/2007 and A/Egypt/1394-NAMRU3/2007 categorized according to frequencies of occurrence in percentage-time graph



Appendix F: 3D structure of HA protein in H5N1 Influenza virus. Positions 170, 62 and 139 are highlighted in red.