bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Next-Gen sequencing of novel pandemic swine flu [A(H1N1)pdm09] virus in India revealed novel mutations across the genome

Paban Kumar Dash[1#], Chitra Pattabiraman[2#], Kundan Tandel[1], Shashi Sharma[1], Jyoti S Kumar[1], Shilpa Siddappa[3] , Malali Gowda[3], Sudhir Krishna[2], and Man Mohan Parida[1*]

Affiliation: [1]Division of Virology, Defence Research & Development Establishment (DRDE), Gwalior-474002, India. [2]National Centre for Biological Sciences, TIFR, Bangalore -560065, India.

[3]Next-Generation Sequencing facility, Centre for Cellular and Applied Molecular Platforms, Bangalore, India.

*** Corresponding author**

Dr M.M.Parida,

Scientist -G, Head, Division of Virology,

Defence R & D Establishment,

Jhansi Road, Gwalior – 474002, India

Fax: + 91-751-2351148

Email-paridamm@rediffmail.com

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

The Influenza A H1N1 virus of 2009 was the first pandemic flu virus of the 21st century. Identifying the emergence of mutations in rapidly mutating Influenza viruses that allow increased transmission or confer resistance are invaluable to global outbreak response. Here we recovered 5 complete Influenza A genomes from 4 oropharygeal swabs and one cell culture isolate from a severe Indian outbreak of flu in early 2015. Multiple amino acids substitutions including those known to confer resistance to Oseltamivir and increased pathogenecity in mice were found in the Neuraminidase gene. Additional mutations both reported and novel were found throughout the genome compared to the vaccine strain (California/04/2009). All eight segments of the complete genomes were found to be genetically related to the 2009 pandemic strain, A(H1N1)pdm09 and belonging to the emerging genogroup 6B. This group was found to be of south East Asian origin by time scale phylogentic analysis. A phylogeographic analysis revealed 39 significant migration events among globally circulating viruses. This study is the first extensive complete genome and phylogeographic analysis of 2015 Indian A(H1N1) pdm09 viruses. We report several novel mutations in the 2015 Indian strains which need to be evaluated for effect on viral replication, transmission and resistance to therapy. The identification of mutant A(H1N1)pdm09 from India warrants continuous monitoring of viral evolution for implementation of suitable medical countermeasures.

Keywords

Pandemic H1N1, India, genotype 6B, deep sequencing, genome, phylogeography

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1. Introduction

The last pandemic of influenza occurred when a quadrouple reassorted Influenza A virus which originated in Mexico spread across the human population in 2009. This novel pandemic virus was found to originate from multiple reassortments of avian, swine, and human viruses and probably circulated in the north American swine population for over a decade (Potdar et al., 2010).

The genome of IAV comprises of 8 segments consisting of 8 negative-sense single-stranded RNA genomes. The 13.5 kb genome encodes 12 proteins: hemagglutinin (HA), M proteins (M1 and M2), neuraminidase (NA), nucleocapsid protein (NP), nonstructural proteins (NS1 and NS2), and polymerase subunits (PA, PA-X, PB1-F1, PB1-F2, and PB2). IAVs are divided into multiple subtypes based on two surface glycoproteins hemagglutinin (H1 to H18) and neuraminidase (N1 to N11) (Matsuzaki et al., 2014; Garten et al., 2009)

The evolution of Influenza viruses, driven by antigenic drift and antigenic shift allows them to circumvent host immunity and leads to occurrence of regular epidemic and occasional pandemic. Work on multiple influenza viruses has demonstrated that mutations in the receptor binding domain of the influenza hemaglutinin protein are one of the key antigenic determinants together with the viral neuraminidase (Park et al., 2009). Characterization of NA further allows identification of drug resistant mutants particularly to widely used zanamivir (Relenza) and oseltamivir (Tamiflu). Novel variations in HA and NA will lead to changes in the antigenic profile and the consequent inadequateness of immune responses to the virus could give rise to more virulent strains (Kosoltanapiwat et al., 2014). In contrast, relatively little is known about mutations in the other 6 segments of the influenza genome. While more recent evidence does suggest a role of mutations in the matrix proteins in pathogenecity, the information on novel mutations in the full-length virus, especially in geographical niches remains poor.

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

In 2015, India saw a large number of human flu cases requiring hospitalization. During the pandemic period (2009-2010), more than 50000 cases with a case fatality of 6% were reported in India. Subsequently, the virus activity subsided during 2011-2014 (Parida et al., 2016; Takashita et al., 2015; Dakhave et al., 2013, Broor et al., 2012). Surprisingly, resurgence of A(H1N1)pdm09 was recorded with more severity than the previous pandemic period in early part of 2015. This nation–wide epidemic affected more than 39000 persons with a case fatality of more than 2,500. This outbreak was characterized by viruses which belong to the emerging groups of H1N1 – genogroup 6B. We have recently published HA gene sequence of some of the viruses circulating in 2015, which highlighted the need to characterize the full- genome of these viruses. Here we recovered complete genomes of viruses from this outbreak both from oral swabs and from an isolate and identified potential amino acid changes throughout the viral genome. A time scale phylogeny was constructed to understand their evolution particularly focussing on the emerging genogroup 6B, since the onset of pandemic in 2009.

2. Material and Methods

2.1 Clinical samples

Acute phase respiratory swab (nasopharyngeal/ throat/ nasal) samples from patients suspected for A(H1N1)pdm09 infection were collected by health administrators of different district of Madhya Pradesh, India. These were referred to DRDE, Gwalior, a nodal reference lab for pandemic swine flu H1N1 molecular diagnosis. WHO guidelines were followed for sample collection, transport and laboratory investigation in a BSL-3 laboratory (WHO, 2009). The detailed clinical history and epidemiological information of the five samples sequenced in this study were provided in Table 1. All these samples were collected at the height of epidemic during Feb-Mar 2015. Ethical clearance for this study was approved by DRDE Biosafety Committee. Informed consent was obtained by district

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

health authorities as part of laboratory diagnosis and treatment.

