Iranian Journal of Microbiology
Volume 12 Number 5 (October 2020) 483-494

ORIGINAL ARTICLE

In silico analysis and molecular characterization of Influenza A (H1N1) pdm09 virus circulating and causing major outbreaks in central India, 2009-2019

Arshi Siddiqui[1,3], Rashmi Chowdhary[1*], Harjeet Singh Maan[2], Sudhir Kumar Goel[1], Nidhi Tripathi[3], Anil Prakash[3]

[1]*Department of Biochemistry, All India Institute of Medical Sciences (AIIMS) Bhopal, Madhya Pradesh, India*
[2]*Department of Microbiology, State Virology Laboratory, Gandhi Medical College (GMC), Bhopal, Madhya Pradesh, India*
[3]*Department of Biotechnology, Barkatullah University, Bhopal, Madhya Pradesh, India*

Received: May 2020**, Accepted:** July 2020

ABSTRACT

Background and Objectives: Influenza A/H1N1pdm09 causes respiratory illness and remains a concern for public health. Since its first emergence in 2009, the virus has been continuously circulating in the form of its genetic variants. Influenza A/H1N1pdm09 surveillance is essential for uncovering emerging variants of epidemiologic and vaccine efficacy. The present study attempts *in silico* analysis and molecular characterization of Influenza A (H1N1) pdm09 virus circulating and causing major outbreaks in central India during 2009-2019.

Materials and Methods: We have investigated the antigenic drift analysis of 96 isolates' hemagglutinin (HA) gene sequences (59 central Indian and 37 local Indian and 28 global reference HA gene sequences) of Influenza A/H1N1pdm09 viruses from 2009 to 2019. The study includes mutational (Multiple sequence Alignment), phylogenetic (Maximum Likelihood Method), and statistical analysis (Covariance and correlation) of HA sequences submitted in NCBI, IRD and GISAID from central India.

Results: Phylogenetic analysis indicated maximum clustering of central Indian HA gene sequences in genogroup 6B. Analysis of amino acid sequence alignment revealed changes in receptor binding site (RBS). The frequency of S220T amino acid substitution was found to be high followed by S202T, K300E A273T, K180Q. The Karl Pearson correlation coefficient (r) and covariance between the number of mutations and the death toll was found +0.246 and +100.3 respectively.

Conclusion: The study identifies the continuous genetic variations in the HA gene sequences of circulating Influenza A/H1N1pdm09 in central India from the year 2009 to 2019. Further suggesting importance of monitoring the gradual evolution of the virus with regards to an increase in virulence, pathogenicity and vaccine efficacy timely.

Keywords: Influenza A virus, H1N1 subtype; Haemagglutinin; Central India

INTRODUCTION

Influenza A (H1N1) pdm09 is one of a subtype of Influenza A viruses of *Orthomyxoviridae* family, first appeared in 2009 in North America and caused a global pandemic in 2015 (1, 2). After emergence of 2009 pandemic, all Influenza A (H1N1) viruses called as Influenza A (H1N1) pdm09 (3). Globally, the annual epidemics have accounted for about 3 to 5 million cases of severe illness and 250,000 to 500,000 deaths worldwide (3). Influenza A (H1N1) pdm09 is now considered as a seasonal influenza virus that co-circulates with another seasonal influenza (H3N2) and influenza B viruses humans (3, 4). Un-

*Corresponding author: Rashmi Chowdhary, Ph.D, Department of Biochemistry, All India Institute of Medical Sciences (AIIMS) Bhopal, Madhya Pradesh, India.
Tel: +917552836116
Email: rashmi.biochemistry@aiimsbhopal.edu.in

like, influenza B and C which infects only humans, influenza A infects avian and mammalian hosts (5). Some aquatic avian species are the natural reservoir of all influenza A subtypes (6.) On the other hand, swine/pigs are susceptible to both avian and mammalian influenza viruses and act as an intermediate host and play an important role in the epidemiology and evolution of the virus (6).

Influenza A (H1N1) pdm09 genome consists of 8 negative-sense RNA segments viz. polymerases PB2 (Basic polymerase-2), PB1 (Basic Polymerase-1), PA (Acid Polymerase), NS1 (Nonstructural protein-1), NS2 (Nonstructural Protein-2) and surface proteins HA (Haemagglutinin), M (Matrix) and NA (Neuraminidase (7). Influenza viruses have their ability to undergo rapid and consistent genetic and antigenic evolution due to point mutations in the genome, especially HA and NA genes (antigenic drift) and reassortment of gene segments from intra-species and inter-species influenza viruses (antigenic shift) (8). Based on the variation of surface glycoproteins, haemagglutinin (HA) and neuraminidase (NA), the viruses are categorized into 18 HA (H1–H18) and 11 NA (N1–N11) subtypes (9). Influenza A (H1N1) pdm09, caused pandemic has evolved as a result of triple reassortment between swine influenza viruses of two distinct lineages and Eurasian avian-like swine (10, 11).

In India, the first positive case of Influenza A (H1N1) pdm09 was reported in 2009 from Hyderabad and continued to spread across the country (12, 13). Thereafter, the virus reappeared in the country during 2012-2013 with increased morbidity and mortality (14). The latest re-emergence in 2014-2015 caused the outbreak in the country with the highest human mortality (14). In central India, Madhya Pradesh has reported 5.74% positive cases in 2009 and 12.27% in the 2015 outbreak (15). So far, from 2010 to 2019 total of 4,790 positive cases and 905 deaths from central India have been recorded (15). Further continuous circulation of Influenza A (H1N1) pdm09 virus in the human population in the last decade (2009-2019) has resulted in its 8 genetic genogroups (4, 16). These data indicate the increasing virulence and pathogenicity of the virus with the time, mainly attributed to the continuous antigenic shift and drift in HA and/or NA genes of the virus (17).

