2013 13th International Conference on Computational Science and Its Applications

Influenza-A Circulation in Vietnam through Data Analysis of Hemagglutinin Entires

Ly Le
School of Biotechnology
International University (VNU-HCM)
Thu Duc District, HCMC, Vietnam
ly.le@hcmiu.edu.vn

Hong-Quang Nguyen
School of Computer Science and Engineering
International University (VNU-HCM)
Thu Duc District, HCMC, Vietnam
nhquang@hcmiu.edu.vn

***Abstract*— Hemagglutinin is a glycoprotein locating on surface of influenza virus. Influenza A human hemagglutinin protein H1, H2, H3, H5, H7, H9 sequences on NCBI database from Vietnam were retrieved and analyzed to find out the pattern of viral circulation within the country in the last 12 years. 1060 sequences were collected and analyzed. The results revealed five main influenza A virus including H5N1, H3N2, H1N1, H6N2, and H9N2 circulating in Vietnam with high level of genetic compatibility, identified as reassortants between a wide variety viral genotypes and share common mechanisms of spread. The most two popular host are human and poultry. The two recent pandemics avian H5N1 in 2005 and Swine H1N1 in 2009 did show significant effects on the viral circulation in Vietnam. H5N1 and H1N1 subtype remain pathogenic and require close surveillance.**

Keywords— Influenza-A Circulation, Viral circulation, Hemagglutinin Entires

I. INTRODUCTION

In the years 1918, 1957 and 1968, the world was attacked by influenza A pandemics H1N1, H2N2 and H3N2 accordingly. The number of casualties reached 50 to 100 million people worldwide. Characterized by their negative strand RNA genomes requiring an RNA dependent RNA polymerase of viral origin for replication, influenza viruses are capable of evading adaptive immune responses of many mammalian species through antigenic shift and antigenic drift. This shows that influenza viruses are much diverged.

Furthermore, sequencing studies show that influenza viruses have the same genetic ancestry. Especially, the surface glycoproteins, hemagglutinin (HA) [1] and neuraminidase (NA)[2] are specific for each subtype and therefore can be used for vaccine design and production. Up to date, there are 16 subtypes of HA and 9 subtypes of NA for influenza A viruses which have been found circulating the population. However, only three HA (H1, H2, H3) and 2 NA (N1 and N2) subtypes are found to cause human epidemics and pandemics.

Hemagglutinin (HA) is a viral surface glycoprotein. Hemagglutinin is the key protein in viral internalization. It facilitates the penetration of viral RNA by the binding between influenza A virus and host sialic acid decorated receptor. This protein also has a rapid evolution which leads to changes in its antigenic structure. However, during the evolutionary changes for host specific adaptation, some regions of hemaggluttin are receptor binding sites between host and virus are likely to maintain their structural properties.

Moreover, those conserved regions can be found to have other function such as facilitating the post translational modification, protein folding. Therefore, it is possible to hypothesize that conserved regions of hemagglutinin protein would be good candidates for rational design in universal vaccine engineering against influenza A virus.

Centers for Disease Control and Prevention (CDC) shows that there are two types of influenza: highly pathogenic and low pathogenic based on molecular genetics and pathogenesis criteria. Influenza A viruses are classified based on the two surface proteins: hemagglutinin (HA) and neuraminidase (NA). For example, H1N1 is understood as the influenza A subtype has HA 1 protein and NA 1 protein. HA has been targeted as a possible molecule for vaccine design and development. However, only few HA molecules in a total 16 HA molecules are truly causing danger for human. Those are H1, H2, H3, H5 and H7. In those, H1, H2 and H3 are currently circulating the human population.

This study aims to analyze the circulation of influenza A virus in Vietnam based on the entries of hemaglutin to NCBI from the country.This research is proposed to help Vietnamese people to find possible ways to minimize the negative effects of influenza A virus on agriculture and public health.