2.2 Extraction of viral RNA

Viral RNA was extracted from the infected nasopharyngeal/throat/nasal samples using the QIAamp viral RNA mini kit (Qiagen, Germany), according to the manufacturer's protocol, Briefly 560 µl of buffer AVL was added to 140 µl of infected material and vortexed for 15 sec and the mixture was incubated at room temperature (15^0C-25^0C) for 10 min. After washing twice finally RNA was eluted in 70 µl of elution buffer in collection tube and stored at –80 0C until use.

2.3 Metagenomic sequencing from cell culture supernatants and oropharngeal samples – RNA from 6 clinical samples, 3 isolates (passaged in MDCK cells) and 1 healthy control was subjected to deep sequencing on the Ion Proton/PGM sequencing platform. Extracted RNA was concentrated and used for library preparation using the Ion Proton Library preparation kit (Ion Total RNA-Seq kit V2), samples were multiplexed in set of 5-6 using IonXpress barcodes (Ion Xpress RNA Seq BC) and template preparation was carried out using Ion PI HI-Q template OT2 200 kit. A water control was included in one chip. Sequencing was carried out with 8 pM of pooled libaries using the Ion PI HIQ 200 kit and pre-primed chips for sequencing (Ion PI chip kit V3). The number of sequences from each sample are provided in Table 2.

2.4 Analysis of sequencing reads

The reads were retrieved after demultiplexing in fastq format. The SNAP alignment tool was used to remove host sequences (human genome and mRNA for samples and canine genome for isolates) and ribosomal RNA sequences (From the SILVA database). The remaining reads were aligned using blast to the viral RefSeq database to provide an unbiased spectrum of viruses. Mapping assembly was performed using the MIRA tool

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(Chevreux et al., 2000) with the A/California/07/2009(H1N1) strain as the reference strain. The resulting consensus sequence was used in all other analysis. The five viral genome sequences reported here were deposited in GenBank under the accession numbers KX078482 to KX078489, KX078490 to KX078497, KX078498 to KX078505, KX078506 to KX078513, and KX078514 to KX078521 (Dash et al., 2018).

Representative sequences of A(H1N1)pdm09 from geographically diverse region circulating between 2009 and 2015 were retrieved from GISAID EpiFlu database (Table S1). Multiple sequence alignment of the viruses was carried out employing MUSCLE (Edgar, 2004) hosted at the EMBL-EBI server. The deduced amino acid was determined from the nucleotide sequence using the EditSeq module of Lasergene 5 software package (DNASTAR Inc, USA). The deduced amino acid sequences of five A(H1N1)pdm09 sequenced in this study along with prototype vaccine strain (A/California/07/2009) and six Indian representative A(H1N1)pdm09 viruses from 2009 to 2014 were compared. The percent nucleotide identity and percent amino acid identity values were calculated as pair-wise p-distances. Concatenated (n=96) and individual gene segment (n=120)-wise phylogenetic analysis was carried out using maximum likelihood algorithm employing GTR+G+I nucleotide substitution model in MEGA5 software programme (Tamura et al., 2011). The robustness of the tree topology was assessed by bootstrap analysis of 1000 pseudo-replicates.

2.5 Selection pressure analysis

Recombination analysis was carried out using GARD analysis available at open-source Datamonkey web server. Selection pressure analysis acting on the codons of hemagglutinin (HA), neuraminidase (NA), nucleoprotein (NP) and matrix protein (MP:M1,M2) of H1N1pdm virus was carried out using HyPhy open-source software package available under the datamonkey (http://www.datamonkey.org/) (Delport et al.,

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2010). Analysis was performed using reference sequences [n=115(HA); n=115(NA); n=115(NP) and n=115 (MP:M1, M2)] including globally circulating H1N1pdm virus. The ratio of non-synonnymous (dN) to synonymous (dS) substitutions per site (dN/dS) were estimated using four different approaches including: single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), mixed effects model of evolution (MEME) and fast unbiased bayesian approximation (FUBAR) (Muerell., 2012). Best nucleotide substitutions model for different data sets as determined through the available tool in Datamonkey server was adopted in the analysis.

2.6 Molecular clock analysis

Ninety six complete genome (13027 nt) sequences (including 86 from genogroup 6B) from eleven representative geographical locations were downloaded from the GISAID EpiFlu database were used for molecular clock analysis. The sampling dates ranged from 2009-2015 and sampling locations were Africa (n=8), Caribbean (n=3), Central Asia (n=1), Europe (n=14), Middle East (n=2), North America (n=16), North Asia (n=3), Oceania (n=7), South America (n=4), South Asia (n=20), and Southeast Asia (n=18). All sequences were aligned using Bioedit and MEGA softwares. Best fit model was selected for the analysis of this data set using MEGA version 6 software using model selection application. Time scaled phylogeny construction was performed using Bayesian MCMC approach (Beast v 1.7.4; http://beast.io.ed.ac.uk) (Drummond and Rambaut, 2007) implementing a HKY85 model using both strict and an uncorrelated lognormal relaxed clock model. Bayesian GMRF skyride population model was selected in order to perform discrete phylogeographic analysis (Lemey et al., 2009). MCMC analysis was conducted for 100 million generation and sampled at every 10,000 steps. Convergence was assessed on the basis of effective sampling size (ESS) after a 10% burn-in using Tracer software v1.5 (http://tree.bio.ed.ac.uk/software/tracer). All parameter estimates was indicated by 95%

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

highest posterior density (95% HPD) intervals. The tree was summarized in another application of beast analysis tree annotator program included in the beast package. Final tree was analyzed in Figtree v 1.4.0 for tree visualization.

2.7 Phylogeographic analysis

The continuous time markov chain (CTMC) process over discrete sampling locations implemented in BEAST (Drummond and Rambaut, 2007) was used for the phylgeographical analysis, implementing BSSVS (Bayesian stochastic search variable selection) model which allows the diffusion rates to be zero with a positive posterior probability. Comparison of the posterior and prior probability of the individual rates being zero provided a formal BF (Bayes factor) for testing the significant linkage between locations. Rates with a BF of >3 were considered well supported and formed the migration pathway. SPREAD application (Bielejec et al., 2011, http://www.phylogeography.org/SPREAD.html) was used to visualize and convert the estimated divergence time on spatial estimates annotated in the MCC tree to a key hole markup language (KML) file.