Vaccination is the principal strategy for the prevention of the disease caused by the influenza virus (18). The surface protein HA helps the virus to attach and evade the host cell, thereby have become the prime targets of the host's neutralizing antibodies (19). Hence, the Influenza vaccine comprises HA antigens from A (H3N2), A (H1N1) pdm09, and one lineage of B (trivalent vaccine) or both lineages (quadrivalent vaccine) (20). However, accumulation of amino acid mutations coupled with N-linked glycosylation at epitope sites diminishes the antibody recognition leading to reduced vaccine efficacy and intermittent seasonal epidemics (21). In 2010, the World Health Organization (WHO) recommended using A/California/7/2009 as the vaccine component for A (H1N1) pdm09 virus (22). Due to the emergence of antigenic drift variants, it was later replaced in 2017 with A/Michigan/45/2015-like virus (23).

Hence, the routine influenza surveillance and continuous monitoring of the genetic changes in the major antigenic sites of these viruses are needed and are important for improving the efficacy of the vaccine for the slowly altering variants of the influenza virus.

The present study aims to investigate the in-silico analysis for antigenic and phylogenetic aspects of influenza A (H1N1) pdm09 viruses circulating in central India during 2009-2019, when there was increased evolution of the virus, focusing on the hemagglutinin protein (HA).

MATERIALS AND METHODS

Sequence retrieval. A total of 96 influenza A (H1N1) pdm09 HA gene sequences from the period of 2009 to 2019 retrieved from the Gen Bank, IRD (Influenza Research Database), GISAID (Global Initiative on Sharing All Influenza Data databases) in FASTA format (Table 1). Out of 96, 59 were central Indian sequences, and rest was reference sequences (27 global and 10 local reference sequences). Both partial and complete HA sequences used in the study.

Phylogenetic analysis. All the 96 HA gene sequences used for phylogenetic analysis. Both partial and complete nucleotide sequences of all the isolates aligned using CLUSTAL-W program. After alignment, the deduced HA sequences trimmed from both sides in Bioedit software and consensus sequences of 688 base pair of HA sequences analyzed to infer the evolutionary history through phylogenetic analysis.

Table 1. Twenty seven global references isolate' HA gene sequences were used in the study whereas fifty nine central Indian isolates' HA gene sequences were used in the study. One north Indian, 3 West Indian, 2 East Indian, 3 North east Indian, 1 Northwest Indian isolates' HA gene sequences were used as a reference isolate.

S. No	Name of isolates	Accession number	Source	Origin	Year	Genogroup
1	A/Czech Republic/112/2011	JQ693492	NCBI	Czech Republic	2011	2
2	A/San Diego/INS202/2009	CY066575	NCBI	San Diego	2009	4
3	A/Wisconsin/66/2009	KC781310	NCBI	Wisconsin	2009	4
4	A/Montana/18/2009	KC781361	NCBI	Montana	2009	3
5	A/Italy/128/2009	CY046061	NCBI	Italy	2009	3
6	A/Finland/18/2010	JN601087	NCBI	Finland	2010	3
7	A/Shanghai/143T/2009	GQ411907	NCBI	Shanghai	2009	1
8	A/California/07/2009	FJ981613	NCBI	California	2009	1
9	A/Osaka/1/2009	GQ219578	NCBI	Osaka	2009	1
10	A/England/197/2009	CY065198	NCBI	England	2009	1
11	A/ASTRAKHAN/1/2011	EPI90787	GISAID	Astrakhan	2011	5
12	A/Wisconsin/15/2011	KC881725	NCBI	Wisconsin	2011	5
13	A/Paris/1878/2012	EPI134406	GISAID	Paris	2012	8
14	A/Norway/120/2013	EPI134121	GISAID	Norway	2013	8
15	A/Wisconsin/14/2012	KC891394	NCBI	Wisconsin	2012	7
16	A/Thailand/ICRC-BKK4/2012	KF732010	NCBI	Thailand	2012	6A
17	A/Arizona/15/2010	KC 881868	NCBI	Arizona	2010	6A
18	A/Oman/SQUH-51/2012	KM213277	NCBI	Oman	2012	6C
19	A/Ontario/02/2014	KP864396	NCBI	Ontario	2014	6B
20	A/Washington/19/2015	KT836815	NCBI	Washington	2015	6B
21	A/Montana/35/2018	MK245822	NCBI	Montana	2018	6B
22	A/Hawai/56/2018`	MK557062	NCBI	Hawaii	2018	6B
23	A/Penninsylvania/511/2018	MK626389	NCBI	Penninsylvania	2018	6B
24	A/California/78/2018	MK556945	NCBI	California	2018	6B
25	A/Wisconsin/15/2016	KX409067	NCBI	Wisconsin	2016	6B
26	A/Michigan/45/2015	KU933493	NCBI	Michigan	2015	6B
27	A/California/80/2015	KT836680	NCBI	California	2015	6B
28	A/Jabalpur/8504/2017	MH160789	IRD	Central India	2017	6B
29	A/Bhopal/1697/2012	KM885031	IRD	Central India	2012	6A
30	A/Betul/6515/2015	KT369729	IRD	Central India	2015	6B
31	A/Bhopal/1613/2011	KM885030	IRD	Central India	2011	6A
32	A/Bhopal/1664/2011	KT241017	IRD	Central India	2011	7
34	A/Bhopal/3500/2015	KT426698	IRD	Central India	2015	6B
35	A/Dewas/4497/2015	KT241021	IRD	Central India	2015	6B
36	A/India/DRDE GWL672/2015	KX078501	IRD	Central India	2015	6B
37	A/India/DRDE GWL703/2015	KT867221	IRD	Central India	2015	6B
38	A/India/DRDE GWL719/2015	KT867219	IRD	Central India	2015	6B
38	A/India/DRDE GWL721/2015	KT867223	IRD	Central India	2015	6B
39	A/India/DRDE GWL812/2015	KT867224	IRD	Central India	2015	6B
40	A/India/DRDE GWL84/2015	KX078485	IRD	Central India	2015	6B
41	A/Harda/5023/2015	KT946863	IRD	Central India	2015	6B
42	A/Bhopal/3440/2015	KT946852	IRD	Central India	2015	6B
43	A/Bhopal/3580/2015	KT936477	IRD	Central India	2015	6B
44	A/Bhopal/3641/2015	KT946855	IRD	Central India	2015	6B
45	A/Bhopal/6804/2015	KT936493	IRD	Central India	2015	6B

