International Journal of Endorsing
HEALTH SCIENCE RESEARCH

IJEHSR

Original Article

Epidemiological assessment of mutational capabilities in common seasonal viruses.

Syeda Tooba Burney[1] & Nisar Ahmed Shar[2]

[1]Department of Biomedical Engineering, NED University of Engineering and Technology, Karachi-Pakistan.

[2]National Center in Big Data & Cloud Computing, NED University of Engineering and Technology, Karachi-Pakistan.

Doi: 10.29052/IJEHSR.v9.i4.2021.448-458

Corresponding Author Email: nisarshar@neduet.edu.pk

Received 20/04/2021
Accepted 25/07/2021
First Published 04/10/2021

OPEN ACCESS

Abstract

Background: Influenza vaccine composition is reviewed prior to every flu season since influenza viruses frequently evolve through antigenic variations. Vaccine strains are selected in expectation of the upcoming influenza season to allow sufficient time for production. The aim of research is to use computational models for predicting the evolution of influenza based on the association of genetic mutations and antigenic traits of circulating viruses may apprise vaccine strain assortment decisions.

Methodology: This study also focuses on the correlation of viruses with spread rate using statistical method. For this method, we have worked on four different viruses Influenza, Ebola, Measles and Dengue. Year wise mutation rate was correlated with the epidemiological data to see the impact of mutations on the disease spread.

Results: We highlight the efficiency of this approach by analyzing the mutation rate and correlating it with its spread rate to find out either mutation in viruses causes disease spread or not. Our study identified mutations in viruses gets high before the outbreak of disease through which we can assess the upcoming outbreak. We can set a threshold value for nucleotide difference that can predict next outbreak of viral disease.

Conclusion: The concept of correlation between the genomic data and epidemic spread leads to the research analysis that mutations does not follow any pattern. Though most of the mutations are random. Our research concluded that some mutations may suppress the virus outbreak, and some mutate to become more resistant than the existing strain that causes outbreak.

Keywords

Computational Models, Genetic Mutations, Vaccine Strains, Viral Outbreak.

Check for updates

Introduction

Human history is full of viral disease epidemics that have devastated societies and whole populations. Well-known instances include the 1918 Spanish flu pandemic, the 2009 influenza H1N1 pandemic. The Ebola virus (EBOV) epidemic in the year 2014-2015 in West Africa, Dengue outbreak in Thailand with 102,510 confirmed cases in 2019[1] and measles outbreak in California year 2014-2015[2]. Not all outbreaks spread pandemic status, some of them are controlled by environmental factors or specific control measures. Outbreak research addresses two aims: To terminate the present outbreak and prevent future ones[3]. For eras, epidemiological methods such as detailed interaction tracing and mathematical modeling have been used to keep up these aims [4,5].

Applying statistical method to traditional epidemiology has significantly upgraded outbreak monitoring and prevention for all categories of viral diseases. These tools include correlating genomic and epidemic data and phenotypic methods to determine the specific strain or form of virus circulating in a population. They can be used to enhance diagnostics, to advice treatment programs and vaccine development, and to find out the spread of pathogens [6,7].

Analysis using genomic data expands our capacity to recognize viral outbreaks even further, subsequently tools of bioinformatics can predict next viral strain in specific viruses. Genomic analysis can lead to resolve disease spread by predicting the next outbreak years using computational methods, and data processing. Correlation of genomic data with spread rate can help in the prediction and control of multiple viral diseases before their outbreak. In this study we have considered four different viruses i.e., Influenza, Ebola, Measles and Dengue for the assessment of impact of mutations in viruses on disease spread.
Influenza, usually known as the flu, is an infectious disease triggered by an influenza virus. Symptoms can be minor to severe. Influenza is the pattern of a viral disease in which sustained evolution of the virus is of supreme importance for annual epidemics and infrequent pandemics of the disease in humans[8]. Out of all the four categories of influenza viruses (A, B, C and D), only influenza A viruses are commonly developed in humans[9]. The surface antigens of influenza viruses go through two types of alterations requiring the modification of vaccine strains almost every year. Antigenic drift are minor variations in the genes of influenza viruses that occur frequently over time as the virus replicates[10]. The key variation arises as a consequence of increase of point mutations in the superficial antigens determined by the immune response; this is represented as antigenic drift[11]. Antigenic shift is a process by which two or more different strains combines to form a new and different subtype[12]. Antigenic shift is activated either by direct transmission of non-human influenza viruses to humans or the re-assortment of genes from diverse influenza viruses that have infected a single cell.