3. **Results**

We have previously reported H1N1 testing and HA gene sequencing for a cohort of patients from India in 2015. Here we performed full genome sequencing for 9 of those and got 5 complete genomes (Table 1). We sequenced 6 oropharyngeal swabs confirmed to be H1N1 by RT-PCR, 3 cell culture (in MDCK cells) isolates at passage 1 and 1 healthy oropharngeal swab at the peak of the Flu outbreak of 2015 in India (January-March, 2015). Unbiased metagenomic sequencing was performed with the aim of capturing variant viral genomes as well as other potentially pathogenic viruses. The reads from IonProton

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

sequencing were mapped using the A/California/07/2009 as the reference strain and complete genomes were recovered from 4 swabs and 1 isolate. The average genome length was 13133 bp, with good coverage across 8 segments/genes (Table 2). Contig statistics show an average read depth of about 100X except for the genome from the isolate which had an average read depth of 16X across all segments (Table 2). These five genomes (henceforth called 2015 genomes) were used for all subsequent analysis. The sequences had a similarity of 98.7 -100% similarity among themselves, with maximum sequence divergence in the Neuraminidase gene wiith the A/India/DRDE_GWL812/2015 showing maximum divergence. All 5 genomes were confirmed to be H1N1 and related to the 2009 pandemic strain.

We then compared the 2015 genomes to the prototype A/California/07/2009 and representative Indian isolates from 2009-2014. Amino acid sequence similarity among the set of 12 viruses was between 95.7-100% depending on protein. Many of the amino acid substitutions were found in residues on the surface of the protein (Fig 1.) 11 amino acid substitutions were consistently found in all 5 2015 genomes in the Hemagluttinin protein. 3 of the 11 substitutions were in the antigenic site of HA. (K180Q in Sa, S202T in Sb and S220T in Ca1) (Fig. 2). 13 mutations, common to all 5 2015 genomes were found in Neuraminidase. The characteristic H275Y Oseltamivir resistance mutation was found in one sample (with pre-exposure to the drug). 2, 7 4, 5, 4 and 5 amino acid substitution were identified in PB1, PB2, NP PA, MP and NS protein were consistently identified. Apart from these, multiple unique mutations were also noted in the different strains (Fig1 and Table 3).

Next, we performed recombination and selection pressure analysis to understand the evolution of different parts of the genome. Selection pressure analysis showed that a large number of codons were under strong purifying selection. 23 sites were identified to be under positive selection in HA, NA, NP and MP (Table 4), with 7 sites in HA, 6 in NA and

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

NP and 4 in MP. Of these position 239 in HA and positions 13, 248, 275 and 386 in NA were selected by all the different approaches used in this analysis.

In order to assess the relatedness of the Indian 2015 genomes to circulating Influenza A H1N1 genomes worldwide (2009-2015), we performed phylogenetic analysis of representative sequences of A(H1N1)pdm09 from distinct geographical regions. A maximum likelihood based phylogenetic tree revealed 10 distinct genogroups (1-5, 6A, 6B, 6C, 7 and 8) (Fig 3). While the prototype strain belongs to genogroup 1, the 2015 Indian genomes, grouped into genogroup 6B along with a large number of viruses circulating across the globe since 2013. Older isolates from India grouped into genogroup 1, 3, 4, 7, 6A and 6C. As predicted from the differences in amino acids, A/India/DRDE_GWL812/2015 grouped into a separate cluster (I) within genogroup 6B compared to the other 4 2015 genomes (cluster II). The individual gene segment wise phylogeny also revealed similar topology (results not shown).

The molecular clock analysis revealed the mean rate of nucleotide substitution among all the A(H1N1)pdm09 sequnces analysed in this study is 2.543×10^{-3} substitutions/site/year (95% HPD, 2.181×10^{-3} to 2.266×10^{-3}).Using a time to most recent common ancestor (TMRCA) approach, we found that the genogroup 6B emerged around 26 Dec 2011 (,with respect to the most recent isolate of 2015. The maximum clade credibility tree showed multiple clades, all supported by high posterior probabilities (>0.95) (Figure 4). This analysis predicted south east Asia as the ancestral state of genogroup 6B and classified it into 3 distinct clades which diverged at 28 Jan 2012, 01 Apr 2012 and 29 Apr 2012 3 (A/India/DRDE_GWL84/2015, A/India/DRDE_GWL897/2015, and A/India/DRDE_GWL989/2015) out of 5 Indian 2015 genomes clustered into clade 2a, 1 (A/India/DRDE_GWL672/2015) in clade 2b and A/India/DRDE_GWL812/2015 grouped into Clade 3 together with an isolate from Bangladesh (A/Bangladesh/01/2015) and a large

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

number of global isolates.

Phylogeographic analysis of the genogroup 6B viruses of A(H1N1)pdm09 using SPREAD showed the spatial diffusion of these viruses (Figure 5). The geograpahical diffusion pattern revealed 39 well supported transitions of viruses (BF>3) under the Bayesian stochastic search variable selection (BSSVS) model. In particular, migration events for the virus were noted between Africa – Europe (BF=36555), Central Asia – middle East (BF=1635), south America – south Asia (BF=1466) with highest credibility. The migration between Africa-North America and Caribbean-north America Europe-south Asia, central Asia-south America were also characterized by high BF values (>250) (Table S2).

4. **Discussion**

Influenza virus A(H1N1)pdm09, which causes swine flu in human, is a rapidly changing virus requiring a seasonal review of circulating strains and evaluation of vaccine effectiveness. The emergence of the quadruple assortment strain in 2009 and the resulting pandemic attest to the serious consequences of emergence of new strains of H1N1. In 2015, an unprecedented number of Influenza cases with increased severity were reported in India. In this study we obtained full length influenza virus sequences from 5 individuals during this outbreak. Analysis of the Hemaglutinin (HA) gene alone revealed 11 substitutions common to all 5 strains when compared to the prototype vaccine strain (A/California/07/2009), 3 of which were in the antigenic site. This is a matter of great concern as the probability of occurrence of a new flu epidemic has been linked either to appearance of >10 amino acid substitutions in the HA protein or > 4 amino acid substitutions in its antigenic sites (Park et al., 2009). Further, HA: K180Q in the Sa antigenic site was also found to be under strong positive selection. 2 of the substitutions that we have identified in the HA gene (HA1: S220T; HA2 E391K) have been reported to increase transmission of the virus.