Table 1. Continuing

46	A/Dewas/5457/2015	KT946865	IRD	Central India	2015	6B
47	A/Dewas/6184/2015	KT936490	IRD	Central India	2015	6B
48	A/Dhar/4849/2015	KT946862	IRD	Central India	2015	6B
49	A/Harda/4725/2015	KT946860	IRD	Central India	2015	6B
50	A/India/DRDE GWL897/2015	KT867220	IRD	Central India	2015	6B
51	A/India/DRDE GWL989/2015	KT867222	IRD	Central India	2015	6B
52	A/India/GWL-01/2011	JQ319658	IRD	Central India	2011	3
53	A/India/GWL-02/2011	JQ319657	IRD	Central India	2011	4
54	A/India/Gwl-06/2012	KC894815	IRD	Central India	2012	7
55	A/India/GWL-13/2013	KF683625	IRD	Central India	2013	4
56	A/India/INDO05/2019	MN061050	NCBI	Central India	2019	6B
57	A/India/K1730217/2017	MG271884	NCBI	East India	2017	6B
58	A/India/K1730225/2017	MG271885	NCBI	East India	2017	6B
59	A/India/P1729358/2017	MG271901	NCBI	Western India	2017	6B
60	A/India/Raj1725726/2017	MF319577	NCBI	Northwest India	2017	6B
61	A/Indore/10/2009	KT241013	IRD	Central India	2009	4
62	A/Indore/59/2009	KM885027	IRD	Central India	2009	7
63	A/Indore/379/2010	KT241014	IRD	Central India	2010	2
64	A/Indore/2683/2013	KF886296	IRD	Central India	2013	6C
65	A/Indore/2820/2013	KM885035	IRD	Central India	2013	6C
66	A/Indore/3415/2015	KT241019	IRD	Central India	2015	6B
67	A/Indore/3598/2015	KT241020	IRD	Central India	2015	6B
68	A/Indore/4181/2015	KT946858	IRD	Central India	2015	6B
69	A/Indore/4911/2015	KT936485	IRD	Central India	2015	6B
70	A/Indore/4961/2015	KT936487	IRD	Central India	2015	6B
71	A/Indore/6002/2015	KT369725	IRD	Central India	2015	6B
72	A/Indore/5991/2015	KT369724	IRD	Central India	2015	6B
73	A/Jabalpur/112/2009	KF886294	IRD	Central India	2009	4
74	A/Jabalpur/1737/2012	KM885033	IRD	Central India	2012	7
75	A/Jabalpur/1758/2012	KM885034	IRD	Central India	2012	7
76	A/Jabalpur/6722/2015	KT936491	IRD	Central India	2015	6B
77	A/Jabalpur/8504/2017	MH160790	IRD	Central India	2017	6B
78	A/Khandwa/3973/2015	KT946856	IRD	Central India	2015	6B
79	A/Khargone/4915/2015	KT936486	IRD	Central India	2015	6B
80	A/Khargone/5377/2015	KT946864	IRD	Central India	2015	6B
81	A/Madhya Pradesh/024/2010	KP317228	IRD	Central India	2010	3
82	A/Satna/6331/2015	KT369726	IRD	Central India	2015	6B
83	A/Shajapur/3712/2015	KT426699	IRD	Central India	2015	6B
84	A/Shajapur/4912/2015	KT369722	IRD	Central India	2015	6B
85	A/Ujjain/6165/2015	KT936488	IRD	Central India	2015	6B
86	A/Ujjain/2558/2013	KT241018	IRD	Central India	2013	6C
87	A/Ujjain/3548/2015	KT946853	IRD	Central India	2015	6B
88	A/Ujjain/4091/2015	KT946857	IRD	Central India	2015	6B
89	A/Ujjain/4154/2015	KT936480	IRD	Central India	2015	6B
90	A/Ujjain/5448/2015	KT369727	IRD	Central India	2015	6B
91	A/Delhi/086/2013	KP317290	NCBI	North India	2013	6C
92	A/Pune/NIV6196/2009	GU292352	NCBI	West India	2009	1
93	A/Lur/NIV24770/2010	CY075915	NCBI	West India	2010	3

Table 1. Continuing

94	A/Assam/2264/2009	KU310626	NCBI	Northeast India	2009	4
95	A/Assam/2590/2010	JN600357	NCBI	Northeast India	2010	3
96	A/Assam/2585/2010	KU310638	NCBI	Northeast India	2010	3

The phylogenetic tree generated by using the Maximum likelihood method and Tamura-Nei method of nucleotide substitution implemented in the MEGA X using bootstrap analysis of 1000 replicates (24, 25).

Mutational analysis. All the Central Indian isolates' HA sequences along with the Michigan and Californian vaccine strain aligned by Multiple Sequence Alignment using MUSCLE in MEGA X software to know the amino acid substitution in receptor binding sites ($RBS_{80\text{-}303}$) and intergenogroup antigenic divergence (24, 26, 27).

Statistical analysis. In order to find the correlation and co-variance between the number of mutations in receptor binding site and death toll during 2009 to 2019, mutations from 151 to 300 amino acid position of HA protein considered for calculation of co-variance and Karl Pearson coefficient of correlation (r) (28, 29). Because of unavailability of sequences in a particular year, we couldn't include it in calculation (Table 2).

RESULTS

Clustering of isolates in phylogenetic tree. The results obtained revealed clustering of central Indian HA gene sequences as follows: 1 isolate in genogroup 2, 2 in genogroup 3, 4 in genogroup 4, 3 in genogroup 6A, 43 in genogroup 6B, 2 in genogroup 6C and 4 in genogroup 7 (Table 1). 2017 and 2019 isolates were clustered in genogroup 6B (Fig. 1).

Formula for calculation of Karl Pearson Coefficient of Correlation

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2}\sqrt{(Y-\bar{Y})^2}}$$

Where, $\bar{X}$ = mean of X variable
$\bar{Y}$ = mean of Y variable

Formula for calculating Co-variance

Table 2. Table showing number of mutations in receptor binding sites and number of death annually.