A threatening epidemic of Zaire Ebola virus has been attacking West Africa since around December 2013, with the primary cases likely taking place in southern Guinea[13]. The responsible Ebola strain is intently associated to a strain related with past EBOV outbreaks in Central Africa and circulating in West Africa for about a decade[14]. Unusually, the existing size of the current EBOV epidemic far exceeds the overall number of cases reported for all former Ebola outbreaks combined. A total of 6,553 cases, with 3,083 deaths, have been reported to the World Health Organization (WHO) as of 23 September 2014. The contributing agent of Ebola Virus is an RNA virus of the family Filoviridae and genus Ebola Virus. There are total five Ebola strains have been recognized, namely Zaire ebolavirus, Tai Forest ebolavirus, Bundibugyo ebolavirus, Sudan ebolavirus and Reston ebolavirus with fruit bats which are the major probable reservoir host[15].

Measles morbillivirus, is a single-stranded, enveloped negative-sense, non-segmented RNA virus of the genus Morbillivirus within the family Paramyxoviridae. WHO reports that worldwide, in an approximate uptrend of cases going on in 2017, measles led to around 110,000 deaths, recorded in children aged under 5 years[16]. In the United States, outbreaks have been well known in Southern

California throughout December 2014 to February 2015 involved at least 125 cases[2]. Around 75% of the United States' cases originate from huge outbreaks in New York state and New York City[17].

Dengue is an infectious disease caused by the transfer of virus from the genre of mosquitoes called the Aedes agypti and Aedes albopictus. These mosquitoes are specially known for living in stagnant water. Dengue is a global burden and a major challenge throughout the world. Every year dengue infects an estimated 50-100 million individuals in tropical and subtropical countries. Dengue hemorrhagic fever may lead to fatal hemorrhagic events. Dengue has become the quickest growing mosquito-borne disease with nearly half the world's population currently in danger. Analysis at the local level in a specific high disease outbreak time frame is very much needed for better control of the dengue spreads locally.

Methodology

The flow of this research can be analyzed by the following schematic diagram.