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Surprisingly, the Neuraminidase (NA) gene was found to be most divergent both from analysis of nucleotides and translated amino acids and also harboured the maximum number of sites under positive selection. The presence of the characteristic Histidine to Tyrosine change in the drug binding pocket -NA: H275Y in A/India/DRDE_GWL84/2015 confirms the circulation of oseltamivir resistant virus in this outbreak. We also identified two novel mutations T72I and R257K in NA protein of this virus which needs further investigation. Additionally, we found the N248D substitution in the drug binding pocket, linked with reduction in sensitivity to Oseltamivir in all 5 strains. In addition to this, 3 substitutions (V241I, N369K, and N386K) identified in all the five Indian viruses were also consistently recorded in oseltamivir - resistant viruses isolated from China and Japan during 2013-14. These mutations were linked to offset the destabilizing effect of H275Y and enhance the virus transmission (Escuret et al., 2008, Huang et al., 2015; Koel et al., 2013; Koel et al., 2015). The NA: N386K in particular, leading to the loss of a glycosylation site, was found to be under strong selection pressure.

Codon specific selection pressures revealed strong purifying selection, as observed in other influenza viruses. The NA: V13I (in 4 of the 2015 strains) and NP: S498N (observed in Indian isolates of 2014 and 2015) were found to be under positive and strong positive selection respectively.

Molecular phylogeny analysis of the concatenated genome divides the strains from the H1N1 pandemic 2009 strains into 10 genogroups. Of these groups, 6B is the most diverse and contains within it most of the viruses isolated since 2013 including the 5 2015 Indian strains. Our analysis estimates that the common ancestor for genogroup 6B arose in end of 2011 with south east Asia as the ancestral state. Consistent with previous results we also dated the emergence of the pandemic 2009 strain to mid 2008. (Fraser et al., 2009, Smith et al., 2009).

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Unlike many other pathogens which cluster geographically, influenza viruses generally follow temporal clustering pattern. Maximum clade credibility (MCC) phylogenetic analysis demonstrates the existence of multiple clusters among genogroup 6B (clusters are confirmed on the basis of more than 90% posterior probability values). The existence of multiple clusters confirms the idea that the virus is evolving continuously, raising the possibility of pandemic strain emergence.

This study is the first extensive complete genome analysis of 2015 Indian A(H1N1)pdm09 viruses. The identification of mutant A(H1N1)pdm09 viruses, particularly the presence of so many mutations in HA and as well as in NA are quite important both from surveillance, prediction of outbreak and therapeutic intervention.

REFERENCES

Bielejec F., Rambaut A., Suchard M.A., Lemey P., 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics, doi:10.1093/bioinformatics/btr481.

Broor S., Krishnan A., Roy D.S., Dhakad S., Kaushik S., Mir M.A., Singh Y., Moen A., Chadha M., Mishra A.C., Lal R.B., 2012. Dynamic patterns of circulating seasonal and pandemic A (H1N1) pdm09 influenza viruses from 2007–2010 in and around Delhi, India. PloS One, doi:10.1371/journal.pone.0029129.

Campbell P.J., Kyriakis C.S., Marshall N., Suppiah S., Seladi-Schulman J., Danzy S., Lowen A.C., Steel J., 2014. Residue 41 of the Eurasian avian-like swine influenza a virus matrix protein modulates virion filament length and efficiency of contact transmission. Journal of Virology, doi:10.1128/JVI.00119-14.

Chevreux B., Pfisterer T., Drescher B., Driesel A.J., Müller W.E., Wetter T., Suhai S., 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Research, doi:10.1101/gr.1917404.

Dakhave M., Khirwale A., Patil K., Kadam A., Potdar V., 2013. Whole-genome sequence analysis of postpandemic influenza A (H1N1) pdm09 virus isolates from India. Genome Announcements, doi:10.1128/genomeA.00727- 13.

Dash PK, Pattabiraman C, Tandel K, Sharma S, Kumar JS, Siddappa S, Gowda M, Krishna S,

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Parida M., 2018. Recovery of Five Complete Influenza A(H1N1)pdm09 Genome Sequences from the 2015 Influenza Outbreak in India by Metagenomic Sequencing. Genom Announcements, doi: 10.1128/genomeA.00511-18.

Delport W., Poon A.F., Frost S.D., Pond S.L., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics, doi:10.1093/bioinformatics/btq429.

Drummond A.J., Rambaut A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, doi:10.1186/1471-2148-7-214.

Edgar R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5):1792-7. PubMed PMID:15034147.

Escuret V., Frobert E., Bouscambert-Duchamp M., Sabatier M., Grog I., Valette M., Lina B., Morfin F., Ferraris O., 2008. Detection of human influenza A (H1N1) and B strains with reduced sensitivity to neuraminidase inhibitors. Journal of Clinical Virology, doi:10.1016/j.jcv.2007.10.019.

Fraser C., Donnelly C.A., Cauchemez S., Hanage W.P., Van Kerkhove M.D., Hollingsworth T.D., Griffin J., Baggaley R.F., Jenkins H.E., Lyons E.J., Jombart T., 2009. Pandemic potential of a strain of influenza A (H1N1): early findings. Science, doi:10.1126/science.1176062.

Garten R.J., Davis C.T., Russell C.A., Shu B., Lindstrom S., Balish A., Sessions W.M., Xu X., Skepner E., Deyde V., Okomo-Adhiambo M., 2009. Antigenic and genetic characteristics of swine-origin 2009 A (H1N1) influenza viruses circulating in humans. Science, doi:10.1126/science.1176225.

Huang W., Li X., Cheng Y., Tan M., Guo J., Wei H., Zhao X., Lan Y., Xiao N., Wang Z., Wang D., 2015. Characteristics of oseltamivir-resistant influenza A (H1N1) pdm09 virus during the 2013–2014 influenza season in Mainland China. Journal of Virology, doi:10.1186/s12985-015-0317-1.