S. No.	Year	Number of Mutations in RBS (X)	Number of Death (Y)
1.	2010	4	110
2.	2011	5	04
3.	2012	8	26
4.	2013	8	32
5.	2014	Sequence not available	09
6.	2015	9	367
7.	2016	Sequence not available	12
8.	2017	13	146
9.	2018	Sequence not available	34
10	2019	4	162

$$\mathrm{Cov}(X,Y) = \frac{\sum(X_i - \bar{X})(Y_j - \bar{Y})}{n-1}$$

- Xi – the values of the X-variable
- Yj – the values of the Y-variable
- X – the mean (average) of the X-variable
- $\bar{Y}$ – the mean (average) of the Y-variable
- n – the number of data points

Mutational analysis. Mutational analysis of 2019 virus compared to A/Michigan/45/2015 revealed N179S, Q180K, T233I, R240Q amino acid substitutions whereas 2017 viruses showed A90V, S91R, N179S, Q180K, T233I, R240Q mutations in receptor binding sites (Table 3). The intergenogroup antigenic divergence investigated with regards to genetic changes in HA gene of H1N1 viruses. Analysis of amino acid sequence alignment revealed changes at two positions (T151A, D239G) at RBS of HA between genogroup 2 and 3, three positions (A151T, S200P, S202T) between genogroup 3 and 4, two positions (N114D, E279G) between 4 and 6A, four positions (N101S, Q180K, G279E, E300K) between 6A and 6B, three positions (S101N, K180Q, I251V) between 6B and 6C genogroup, four positions (G101S, T214A, V251I, K300E) between 6C and 7 genogroup (Table 4).

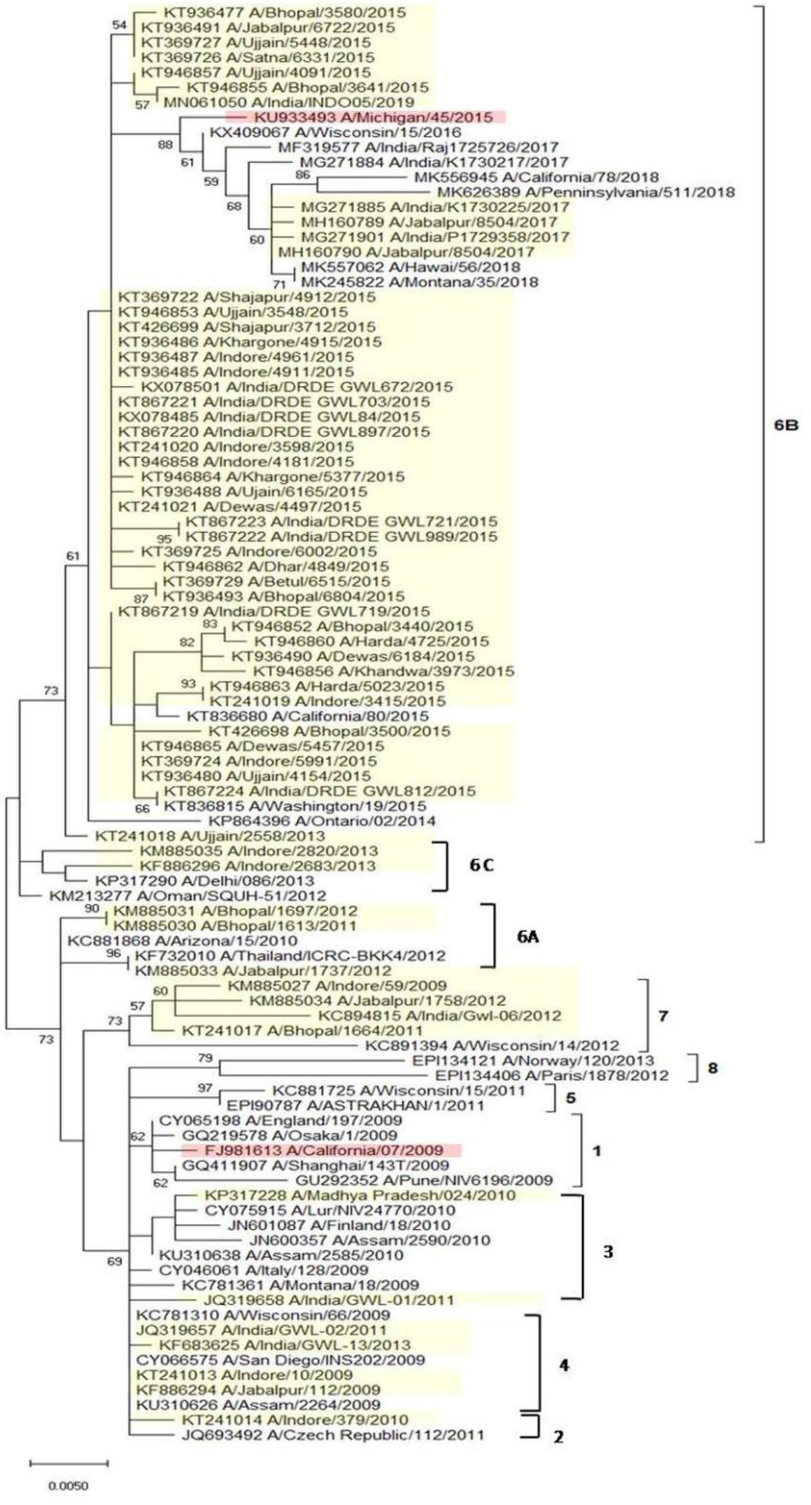

Fig. 1. Phylogenetic tree of H1N1 influenza A virus from Indian and global strains reported from 2009 to 2019. Phylogenetic tree of HA gene constructed from 28 global strains and 68 Indian strains, of which 59 strains includes form central India representing different genogroups. The tree was generated with the MEGA X programme using the maximum Likelihood method based on Tamura-Nei model of nucleotide substitution. Central India strains are indicated in light yellow shaded box, while A/California/07/2009 vaccine strains and A/Michigan/45/2015 strain in light pink colour.

Table 3. Amino acid changes on receptor binding sites of HA1 of 2017 and 2019 isolates compared to A/Michigan/45/2015 and A/California/07/2009 vaccine strain.