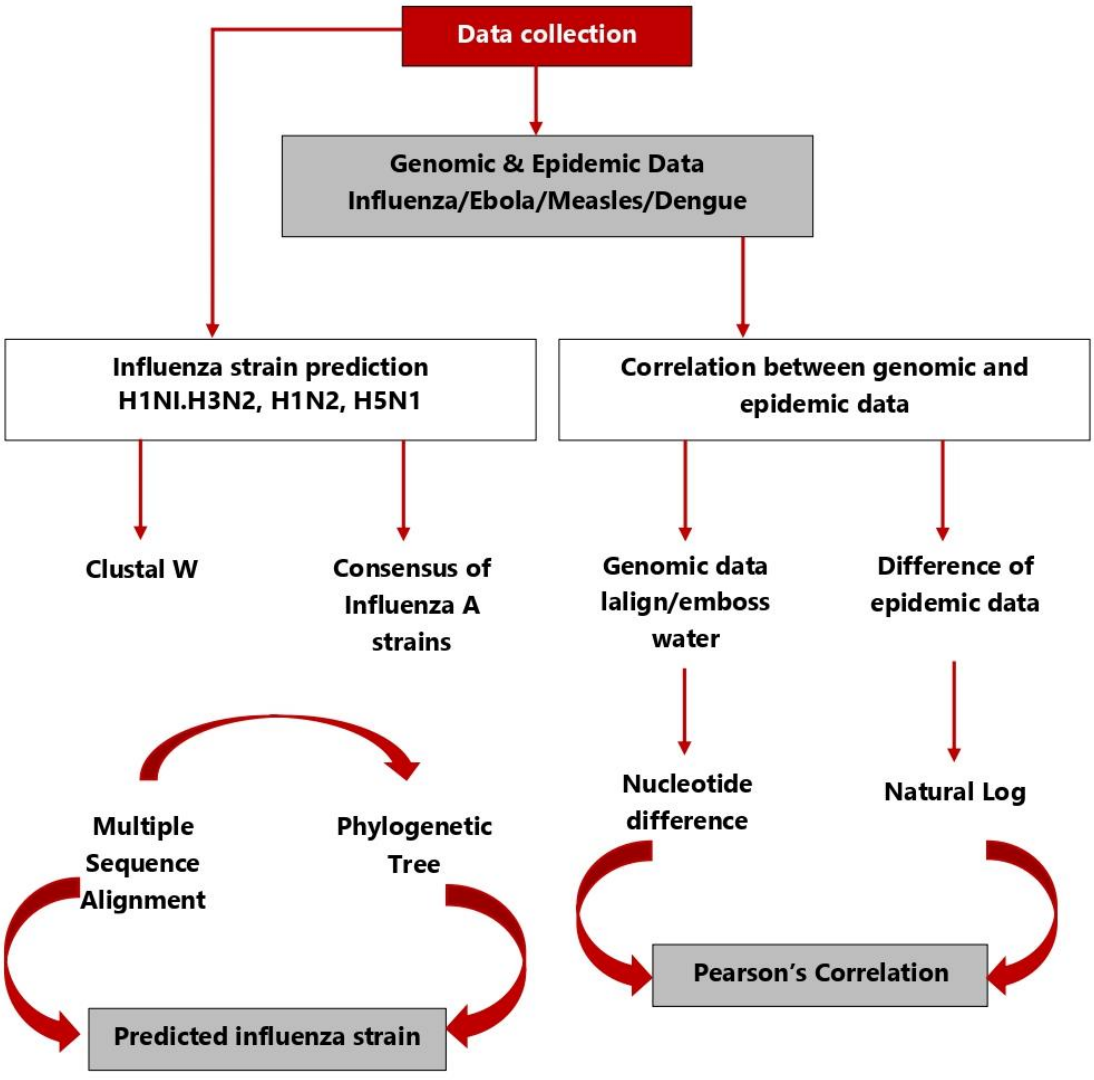

Figure 1: Schematic diagram of methodology followed in this research.

Influenza genomic data is utilized for two different methods one is related with the prediction of upcoming influenza strain and other one is to assess the impact of mutation in viruses including influenza, Ebola, Measles and Dengue on the spread of disease.

Influenza Research database (IRD) has search tools that allows user to search several different databases[18]. It is a huge database through which genomic data of hemagglutinin (HA1-HA18) and Neuraminidase (N1-N11) have been retrieved. Furthermore, different irregular genomic data points of H1N1 1918-1999 and continuous genomic data points 2000-2019 of USA have been collected. Other types of influenza A virus genomic data i.e. H3N2 2004-2019 USA, H5N1 2003-2015 China and H1N2 2009-2012 and 2017-2018 of USA have also been collected. The genomic data of Ebola, Measles and Dengue has been taken from virus pathogen resource[19].

Epidemic Data of Human influenza H1N1 have been collected on yearly basis from center of disease control[20]. Influenza cases from 2003 to 2019 cumulative and unadjusted incidence rate per 100,000 population data has been retrieved from CDC. Few epidemic cases and death reports have been collected by the help of California department of public health[21]. Data from 2003 to 2019 for hospitalizations per week and death cases per week have been collected from California Department of Public Health.

The collection of epidemic data of Ebola is from humanitarian data exchange[22] and center of disease control[23]. Epidemic data of measles has been taken from center of disease control[24]. Epidemic data of dengue has been collected from WHO[25].

Table 1: Year wise genomic and epidemiological data considered for four different viruses from different regions.

Virus	Year	Region	Reference
Influenza	2005-2019	California	18, 20, 21
Ebola	1994-1995	Democratic Republic of Congo	19, 22, 23
	1995-1996		
	2001-2002		
	2002-2003		
	2007-2008		
	2017-2018		
Measles	2012-2019	California	19, 24
Dengue	2000-2017	Thailand	19, 25

Techniques of Influenza Strain Prediction

The method designed for influenza strain prediction requires specific designed softwares. Clustal Omega is a multiple sequence alignment tool[26] that uses seeded guide trees and HMM (Hidden Markov Model) profile-profile techniques to create alignments between more than two sequences.

In bioinformatics, the consensus sequence is the calculated order of most frequent residues, either nucleotide or amino acid, found at each position in a sequence alignment[27]. The HIV sequence Database[28] makes a consensus of a submitted alignment using common consensus conventions.

Nucleotide difference is calculated by using webserver emboss water[29] or LALIGN[30]. It is known as William Pearson's Lalign program. Lalign program is based on Huang and Miller algorithm. It is a part of sequence analysis program. EMBOSS Water uses the Smith Waterman algorithm to calculate the local alignment of two sequences. By taking the difference of genomic segments of two years we can find out the change in nucleotide within each year.