Human Infection with Pandemic (H1N1) 2009 Virus: Updated Interim WHO Guidance on Global Surveillance". Available from http://www.who.int/csr/disease/swineflu/guidance/surveillance/WHO_case_definition_swine_flu_2009_04_29.pdf?ua=1) accessed on 22 aug 2016.

Koel B.F., Burke D.F., Bestebroer T.M., van der Vliet S., Zondag G.C., Vervaet G., Skepner E., Lewis N.S., Spronken M.I., Russell C.A., Eropkin M.Y., 2013. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science, doi:10.1126/science.1244730.

Koel B.F., Mogling R., Chutinimitkul S., Fraaij P.L., Burke D.F., van der Vliet S., de Wit E., Bestebroer T.M., Rimmelzwaan G.F., Osterhaus A.D., Smith D.J., 2015. Identification of amino acid substitutions supporting antigenic change of influenza A (H1N1) pdm09 viruses.

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Journal of Virology, doi:10.1128/JVI.02962-14.

Kosoltanapiwat N., Pooruk P., Iamsirithaworn S., Mungaomklang A., Chokephaibulkit K., Auewarakul P., Puthavathana P., 2014. Amino acid substitutions in hemagglutinin of the 2009 pandemic influenza A (H1N1) viruses that might affect the viral antigenicity. BMC Research Notes, doi:10.1186/1756- 0500-7-951.

Lemey P., Rambaut A., Drummond A.J., Suchard M.A., 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology, doi:10.1371/journal.pcbi.1000520.

Matsuzaki Y., Sugawara K., Nakauchi M., Takahashi Y., Onodera T., Tsunetsugu- Yokota Y., Matsumura T., Ato M., Kobayashi K., Shimotai Y., Mizuta K., 2014. Epitope mapping of the hemagglutinin molecule of A/(H1N1) pdm09 influenza virus by using monoclonal antibody escape mutants. Journal of Virology, doi:10.1128/JVI.01381-14.

Murrell B., Moola S., Mabona A., Weighill T., Sheward D., Pond S.L., Scheffler K., 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Molecular Biology and Evolution, doi.org/10.1093/molbev/mst030.

Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team. 2009. Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans. New England Journal of Medicine, doi:10.1056/NEJMoa0903810.

Parida M.M., Dash P.K., Kumar J.S., Joshi G., Tandel K., Sharma S., Ambuj., Agarwal A., Saha A., Saraswat S., Karothia D.,Malviya V., 2016. Emergence of Influenza A (H1N1) pdm09 Genogroup 6B and Drug Resistant Virus, India, January to May 2015. Euro Surveillance, doi:10.2807/1560-7917.ES.2016.21.5.30124.

Park A.W., Daly J.M., Lewis N.S., Smith D.J., Wood J.L., Grenfell B.T., 2009. Quantifying the impact of immune escape on transmission dynamics of influenza. Science, doi:10.1126/science.1175980.

Potdar V.A., Chadha M.S., Jadhav S.M., Mullick J., Cherian S.S., Mishra A.C., 2010. Genetic characterization of the influenza A pandemic (H1N1) 2009 virus isolates from India. PloS One, doi:10.1371/journal.pone.0009693.

Smith G.J., Vijaykrishna D., Bahl J., Lycett S.J., Worobey M., Pybus O.G., Ma S.K., Cheung C.L., Raghwani J., Bhatt S., Peiris J.M., 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature, doi:10.1038/nature08182.

Takashita E., Meijer A., Lackenby A., Gubareva L., Rebelo-de-Andrade H., Besselaar T., Fry A., Gregory V., Leang S.K., Huang W., Lo J., 2015. Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors, 2013– 2014. Antiviral Research, doi:10.1016/j.antiviral.2015.02.003.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S., 2011. MEGA5: molecular

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, doi:10.1093/molbev/msr121.

Acknowledgements

This work was funded by the DRDO. The authors are thankful to Dr D.K. Dubey, Director, Defence Research and Development Establishment, Ministry of Defence, Government of India, for his support, constant inspiration and providing the necessary facilities for this study. The authors are also thankful to the Commissioner, Health Services, Govt of MP, IDSP network and CMHO, respective districts of MP for providing the clinical samples. The authors are thankful to GISAID and all the contributors of A(H1N1)pdm09 sequences, which were utilised in this study.

Figure Legends

Fig 1. 3D structure of prototype A/California/04/2009 strain showing the critical amino acid substitutions. Figure shows the location of amino acid substitutions for Influenza Hemagglutinin (A and B – different rotational views) and Neuraminidase (C) . Homology modeling was performed using SWISS-MODEL and the best model for the California/04/2009 sequence was chosen for showing the substitutions. The protein surfaces as visualized by PyMOL are shown in light brown, amino acid that were found to be different in the Gwalior 2015 strains are colored by elements CHNOS. * represents amino acid not found on the surface. ** represent amino acid inside an accessible pocket. 13/16 substitutions for Hemagglutinin and 14/21 substitutions for Neuraminidase are shown, as the other residues were not present in the model.

Fig 2. Figure showing amino acid substitutions in HA1 protein and Neuraminidase (NA) of 2015 Indian A(H1N1)pdm09 viruses in comparison to prototype

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A/California/07/2009 vaccine strain.

Fig 3. Phylogenetic tree among A(H1N1)pdm09 viruses generated by maximum likelihood method based on the nucleotide sequence of concatenated whole genome of geographically diverse viruses.

Fig 4. Maximum clade credibility (MCC) tree of A(H1N1)pdm09 viruses constructed using BEAST version 1.7.4. Colors of branches indicate geographic locations as per the color key. Branch lengths correspond to lengths of time. The letters at the nodes correspond to posterior probability values and divergence date.

Fig. 5. Map showing the global dispersal of cosmopolitan genotype of A(H1N1)pdm09 viruses, belonging to genogroup 6B. The migration route is inferred from branching pattern of maximum clade credibility (MCC) tree, as prepared using SPREAD software. The migration route supported by BF values > 3 is considered significant and is shown in this figure.