S. No.	Accession Number	Strain name	90	91	179	180	181	233	240
1.	KU933493	A/Michigan/45/2015	A	S	N	Q	S	T	R
2.	MN061050	A/India/INDO/2019	-	-	S	K	S	I	Q
3.	MH160789	A/Jabalpur/8504/2017	V	R	S	K	T	I	Q
4.	MH160790	A/Jabalpur/8504/2017	V	R	S	K	T	I	Q
5.	FJ981613	A/California/07/2009	A	S	S	K	S	I	Q

Table 4. Amino acid changes on antigenic sites of HA1 among genogroups of Influenza A H1N1 human influenza viruses.

Position	Geno group 2	Geno group 3	Geno group 4	Geno group 6A	Geno group 6B	Geno group 6C	Geno group 7
90	A	A	A	A	A	A	A
91	S	S	S	S	S	S	S
100	S	S	S	S	S	S	S
101	S	S	S	S	N	S	G
114	D	D	D	N	D	D	D
131	F	F	F	F	F	F	F
142	N	N	N	N	N	N	N
146	N	N	N	N	N	N	N
151	A	T	A	A	A	A	A
155	H	H	H	H	H	H	H
160	S	S	S	S	S	S	S
166	I	I	I	I	I	I	I
172	G	G	G	G	G	G	G
175	Y	Y	Y	Y	Y	Y	Y
177	K	K	K	K	K	K	K
179	S	S	S	S	S	S	S
180	K	K	K	K	Q	K	K
181	S	S	S	S	S	S	S
200	S	P	S	S	S	S	S
202	T	T	S	T	T	T	T
214	A	A	A	A	A	A	T
216	V	V	V	V	V	V	V
220	T	T	T	T	T	T	T
221	S	S	S	S	S	S	S
233	I	I	I	I	I	I	I
239	G	D	D	D	D	D	D
240	Q	Q	Q	Q	Q	Q	Q
245	N	N	N	N	N	N	N
246	Y	Y	Y	Y	Y	Y	Y
251	V	V	V	V	V	I	V
262	T	T	T	T	T	T	T
266	T	T	T	T	T	T	T
273	A	A	A	A	A	A	A
279	G	G	G	E	G	G	G
300	K	K	K	K	E	E	K

Frequency of variation of amino acid. Alignment of 59 HA protein sequences during the period of 2009 to 2019 was investigated to know the frequency of variation of amino acid at different key positions (151, 155, 166, 172, 175, 177, 179, 180, 181, 200, 202, 214, 216, 220, 221, 233, 239, 245, 246, 251, 262, 266, 273, 279, 300) of HA protein. The study included both partial and complete sequences and to keep up homogeneity, amino acid from position 151 to 300 used in study which revealed S220T highly frequent followed by S202T, K300E, A273T, K180Q and so on (Fig. 2).

Correlation and co-variance. Correlation between annual death toll and number of amino acid substitutions was calculated. The Karl Pearson Coefficient of correlation (r) came out +0.246. The co-variance between number of amino acid substitution and death toll came out +100.3.

DISCUSSION

Since its emergence in 2009, Influenza A/H1N1pdm09 is causing menace continuously (2). During 2009-2019, India saw a varied number of positive cases and deaths due to many factors (15). One of them is the mutations accumulated in the HA protein of Influenza A/H1N1pdm09 virus (21). HA protein is present as H0 which cleaved into HA1 and HA2 by the host cell (30). Influenza A/H1N1pdm09 HA1 subunit contains five antigenic sites Sb, Ca1, Ca2 and Cb (30). In addition to N-glycosylation, mutations in these sites change the antigenicity of influenza A/H1N1pdm09 HA1 and generate different variants that escape the neutralizing antibodies (21). *In silico* analysis (Phylogenetic analysis, mutational analysis) of HA amino acid sequences is a convenient method to study genetic variations which leads to evolution (10, 16).

For *in silico* analysis, we retrieved 96 HA gene sequences (59 central Indian sequences, 27 global reference HA gene sequences and 10 local Indian reference sequences) from IRD, NCBI, GISAID, during the period of 2009 to 2019. Out of 59 central Indian sequences, 26 were partial sequences and rest was complete sequences. We performed phylogenetic, mutational and statistical analysis using receptor binding sites (RBS80-303) (27).

There are 8 genogroups evolved globally (16). Phylogenetic analysis (using Maximum Likelihood method Tamura-Nei method in MEGA X software) results revealed five genogroups (2, 3, 4, 6 (6A, 6B, 6C) and 7 genogroups) evolved between 2009 to 2019 in central India (31, 32) with 1.69%, 3.38%, 6.77%, 5.08%, 72%, 3.38%, 6.77% HA gene sequences clustered in genogroup 2, 3, 4, 6A, 6B, 6C and 7 respectively. Recent study conducted in Middle East and North Afri-