II. MATERIALS AND METHODS

Our research was conducted using the latest real dataset on hemagglutinin influenza A viruses sequences which have been collected in Vietnam and published on NCBI influenza virus resources [3, 7]. In total, 1060 sequences reported during 11 years (2001-2012) were analyzed for our experiments. The dataset was populated into a relational database using MySQL. Among these, there were 26 distinct subtypes, including H5N1, H3N2, H1N1, H6N2, H9N2 and other subtypes. Each

© IEEE 2013. This article is free to access and download, along with rights for full text and data mining, re-use and analysis.

reported case is examined based on 10 main characteristics: accession, flu name, sequence length, flu host, protein, subtypes, location/country (in this case, Vietnam), reporting date, gender, and specific host (for example, Muscovy duck or Chicken). From 2001 to 2012, 36 cities and provinces in Vietnam were found to be effected by virus infections. However, the ages in the reported cases were not documented. Only 71 cases clearly indicate the genders (male or female) of the cases while the genders in the other 989 cases were not reported.

The dataset has been collected by different groups of researchers and professionals, the data values of the dataset were heterogeneous with different uses of capitalization, white spaces and abbreviations. Therefore, the dataset is required to be pre-processed for further summarization.

Host names and locations were normalized so that the data entries could be consistent across the entire dataset. Host names, including "Chicken", were normalized from various data values such as "Ck", "chicken" and "Chicken". The proper names of 19 cities and provinces in Vietnam were normalized across the database. For example, the capital city "Ha Noi" written in various forms "Hanoi", "HaNoi", "HANOI" and "Ha Noi", was normalized into "Ha Noi".

After the dataset was normalized, tables were created on MySQL and populated with the corresponding data values. To generate various kinds of charts and graphs, different aggregate functions (including Average, Sum, Minimum, Maximum and Count) were applied to calculate these values as described in the subsequent sections.

III. Results and Discussion

Antigenic characteristics of hemagglutinin influenza A viruses sequences collected in Vietnam have shown highly emergence for a new epidemic or pandemic that are taking a large part in three main subtypes A(H1N1), A(H3N2) and A(H5N1) during 2001-2012 that totally possess 93% (Figure 1) all of others subtypes, in which Muscovy duck host is 5% involving 176 cases, chicken host is 17% with 176 cases, duck host is 26% with 278 cases and a largest percentage has been belonged to human host in 50% with 528 cases in report (Figure 2).

In 2003, the influenza A virus outbreak was recorded in poultry by A (H5N1) in Vietnam, specially, which has closely familiar with next two outbreaks during 2004 to 2005 with 245 cases (Figure 3). They were basis ground for evolution of new A(H5N1) strain that parallel existed with old strains in more than 20 provinces in 2007 (WHO). Additionally, the A (H5N1) strains isolated from human cases are often partially identical from poultry. Up to date, they are reservoir of A (H1N1), A (H3N2) A (H10N7), A (H11N3), A (H12N5) and other subtypes.

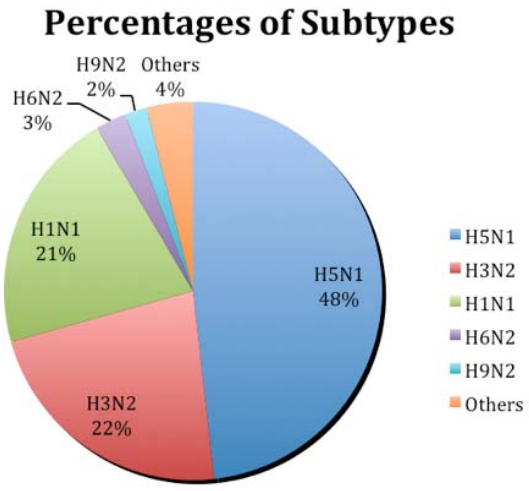

Figure 1: *The percentages of subtypes are mainly included by three subtypes: H5N1 (48%), H1N1 (21%) and H3N2 (22%). Remained subtypes are H6N2, H9N2 and others which take a little percentage.*