Fig 1

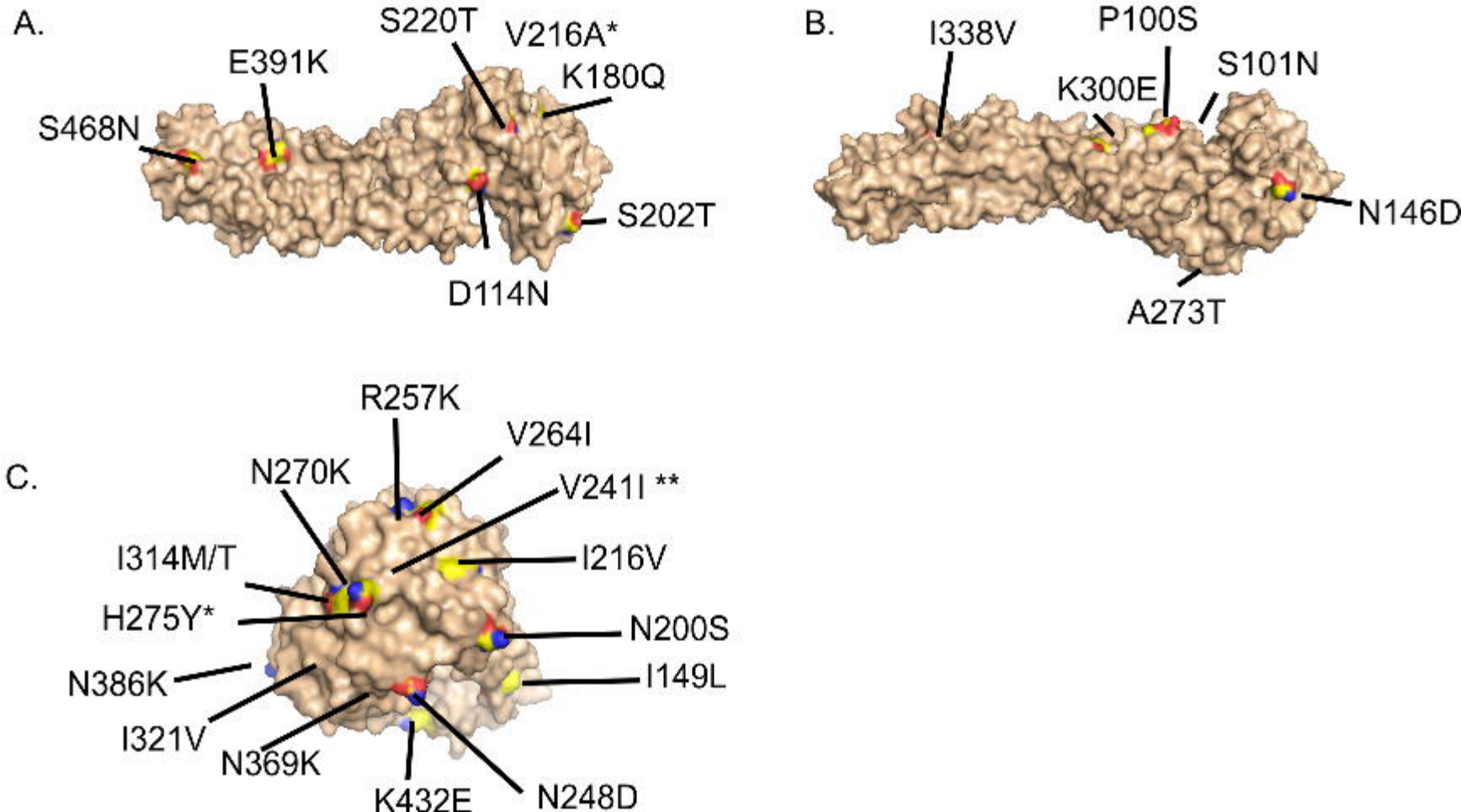
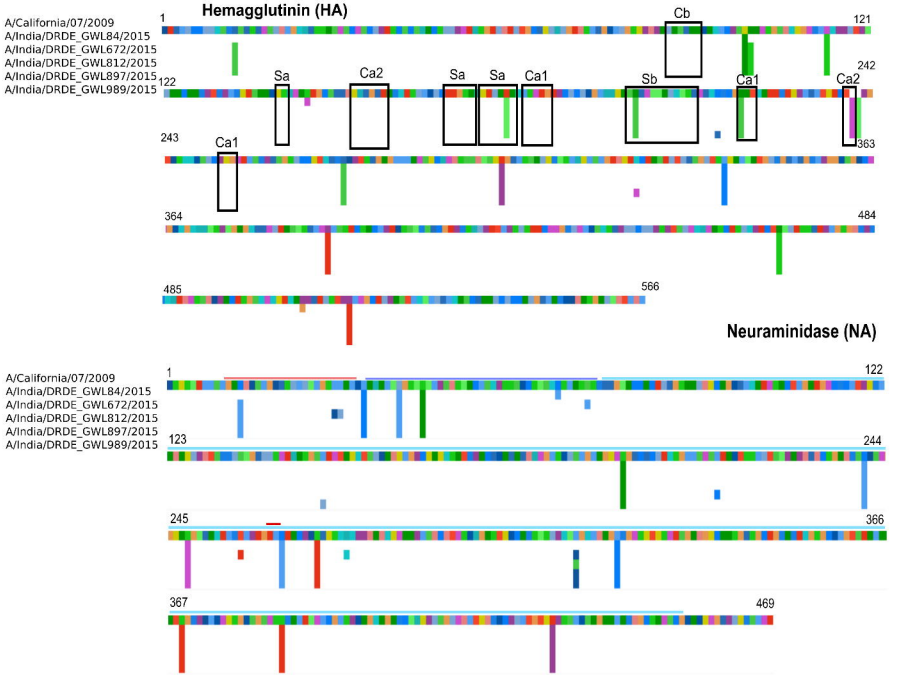