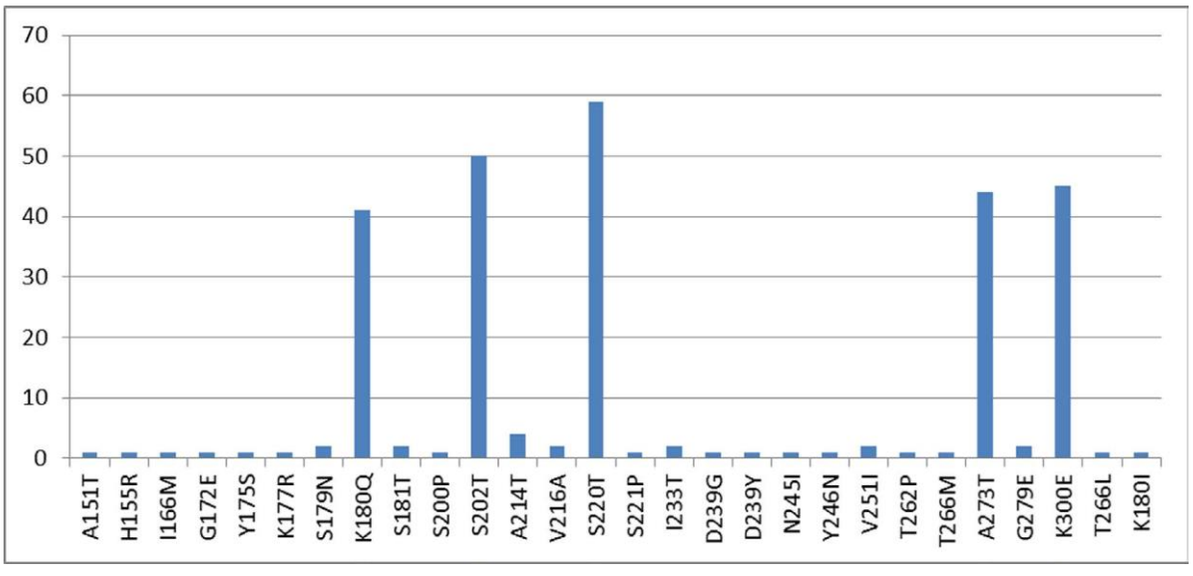

Fig. 2. Frequency distribution of variation of amino acid from 2009 to 2019 at different key positions of HA protein among 59 studied H1N1 viruses isolated in Central India (X-axis represents frequency and Y-axis represents type of amino acid substitutions)

ca (MENA) reported evolution of seven genogroups (33). In our study, after 2015, maximum HA gene sequences clustered in 6B genogroup which is similar to the reports published by WHO and other studies (33-37). Due to antigenic variations, 6B genogroup variants are reported to associate with increased morbidity compared to other genogroup variants (4, 35). For *in-silico* analysis, no clustering of HA gene sequences recorded for genogroup 1, 5 and 8 probably due to unavailability of data from central India. In 2009, HA sequences clustered in genogroup 4 (3.38%) and 7 (1.69%). In 2010, HA sequences clustered in genogroup 2 (1.69%) and 3 (1.69%). In 2011, HA sequences clustered in genogroup 4 (1.69%), 7 (1.69%), 6A (1.69%) and 3 (1.69%). 2012 HA protein sequences clustered in 7 (3.38%) and 6A (1.69%) genogroup. All of the sequences of 2015 (66.1%), 2017 (5.08%) and 2019 (1.69%) clustered in 6B genogroup. D114N, S202 T, S220 T and K300E amino acid substitutions in HA protein are characteristics of genogroup 6 (38).

Mutational analysis of HA amino acid sequences (Multiple sequence alignment by MUSCLE program in MEGA X software) of period 2009-2019 were investigated. Amino acid substitutions at positions 90, 91, 179, 180, 181, 233 and 240 of 2017 HA protein sequences from central India were observed when compared to A/Michigan/45/2015 and A/California/07/2009 has also been analyzed before (35, 38-41). Amino acid substitution at positions 179, 180, 233, 240 from 2017 and 2019 HA gene sequences of central India were found when compared to A/Michigan/45/2015, but when compared with A/California/07/2009, there were no variations (Table 3). This suggests that, these positions are mutated as compare to A/Michigan/45/2015 vaccine strain, but were found similar with A/California/07/2009 vaccine strain (39).

We performed intergenogroup antigenic divergence investigation (by using Multiple sequence alignment and comparing the amino acid substitutions in HA gene sequences between different genogroups found in central India at different positions described in results) during the period of 2009-2019. It revealed many changes which showed that virus is mutating continuously comparable to other studies (4, 35). As such, intergenogroup antigenic divergence investigation is not reported from studies on H1N1 influenza virus, but can observed in H5N1 influenza virus studies (26). Amino acid substitutions (S200, S202, D202, A214, I233) in receptor binding site envisaged to vary during adaptation process to α2-6-linked sialic acids receptors of human (42). Out of these, 4 sites were found mutated (S200, S202, A214, I233) in our study as similar to study conducted elsewhere (33). I223T amino acid substitution linked to increased binding affinity to human α2-6-linked sialic acids receptors (33). S200P and S202 T substitutions are responsible for enhancement of receptor-binding avidity whereas A214T substitution linked to decrease binding avidity (43).

Frequency distribution of amino acid substitution from 2009 to 2019 at different key positions of HA protein among 59 studied H1N1 viruses isolated in Central India resulted that S220T (100%) amino acid substitution was highly frequent similar to study conducted in MEENA (33) followed by S202T (84.7%), K300E (76.27%), A273T (74.57%), K180Q (69.4%), A214T (6.77%), S179N (3.38%), I233T (3.38%), V251I (3.38%), G279E (3.38%), S181T (3.38%), V216 (3.38%), A151T (1.69%), H155R (1.69%), I166M (1.69%), G172E (1.69%), Y175S (1.69%), K177R (1.69%), S200P (1.69%), S221P (1.69%), D239G (1.69%), D239Y (1.69%), N245I (1.69%), Y246N (1.69%), T262P (1.69%), T266M (1.69%) (Fig. 2). These substitutions have significant implications as they appeared in receptor binding site. K180Q substitution triggers conformational variation to ligand binding which might important for virulence (38). S179 N associated with glycosylation is responsible for enhanced pathogenicity of virus by prevention of antigenic sites for immune recognition (33). D239 amino acid substitution has deleterious effect on HA (33). S200P alter receptor binding affinity. S181 leads to changed glycan specificity (38). The amino acid substitutions such as P100S, S101N, D114N, K180Q, S181T, S202T, S220T, I233T, A273T, K300E present in the isolates from Central India in the HA gene are also reflected in the recent studies from other parts of India and world (38, 39, 44). Number of positive cases in the summer in central India, also shows that, the virus is getting more heat -resistant which might be due to the antigenic drift in H1N1 virus (45).

In our study, we reported a conceivable description for correlation between number of deaths and number of amino acid substitutions in HA gene of Influenza A/H1N1pdm09 virus in central India which may direct towards mortality. However in-depth lab study required for making these results operative. This type of study is yet to be enunciated in the literature. We

calculated correlation and co-variance (Karl Pearson coefficient of correlation and Co-variance) of number of amino acid substitutions and number of deaths in central India. Positive correlation and co-variance indicated that there is a correlation between number of deaths and number of mutations in HA protein. Very less researches are in line with the possible cause of mortality. However, Wu et al. recognized quantitative relationship between amino acid substitution of human influenza virus and mortality. They reported positive correlation between mortality and antigenic distance to the first antigenic strain (46). Adam et al. (2019) reported the qualitative relationship of amino acid substitution and mortality. According to him, S202T and D239G are responsible for increased mortality and morbidity (47). However, the factor affecting mortality also includes age, gender, severity of infection, mutations in other genes of influenza like NA and PB2 (47-49).