Calculation of Nucleotide Difference and Epidemic Spread

- Nucleotide differences can be calculated by taking difference of one genomic segment from another
- Length 1- Length 2 =XYZ
- Take difference of small nucleotide segment with overlapping
- Small Length- Overlapping= ABC
- Take difference of both calculated values to calculate the final nucleotide difference.
- XYZ-ABC
- To calculate variation in number of cases, difference of two years is taken, and natural log is applied to arrange data set for correlation with nucleotide difference.
- For example
- 2001-2002 (Epidemic data of two years)
- 4324 – 2345 = 1979 (difference of epidemic data)
- 2.755 (Natural log of difference)
- Pearson's correlation is used between genomic and epidemic data which is the most useful technique to find out correlation between the two variables.

Results

Influenza A strains evolves and antigenically shifts, or drifts more rapidly as compared to other viruses. The Phylogenetic tree in (Figure 2B) represents the consensus of selected years (2009-2010-2011-2012-2014 and 2018) which was constructed by understanding the branching pattern in (Figure 2A). In (Figure 2B), we selected few years (1918-1934-1943) and aligned them with sequence of 2019 along with consensus sequence and analyzed that it was almost similar (Antigenic drift) to the consensus of past few years (2009 to 2011) and (2014 to 2018).

We concluded that consensus of selected years can predict the antigenic shift/drift of upcoming year (2020). Year 2019 in (Figure 2B) shows antigenic drift which clearly represents that there is no major change in the influenza H1N1 strain since past five years. Same methodology can be used for the prediction of influenza 2020 strain.

Phylogenetic tree in (Figure 2C) represents the co-evolution in virus. H1N1 strain of 2017 is related with the consensus of H3N2, H1N2, H5N1 and H1N1 than 2018 and 2019. Antigenic drift and antigenic shift are important parameters on which vaccination of influenza strain is designed.

It is essential to know the strain is similar to previous year or changed before designing of vaccination. Following results represents the phylogenetic trees along with consensus of multiple years and different strains of influenza.