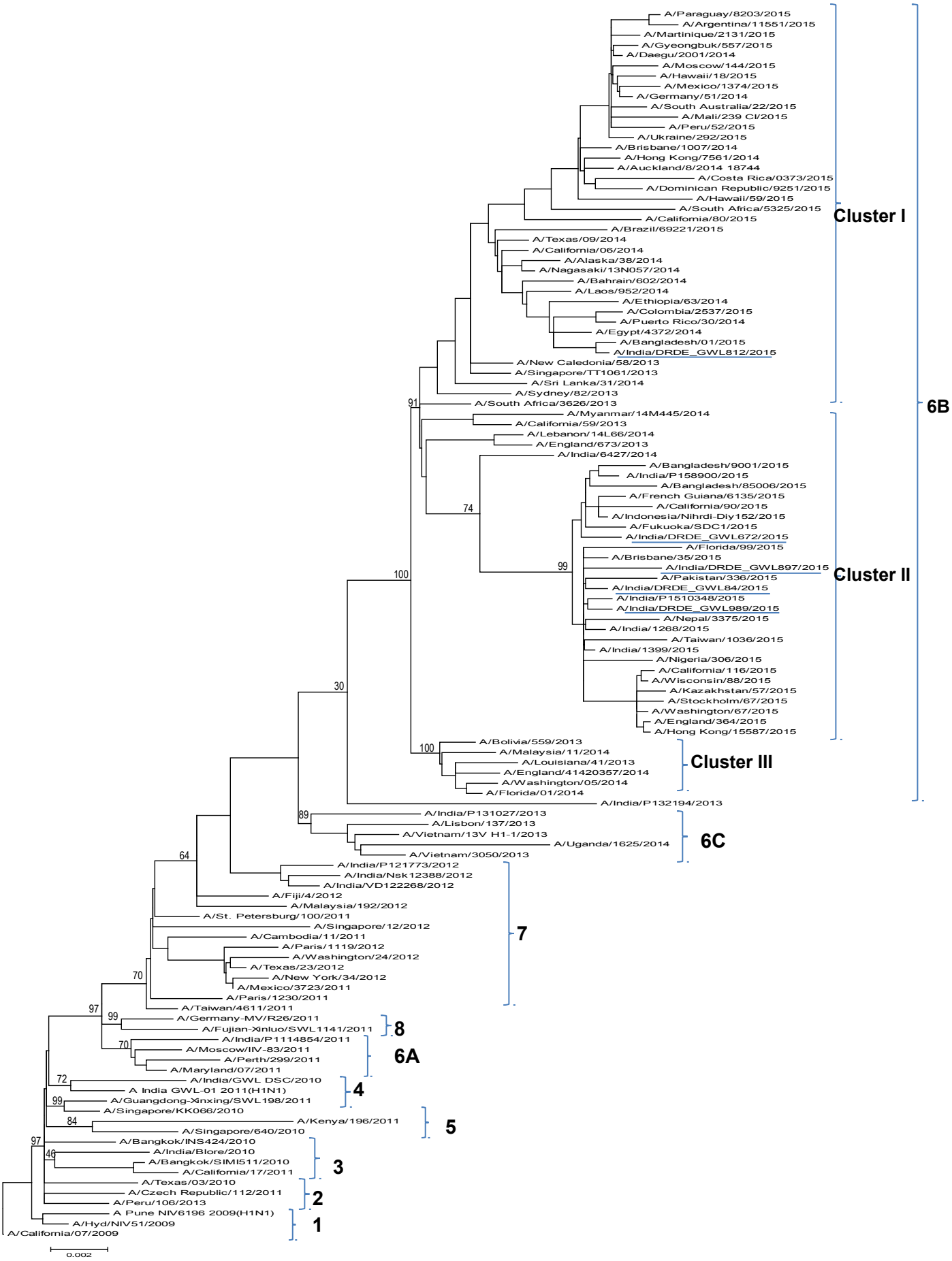
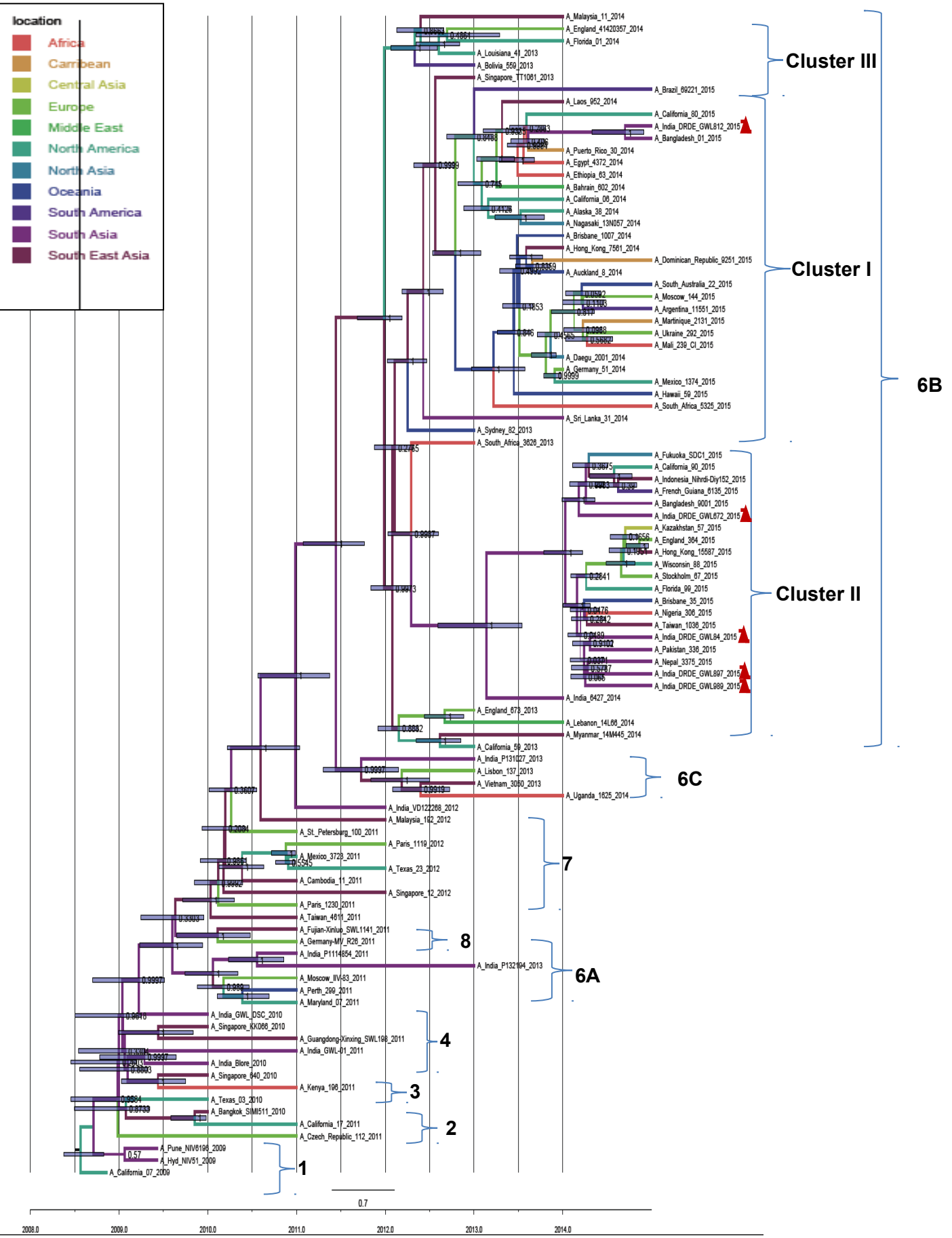