In first months of 2012, A (H5N1) viruses infected to human cases occurred in Quang Tri, Thanh Hoa, and Soc Trang. Beside, non-reported cases could be predicted come from Thai Nguyen, Bac Lieu, Ha Noi, Nghe An, Kien Giang. From 2003 to 2012, although A(H5N1) avian influenza viruses has played as predominant viruses for outbreaks in Vietnam, the diversity of genotypes present in Vietnam suggest a high level of genetic compatibility, identified as reassortants between a wide variety viral genotypes and share common mechanisms of spread. In 2006, for instance, new A (H1N1) strain had shared hemagglutinin structure with old "Spanish" A flu which was discovered in 1918, similar to human flu.

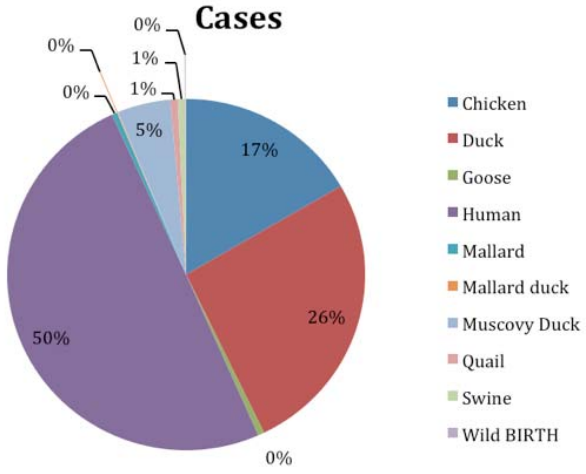

Figure 2: *The pie chart representative percentages of host cases from 2001 to 2012, in which Muscovy duck*