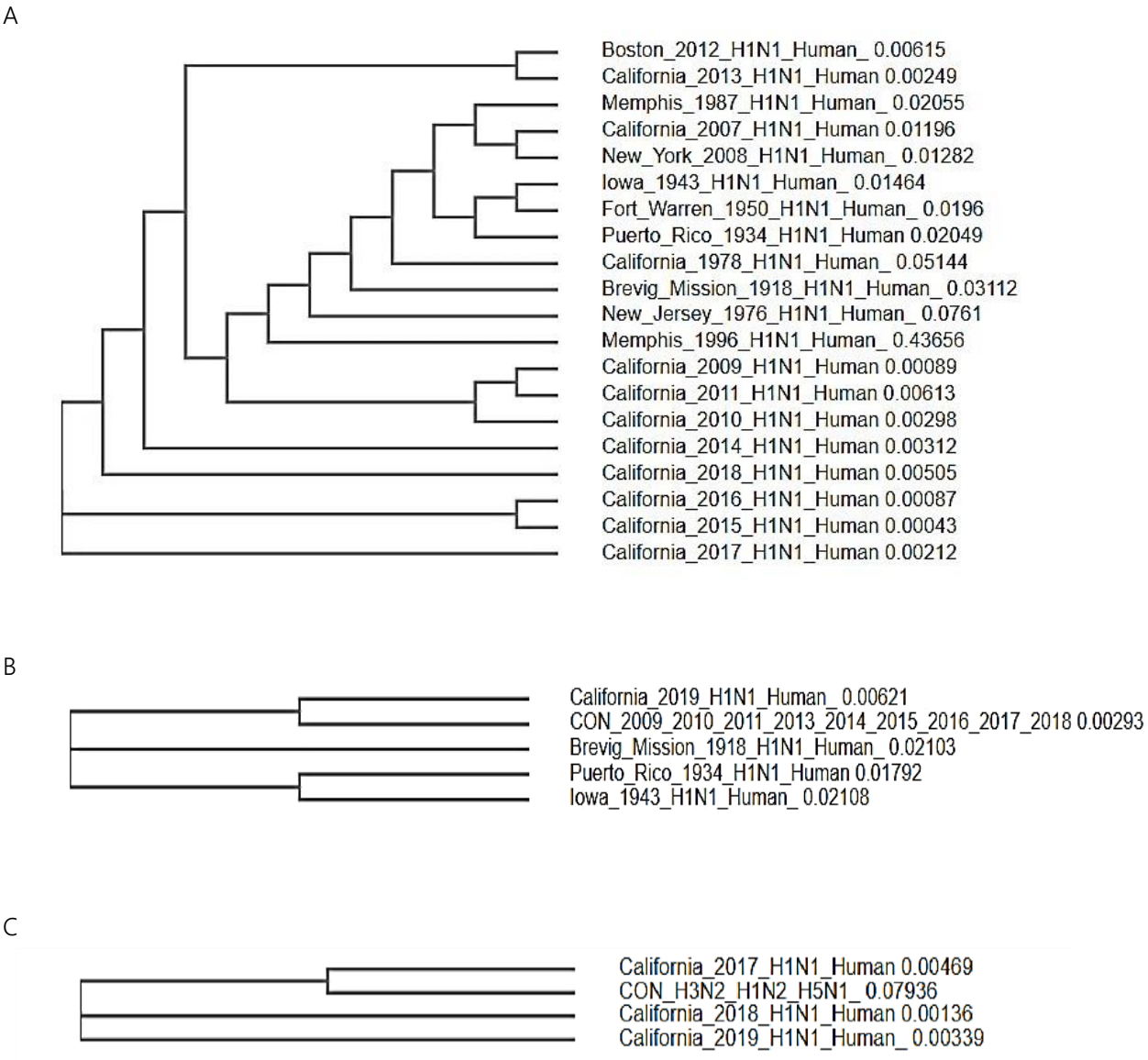

Figure 2: Phylogenetic Tree 2(A) represents comparison of H1N1 from 1918 to 2018 which helps in the selection of consensus years for influenza strain prediction. Phylogenetic tree 2(B) comparing different types of influenza A viruses. H1N1 of year 1918-1934-1943 along with consensus of selected years representing antigenic drift in the year 2019. Phylogenetic tree 2(C) represents coevolution in viruses which means that there must be some antigenic shift before 2017 and virus of other strain evolved into H1N1.

Impact of Mutation in Different Viruses

Selected outbreak years in four different viruses represents that the impact of mutation in virus is strong that took part in the major outbreak of disease. Increase in mutation count increases the spread rate. The mutations in viruses gets high before the outbreak of disease through which we can assess the upcoming outbreak. We can set a threshold value for nucleotide difference that can predict next outbreak of viral disease.

Table 2: Outbreak years of viruses along with their epidemic detail and mutation count.

Virus	Outbreak Years	Mutation	Epidemic Cases
Influenza	2009-2010	6	8.79
	2014-2015	50	8.98
	2016-2017	6	9.05
	2017-2018	6	6.38
	2018-2019	6	6.42
Ebola	1994-1995	180	5.19
	1995-1996	193	5.26
	2017-2018	300	5.7
Measles	2013-2014	887	6.17
	2014-2015	887	6.17
	2018-2019	40	6.73
Dengue	2000-2001	116	11.7
	2008-2009	26	11.07
	2009-2010	25	11.42
	2013-2014	124	11.76
	2014-2015	141	11.76
	2015-2016	494	11.73