There are limitations in this study. Due to unavailability of HA gene sequences submitted to the GenBank from 2018, 2016 and 2014, the original evolution of Influenza A/H1N1pdm09 in central India is not completely explicit in this *in-silico* analysis. Moreover, unavailability of complete sequences of all HA gene is also a limitation for extensive molecular analysis. In addition to number of mutations in HA gene, mortality also depends on other factors described above. Therefore, consideration of other factors and in-depth wet lab study is required for making correlation perspicuous between mutations and mortality. Less data on HA gene sequences prevent investigating original genetic diversity of Influenza A/H1N1pdm09 in the region. This study has provided some direction towards changes that are occurring in the HA gene sequences of circulating Influenza A/H1N1pdm09 in the region and surely help scientific community to compare and further analyze their generated results to correlate findings about its outbreak.

CONCLUSION

In conclusion, these *in silico* findings direct a quickly changing Influenza A (H1N1) pdm09 virus during 2009 to 2019 in central India emphasizing the prerequisite for continuous surveillance together with molecular and antigenic analyses, to recommend suitable and proper influenza vaccine update. Additionally, detailed laboratory studies is needed for making correlation clearer between mutations and mortality.

ACKNOWLEDGEMENTS

We would like to thank Director AIIMS Bhopal, Department of Biochemistry AIIMS Bhopal for supporting to conduct this study. We thank Dr Ekta Makhija for helping in doing proof reading.

REFERENCES

1. World Health Organization. WHO | Influenza. Available at: https://www.who.int/biologicals/vaccines/influenza/en/ Accessed on 15th June 2020.
2. Rafeek RAM, Divarathna MVM, Noordeen F. History and current trends in influenza virus infections with special reference to Sri Lanka. *Virusdisease* 2017;28:225-232.
3. World Health Organization. WHO | Influenza (Seasonal) Available at: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal) Accessed on 12th June 2020.
4. Hsieh YC, Tsao KC, Huang CT, Chang KY, Huang YC, Gong YN. Clinical characteristics of patients with laboratory-confirmed influenza A(H1N1)pdm09 during the 2013/2014 and 2015/2016 clade 6B/6B.1/6B.2-predominant outbreaks. *Sci Rep* 2018;8:15636.
5. Mostafa A, Abdelwhab EM, Mettenleiter TC, Pleschka S. Zoonotic Potential of Influenza A Viruses: A comprehensive overview. *Viruses* 2018;10:497.
6. Gabriele Neumann, Takeshi Noda, Yoshihiro Kawaoka. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* 2009;459:931-939.
7. Dou D, Revol R, Östbye H, Wang H, Daniels R. Influenza A virus cell entry, replication, virion assembly and movement. *Front Immunol* 2018; 9:1581.
8. Kim H, Webster RG, Webby RJ. Influenza virus: Dealing with a drifting and shifting pathogen. *Viral Immunol* 2018;31:174-183.
9. Jang J, Bae SE. Comparative co-evolution analysis between the HA and NA genes of Influenza A virus. *Virology (Auckl)* 2018;9: 1178122X18788328.
10. Korsun N, Angelova S, Gregory V, Daniels R, Georgieva I, McCauley J. Antigenic and genetic characterization of influenza viruses circulating in Bulgaria during the 2015/2016 season. *Infect Genet Evol* 2017;49:241-250.

11. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, et al. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* 2009;325:197-201.
12. Kulkarni PS, Raut SK, Dhere RM. A post-marketing surveillance study of a human live-virus pandemic influenza A (H1N1) vaccine (Nasovac (®)) in India. *Hum Vaccin Immunother* 2013;9:122-124.
13. Choudhry A, Singh S, Khare S, Rai A, Rawat DS, Aggarwal RK, et al. Emergence of pandemic 2009 influenza A H1N1, India. *Indian J Med Res* 2012;135:534-537.
14. Mishra B. Resurgence of Influenza A (H1N1) 09: smoldering pandemic in India?. *J Glob Infect Dis* 2015;7:56-59.
15. Ministry of Health and Family Welfare, Government of India. Press Releases. Update on Influenza a (H1N1) as on 27th October 2019 Available on https://ncdc.gov.in/showfile.php?lid=280. Accessed on 1st November 2019.
16. Mukherjee A, Nayak MK, Dutta S, Panda S, Satpathi BR, Chawla-Sarkar M. Genetic Characterization of circulating 2015 A(H1N1)pdm09 Influenza viruses from Eastern India. *PLoS One* 2016; 11(12): e0168464.
17. Shao W, Li X, Goraya MU, Wang S, Chen JL. Evolution of Influenza A virus by mutation and re-assortment. *Int J Mol Sci* 2017;18:1650.
18. World Health Organization. WHO | Global Influenza Strategy 2019-2030 Available at: https://apps.who.int/iris/bitstream/handle/10665/311184/9789241515320-eng.pdf?ua=1
19. Chen X, Liu S, Goraya MU, Maarouf M, Huang S, Chen JL. Host immune response to influenza A virus infection. *Front Immunol* 2018;9:320.
20. World Health Organization. WHO | Recommended composition of influenza virus vaccines for use in the 2019- 2020 northern hemisphere influenza season. Available at: https://www.who.int/influenza/vaccines/virus/recommendations/201902_recommendation.pdf?ua=1
19. Chang Y, Xu C, Zhang H, Liu GD, Xue C, Cao Y. Targeting Hemagglutinin: approaches for broad protection against the influenza A virus. *Viruses* 2019;11:405.
22. World Health Organization. WHO | Recommended composition of influenza virus vaccines for use in the 2011-2012 northern hemisphere influenza season. Available at: https://www.who.int/influenza/vaccines/2011_02_recommendation.pdf
23. World Health Organization. WHO | Recommended composition of influenza virus vaccines for use in the 2018 southern hemisphere influenza season 2017. Available at: https://www.who.int/influenza/vaccines/virus/recommendations/201709_recommendation.pdf?ua=1
24. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018; 35:1547-1549.
25. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993; 10:512-526.
26. Bhat S, Nagarajan S, Kumar M, Murugkar HV, Kalaiyarasu S, Venkatesh G, et al. Antigenic characterization of H5N1 highly pathogenic avian influenza viruses isolated from poultry in India, 2006-2015. *Arch Virol* 2017;162:487-494.
27. Sarmah K, Borkakoty B, Sarma K, Hazarika R, Das PK, Jakharia A, et al. Genetic variations of the hemagglutinin gene of pandemic Influenza A (H1N1) viruses in Assam, India during 2016. *3 Biotech* 2018;8:408.
28. EDUCBA. Available at: https://www.educba.com/covariance-formula/
29. Karl Pearson coefficient of correlation. Available at: https://businessjargons.com/karl-pearsons-coefficient-of-correlation.html
30. Zolotarova O, Budzanivska I, Leibenko L, Radchenko L, Mironenko A. Antigenic site variation in the hemagglutinin of pandemic Influenza A(H1N1)pdm09 viruses between 2009-2017 in Ukraine. *Pathogens* 2019;8:194.
31. Sahu M, Singh N, Shukla MK, Potdar VA, Sharma RK, Sahare LK, et al. Molecular and epidemiological analysis of pandemic and post-pandemic influenza A(H1N1)pdm09 virus from central India. *J Med Virol* 2018;90:447-455.
32. Pandey S, Sahu M, Potdar V, Barde P. Molecular analysis of influenza A H1N1pdm09 virus circulating in Madhya Pradesh, India in the year 2017. *Virusdisease* 2018;29:380-384.
33. Al Khatib HA, Al Thani AA, Gallouzi I, Yassine HM. Epidemiological and genetic characterization of pH1N1 and H3N2 influenza viruses circulated in MENA region during 2009-2017. *BMC Infect Dis* 2019;19:314.
34. Sullivan SJ, Jacobson RM, Dowdle WR, Poland GA. 2009 H1N1 influenza. *Mayo Clin Proc* 2010;85:64-76.
35. Opanda S, Bulimo W, Gachara G, Ekuttan C, Amukoye E. Assessing antigenic drift and phylogeny of influenza A (H1N1) pdm09 virus in Kenya using HA1 sub-unit of the hemagglutinin gene. *PLoS One* 2020;15(2):e0228029.
36. Monamele CG, Munshili Njifon HL, Vernet MA, Njankouo MR, Kenmoe S, Yahaya AA, et al. Molecular characterization of influenza A(H1N1)pdm09 in Cameroon during the 2014-2016 influenza seasons. *PLoS One* 2019;14(1): e0210119.
37. Bakri M, Samuh M, Azzeh M. Molecular epidemiology survey and characterization of human influenza A viruses circulating among Palestinians in East

Jerusalem and the West Bank in 2015. *PLoS One* 2019;14(3):e0213290.
38. Jones S, Nelson-Sathi S, Wang Y, Prasad R, Rayen S, Nandel V, et al. Evolutionary, genetic, structural characterization and its functional implications for the influenza A (H1N1) infection outbreak in India from 2009 to 2017. *Sci Rep* 2019; 9:14690.
39. Prasad R, Mohanakumari VV, Sasi RV, Nair R, Jones S, Pillai MR. Complete genome analysis of Influenza A(H1N1) viruses isolated in Kerala, India. *Microbiol Resour Announc* 2020;9(12): e00062-20.
40. Pandey S, Sahu M, Potdar V, Barde P. Molecular analysis of influenza A H1N1pdm09 virus circulating in Madhya Pradesh, India in the year 2017. *Virusdisease* 2018;29:380-384.
41. Sahu M, Singh N, Shukla MK, Potdar VA, Sharma RK, Sahare LK, et al. Molecular and epidemiological analysis of pandemic and post-pandemic influenza A(H1N1)pdm09 virus from central India. *J Med Virol* 2018;90:447-455.
42. Maines TR, Jayaraman A, Belser JA, Wadford DA, Pappas C, Zeng H, et al. Transmission and pathogenesis of swine-origin 2009 a(H1N1) influenza viruses in ferrets and mice. *Science* 2009;325:484-487.
43. de Vries RP, de Vries E, Martínez-Romero C, McBride R, Kuppeveld FJ, Rottier PJM, et al. Evolution of the hemagglutinin protein of the new pandemic H1N1 influenza virus: maintaining optimal receptor binding by compensatory substitutions. *J Virol* 2013; 87: 13868-13877.
44. Biswas D, Dutta M, Sarmah K, Yadav K, Buragohain M, Sarma K, et al. Genetic characterisation of influenza A(H1N1)pdm09 viruses circulating in Assam, Northeast India during 2009–2015. *Indian J Med Microbiol* 2019;37:42-49.
45. Science. The wire. https://science.thewire.in/health/influenza-niv-h1n1-h3n2-reassortant/
46. Wu A, Peng Y, Du X, Shu Y, Jiang T. Correlation of influenza virus excess mortality with antigenic variation: application to rapid estimation of influenza mortality burden. *PLoS Comput Biol* 2010;6(8):e1000882.
47. Adam DC, Scotch M, MacIntyre CR. Phylodynamics of Influenza A/H1N1pdm09 in India reveals circulation patterns and increased selection for clade 6b residues and other high mortality mutants. *Viruses* 2019;11:791.
48. Parida M, Dash P, Kumar J, Joshi G, Tandel K, Sharma S, et al. Emergence of influenza A(H1N1) pdm09 genogroup 6B and drug resistant virus, India, January to May 2015. *Euro Surveill* 2016;21:6-11.
49. Gao F, Loring C, Laviolette M, Bolton D, Daly ER, Bean C. Detection of 2009 pandemic influenza A(H1N1) virus Infection in different age groups by using rapid influenza diagnostic tests. *Influenza Other Respir Viruses* 2012;6(3): e30-e34.