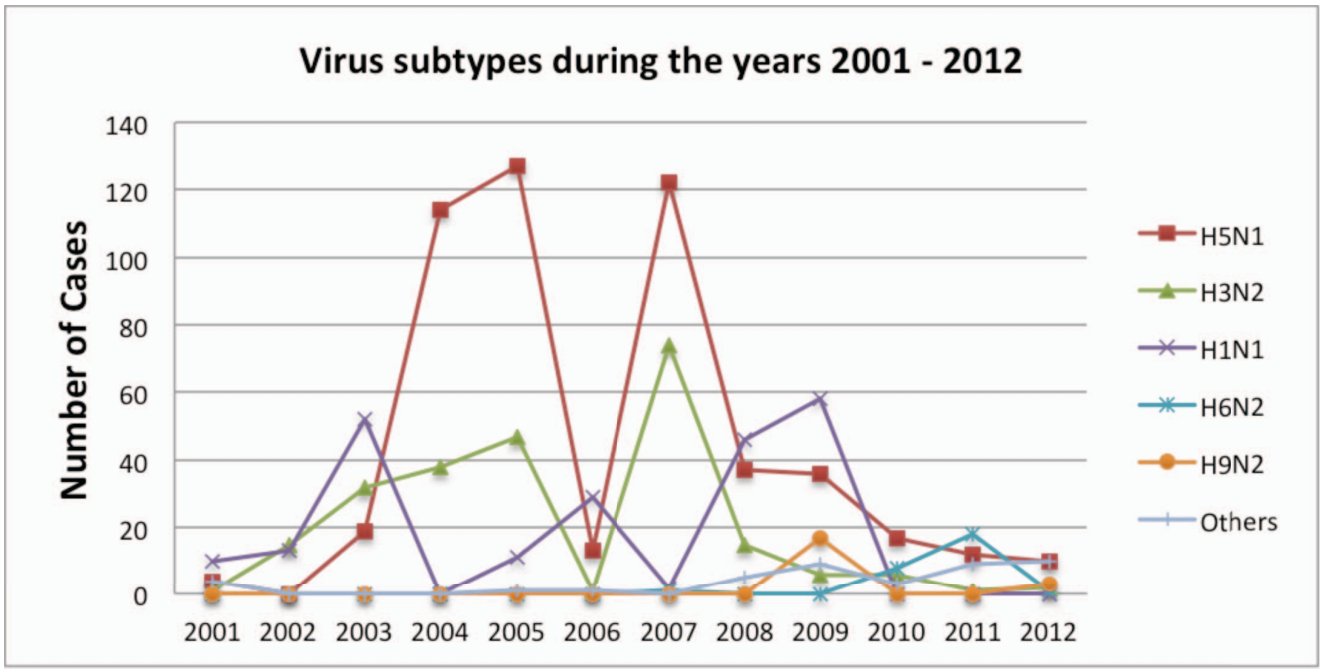

Figure 3: *The line chart shows infection trends of five key specific influenza A viruses during 2001-2012.*

approximately took 5%, chicken was 17% while 26% in duck pie, human cases was 50% and remained others.

There are five main subtypes including H5N1, H3N2, H1N1, H6N2, and H9N2 of influenza A viruses circulating in Vietnam since 2001 as shown in Figure 3. The number of cases of H3N2 is reasonable since this is the most popular subtype in human all over the world.

The facts that number of cases for H5N1 and H1N1 are very high in the period from 2004 to 2008 and in 2009 respectively have proven that the two recent pandemics, avian H5N1[4] and swine 2009[5] did affected the influenza A virus circulation in Vietnam. In addition, Vietnam is the first country that reported Tamiflu resistance to H5N1 subtype in 2005[6]. These two subtypes remain pathogenic and need to be closed monitoring to avoid local or global pandemics in the future.

IV. Conclusion

The analysis of data imported from Vietnam in NCBI for influenza A virus have revealed that there are five main influenza A virus including H5N1, H3N2, H1N1, H6N2, and H9N2 circulating in most areas of Vietnam from the North to the South. the diversity of genotypes present in Vietnam suggest a high level of genetic compatibility, identified as reassortants between a wide variety viral genotypes and share common mechanisms of spread. The most two popular host are human and poultry. The two recent pandemics avian H5N1 in 2005 and Swine H1N1 in 2009 did show significant effects on the viral circulation in Vietnam. These two subtypes are still pathogenic and require close disease control and management.

Acknowledgement

The authors acknowledge research grant No. T2012-13-SCSE by the International University (VNU-HCM).

References

1. Air GM, Laver WG (1989) The neuraminidase of influenza virus. Proteins 6: 341–356.
2. Sahini L, Tempczyk-Russell A, Agarwal R (2010) Large-Scale Sequence Analysis of Hemagglutinin of Influenza A Virus Identifies Conserved Regions Suitable for Targeting an Anti-Viral Response. PLoS ONE 5(2): e9268. doi:10.1371/journal.pone.0009268.
3. Bao Y., 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. J. Virol. 2008 Jan; 82(2): 596-601.
4. Abdel-Ghafar AN, Chotpitayasunondh T, Gao Z, Hayden FG, Nguyen DH, et al. (2008) Update on avian influenza A (H5N1) virus infection in humans.N Engl J Med 358: 261–273.
5. Butler, D., Swine flu goes global. Nature. 2009, (458), 1082-1083.
6. Le, Q. M.; Kiso, M.; Someya, K.; Sakai, Y. T.; Nguyen, T. H.; Nguyen, K. H. L.; Pham, N. D.; Nguyen, H. H.; Yamada, S.; Muramoto, Y.; Horimoto, T.; Takada, A.; Goto, H.; Suzuki, T.; Suzuki, Y.; Kawaoka, Y., Avian flu: Isolation of drug-resistant H5N1 virus. Nature (London, U. K.) 2005, 437 (7062), 1108.