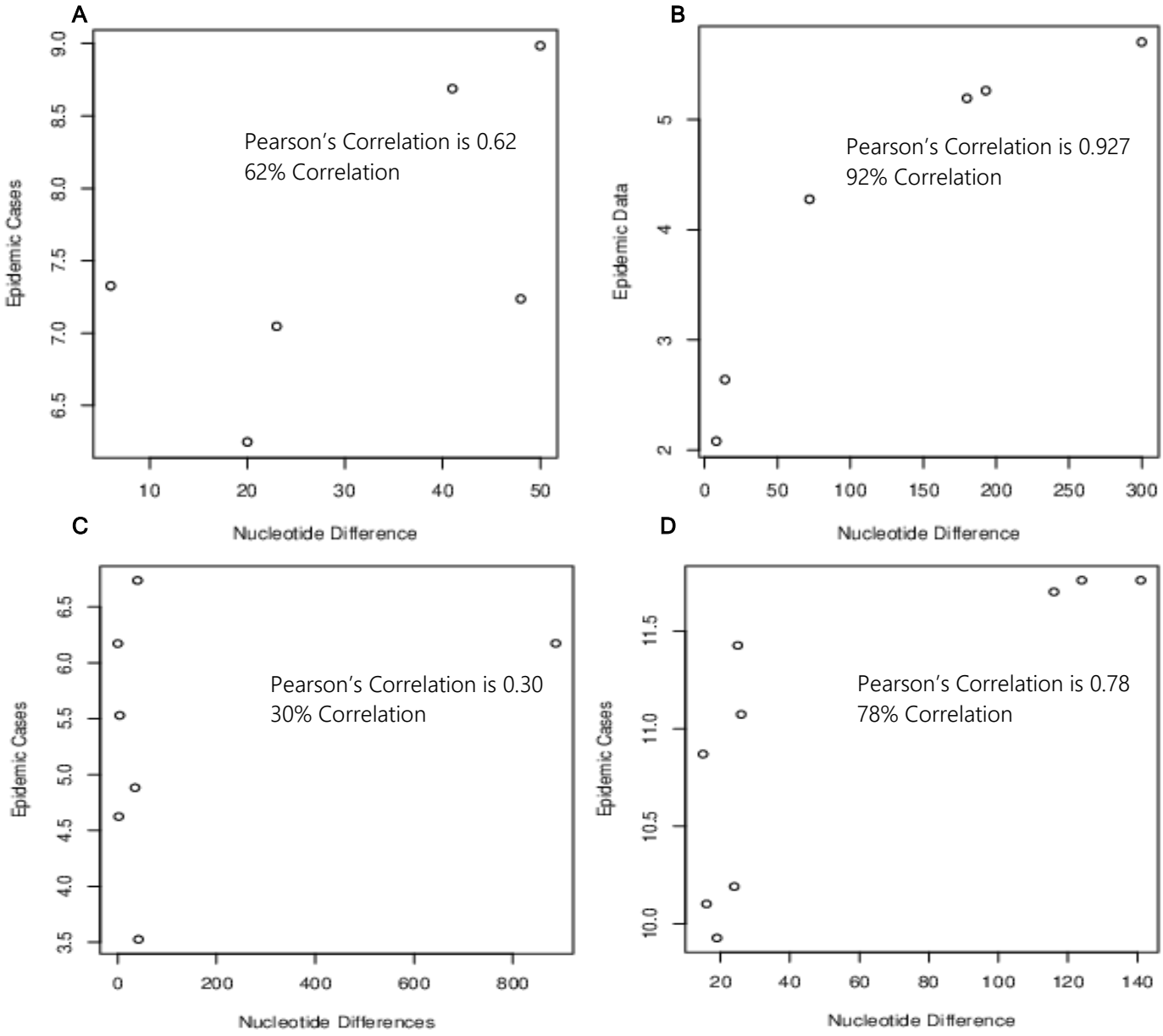

Figure 3: In plot (A) Nucleotide differences (X-axis) from 2005-2006 to 2007-2008 and 2013-2014 to 2015-2016, and the epidemic cases (Y axis) of influenza from 2005-2006 to 2007-2008 and 2013-2014 to 2015-2016 are correlated. In plot (B) Nucleotide differences (X axis) from 1994-1995 to 2017-2018 and epidemic cases (Y-axis) of Ebola from 2017-2018 then from 2013-2014. In plot (C) Nucleotide differences (X-axis) of 2013-2013 and then from 2015-2016 to 2018-2019 and epidemic cases (Y-axis) of measles of 2013-2013 and then from 2015-2016 to 2018-2019. In plot (D) Nucleotide differences (X-axis) of 2000-2001, 2003-2004, 2006-2007 to 2010-2011, 2013-2014, 2014-2015 and epidemic cases (Y-axis) of dengue 2000-2001, 2003-2004, 2006-2007 to 2010-2011, 2013-2014, 2014-2015.

Discussion

The virulence of disease can be distinguished by its genetic patterns with the help of genome sequencing along with the genetic factors that support to successful vaccine response[31]. Viruses that are closely associated to each other typically share the same antigenic properties and an immune system exposed to an analogous virus will usually identify it and respond[32]. But these minor genetic changes can gather over time and result in viruses that are antigenically different (away on the phylogenetic tree). When it occurs, the body's immune system may not identify those viruses. Mutational and evolutionary techniques associated to influenza viruses can help us develop prediction models[33]. Prediction models are extensively required for vaccine production[34]. The core research in epidemic and genomic parameters has the potential to efficiently develop vaccine before the outbreak[35].

The prediction of strain and concept of co-evolution can be verified by consensus of genomic segments[36]. In (Figure 2B) consensus of selected years was made with the help of (Figure 2A) which represents phylogenetic tree of influenza H1N1 from (1918 to 2019) with regular data points from (2000 to 2019). Phylogenetic tree in (Figure 2C) represents the co-evolution in virus. H1N1 strain of 2017 is related with the consensus of H3N2, H1N2, H5N1 and H1N1 than 2018 and 2019. Antigenic drift and antigenic shift are important parameters on which vaccination of influenza strain is designed[37]. It is essential to know the strain is similar to previous year or changed before designing of vaccination[38].

In (Figure 2B) we selected few years (1918-1934-1943) and aligned them with sequence of 2019 along with consensus sequence and analyzed that it was almost similar (antigenic drift) to the consensus of past few years (2009 to 2011) and (2014 to 2018) which clearly represents that there is no major change in the strain since past 4 years which can be verified by the epidemic cases as well. We concluded that consensus of selected years can predict the antigenic shift/drift of the year (2020).

The concept of correlation between the genomic data and epidemic spread leads to the research analysis that mutations does not follow any pattern. Though most of the mutations are random. Some mutations may suppress the virus outbreak, and some mutate to become more resistant than the existing strain that causes outbreak. Since the flu virus commonly drifts in its genetic alignment, you have to reformulate the vaccine, and this is one of the causes that individuals have to get a flu shot on yearly basis[39]. It has been observed that mutations in virus have strong impact in few years only in different viruses.

Conclusion

We have concluded that influenza epidemic and genomic data of the year 2005-2006, 2006-2007, 2007-2008, 2013-2014, 2015-2016 have 62% correlation which shows that increase in mutation count cause increase in spread rate. Strong correlation justifies the concept of antigenic drift and antigenic shift in case of influenza. Ebola genomic and epidemic data of the years 1994-1995, 1995-1996, 2001-2002, 2002-2003, 2007-2008 shows 92% correlation which represents that the impact of mutation in virus was strong that took part in the major outbreak of disease. Measles genomic and epidemic data of the years 2012-2013, 2015-2016, 2016-2017, 2017-2018 shows 30% correlation which represents that mutation in measles virus have moderate effect on disease outbreak. Dengue genomic and epidemic data of the year 2000-2001, 2003-2004, 2006-2007,20072008, 2008-2009, 2009-2010, 2010-2011, 2013-2014, 2014-2015 shows 78% correlation which represents that the impact of mutation in virus was strong that took part in the major outbreak of disease. Our developed method can be applied on Bacteria and other viruses. This research can help us to study how viruses develop resistance against drugs.

Conflicts of Interest

The authors have declared that no competing interests exist.

Acknowledgement

The authors acknowledge the funding from the Higher Education Commission (HEC) of Pakistan and the Ministry of Planning Development and Reforms under the National Center in Big Data and Cloud Computing at Exascale Open Data Analytics Lab (Genomics Lab), NED University of Engineering and Technology, Karachi, Pakistan..

Funding

The study has been funded by the Higher Education Commission (HEC) of Pakistan and the Ministry of Planning Development.

References

1. Thailand: Dengue cases top 100K, 250 additional Chikungunya cases reported - Outbreak News Today. Available at: http://outbreaknewstoday.com/thailanddengue-cases-top-100k-250-additional-chikungunya-cases-reported-21520/
2. Zipprich J, Winter K, Hacker J, Xia D, Watt J, Harriman K. Measles outbreak—California, december 2014–february 2015. MMWR. 2015;64(6):153.
3. Mitchell L, Ross JV. A data-driven model for influenza transmission incorporating media effects. R Soc Open Sci. 2016;3(10):160481.
4. Longini Jr IM, Koopman JS, Monto AS, Fox JP. Estimating household and community transmission parameters for influenza. Am J Epidemiol. 1982;115(5):736-751.
5. Monto AS. Studies of the community and family: acute respiratory illness and infection. Epidemiol Rev. 1994;16(2):351–373.
6. Webster RG. Influenza: an emerging disease. Emerg Infect Dis. 2019;4(3):436–441.
7. Shao W, Li X, Goraya MU, Wang S, Chen J-L. Evolution of Influenza A Virus by Mutation and Re-Assortment. Int J Mol Sci. 2017;18(8):1650.
8. Yan Y, Li C, Lu R. Meridional Displacement of the East Asian Upper-tropospheric Westerly Jet and Its Relationship with the East Asian Summer Rainfall in CMIP5 Simulations. Adv Atmos Sci. 2019;36(11):1203–1216.
9. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. Nature. 2009;459:931–939.
10. Bouvier NM, Palese P. The biology of influenza viruses. Vaccine. 2008;26(4):D49-53
11. Palese P. Making better influenza virus vaccines? Emerging Infectious Diseases. Centers for Disease Control and Prevention (CDC). 2006;12: 61–65.
12. Zambon MC. Epidemiology and pathogenesis of influenza. J. Antimicrob. Chemother. Oxford University Press. 1999:3–9.
13. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba NF, Soropogui B, Sow MS, Keïta S, De Clerck H, Tiffany A. Emergence of zaire ebola virus disease in guinea. N Engl J Med. 2014;371(15):1418–1425.
14. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, Wohl S. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014;345(6202):1369–1372.
15. Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Délicat A, Paweska JT, Gonzalez JP, Swanepoel R. Fruit bats as reservoirs of Ebola virus. Bat species eaten by people in central Africa show evidence of symptomless Ebola infection. Nature. 2005;438(7068):575–576.
16. World Health Organziation. Measles. [Updated 5th December 2019]. Available at: https://www.who.int/news-room/factsheets/detail/measles
17. CIDRAP - Center for Infectious Disease Research and Policy. Global measles outbreaks make 2019 a record-setting year. Available at: http://www.cidrap.umn.edu/news-perspective/2019/08/global-measlesoutbreaks-make-2019-record-setting-year
18. Influenza Research Database - IRD Search Tools. Available at: https://www.fludb.org/brc/search_landing.spg?decorator=influenza
19. Virus Pathogen Database and Analysis Resource (ViPR) - Genome database with visualization and analysis tools.
20. Weekly U.S. Influenza Surveillance Available at: Report://www.cdc.gov/flu/weekly/
21. California Department of Public Health (CDPH) Flu Reports. [Updated June 22, 2021]. Available at: https://www.cdph.ca.gov/Programs/CID/DCDC/pages/immunization/flu-reports.aspx
22. Ebola Cases and Deaths in the North Kivu Ebola Outbreak in the Democratic Republic of the Congo (DRC) - Humanitarian Data Exchange. Available at: https://data.humdata.org/dataset/ebola-cases-and-deaths-drc-north-kivu
23. Ebola Outbreak in Democratic Republic of the Congo, 2014: https://www.cdc.gov/vhf/ebola/outbreaks/drc/2014-august.html

24. Measles Cases and Outbreaks: https://www.cdc.gov/measles/cases-outbreaks.html
25. Bhutan B. WHO South-East Asia Region: Reported Cases and Deaths of Dengue from 2003 to 2012.
26. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD. Fast, scalable generation of high - quality protein multiple sequence alignments using Clustal Omega. Mol Sys Bio. 2011;7(539).
27. Isaroiu M, Dan Serbanati L. Comparison of result differences in multiple implementations of a stochastic motif finding algorithm. In: 2015 E-Health and Bioengineering Conference, EHB 2015. Institute of Electrical and Electronics Engineers Inc.; 2016.
28. Kuiken C, Korber B, Shafer RW. HIV sequence databases. AIDS Rev. 2003;5(1):52-61.
29. Madeira F, mi Park. Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;47(W1):W636-W641.
30. Huang and Miller, published in Adv. Appl. Math. 1991;12:337-357.
31. Gouvêa dos Santos W. NC-ND license Impact of virus genetic variability and host immunity for the success of COVID-19 vaccines. Published online 2021.
32. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, Okomo-Adhiambo M. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science. 2009;325(5937):197-201.
33. Castro LA, Bedford T, Meyers LA. Early prediction of antigenic transitions for influenza A/H3N2. PLOS Computational Biology. 2020;16(2):e1007683.
34. Voskarides K, Christaki E, Nikolopoulos GK. Influenza virus-host co-evolution. A predatorprey relationship? Front Immunol. Frontiers Media S.A. 2018;9.
35. The National Academies Keck Futures Initiative The Genomic Revolution -- Implications for Treatment and Control of Infectious Disease. The National Academies Keck Futures Initiative The Genomic Revolution -- Implications for Treatment and Control of Infectious Disease. 2006.
36. Pappas N, Roux S, Hölzer M, Lamkiewicz K, Mock F, Marz M, Dutilh BE. Virus Bioinformatics. Encyclopedia of Virology. Published online 2021:124.
37. Tenforde MW, Kondor RJ, Chung JR, Zimmerman RK, Nowalk MP, Jackson ML, Jackson LA, Monto AS, Martin ET, Belongia EA, McLean HQ. Effect of Antigenic Drift on Influenza Vaccine Effectiveness in the United States—2019–2020. Clin. Infect. Dis. 2020.
38. Rella SA, Kulikova YA, Dermitzakis ET, Kondrashov FA. Rates of SARS-CoV-2 transmission and vaccination impact the fate of vaccine-resistant strains. Sci. Rep. 2021 11:1. 2021;11(1):1-10.
39. Agor JK, Özaltın OY. Models for predicting the evolution of influenza to inform vaccine strain selection. Hum Vaccin Immunother. 2018;14(3):678.