A.
S220T
V216A*
E391K
K180Q
S468N
S202T
D114N
B.
I338V
P100S
S101N
K300E
N146D
A273T
C.
R257K
V264I
N270K
V241I **
I314M/T
I216V
H275Y*
N200S
N386K
I149L
I321V
N369K
K432E
N248D

Fig 2


Hemagglutinin (HA)
A/California/07/2009
A/India/DRDE_GWL84/2015
A/India/DRDE_GWL672/2015
A/India/DRDE_GWL812/2015
A/India/DRDE_GWL897/2015
A/India/DRDE_GWL989/2015
1
121
122
242
243
363
364
484
485
566
Cb
Sa
Ca2
Sa
Sa
Ca1
Sb
Ca1
Ca2
Ca1
Neuraminidase (NA)
A/California/07/2009
A/India/DRDE_GWL84/2015
A/India/DRDE_GWL672/2015
A/India/DRDE_GWL812/2015
A/India/DRDE_GWL897/2015
A/India/DRDE_GWL989/2015
1
122
123
244
245
366
367
469


A/Paraguay/8203/2015
A/Argentina/11551/2015
A/Martinique/2131/2015
A/Gyeongbuk/557/2015
A/Daegu/2001/2014
A/Moscow/144/2015
A/Hawaii/18/2015
A/Mexico/1374/2015
A/Germany/51/2014
A/South Australia/22/2015
A/Mali/239 CI/2015
A/Peru/52/2015
A/Ukraine/292/2015
A/Brisbane/1007/2014
A/Hong Kong/7561/2014
A/Auckland/8/2014 18744
A/Costa Rica/0373/2015
A/Dominican Republic/9251/2015
A/Hawaii/59/2015
A/South Africa/5325/2015
A/California/80/2015
A/Brazil/69221/2015
A/Texas/09/2014
A/California/06/2014
A/Alaska/38/2014
A/Nagasaki/13N057/2014
A/Bahrain/602/2014
A/Laos/952/2014
A/Ethiopia/63/2014
A/Colombia/2537/2015
A/Puerto Rico/30/2014
A/Egypt/4372/2014
A/Bangladesh/01/2015
A/India/DRDE_GWL812/2015
A/New Caledonia/58/2013
A/Singapore/TT1061/2013
A/Sri Lanka/31/2014
A/Sydney/82/2013
A/South Africa/3626/2013
A/Myanmar/14M445/2014
A/California/59/2013
A/Lebanon/14L66/2014
A/England/673/2013
A/India/6427/2014
A/Bangladesh/9001/2015
A/India/P158900/2015
A/Bangladesh/85006/2015
A/French Guiana/6135/2015
A/California/90/2015
A/Indonesia/Nihrdi-Diy152/2015
A/Fukuoka/SDC1/2015
A/India/DRDE_GWL672/2015
A/Florida/99/2015
A/Brisbane/35/2015
A/India/DRDE_GWL897/2015
A/Pakistan/336/2015
A/India/DRDE_GWL84/2015
A/India/P1510348/2015
A/India/DRDE_GWL989/2015
A/Nepal/3375/2015
A/India/1268/2015
A/Taiwan/1036/2015
A/India/1399/2015
A/Nigeria/306/2015
A/California/116/2015
A/Wisconsin/88/2015
A/Kazakhstan/57/2015
A/Stockholm/67/2015
A/Washington/67/2015
A/England/364/2015
A/Hong Kong/15587/2015
A/Bolivia/559/2013
A/Malaysia/11/2014
A/Louisiana/41/2013
A/England/41420357/2014
A/Washington/05/2014
A/Florida/01/2014
A/India/P132194/2013
A/India/P131027/2013
A/Lisbon/137/2013
A/Vietnam/13V H1-1/2013
A/Uganda/1625/2014
A/Vietnam/3050/2013
A/India/P121773/2012
A/India/Nsk12388/2012
A/India/VD122268/2012
A/Fiji/4/2012
A/Malaysia/192/2012
A/St. Petersburg/100/2011
A/Singapore/12/2012
A/Cambodia/11/2011
A/Paris/1119/2012
A/Washington/24/2012
A/Texas/23/2012
A/New York/34/2012
A/Mexico/3723/2011
A/Paris/1230/2011
A/Taiwan/4611/2011
A/Germany-MV/R26/2011
A/Fujian-Xinluo/SWL1141/2011
A/India/P1114854/2011
A/Moscow/IIV-83/2011
A/Perth/299/2011
A/Maryland/07/2011
A/India/GWL DSC/2010
A India GWL-01 2011(H1N1)
A/Guangdong-Xinxing/SWL198/2011
A/Singapore/KK066/2010
A/Kenya/196/2011
A/Singapore/640/2010
A/Bangkok/INS424/2010
A/India/Blore/2010
A/Bangkok/SIMI511/2010
A/California/17/2011
A/Texas/03/2010
A/Czech Republic/112/2011
A/Peru/106/2013
A Pune NIV6196 2009(H1N1)
A/Hyd/NIV51/2009
A/California/07/2009
Cluster I
Cluster II
Cluster III
6B
6C
7
8
6A
4
5
3
2
1
91
74
99
100
100
30
89
64
70
97
99
70
72
99
84
97
46
0.002


location
Africa
Caribbean
Central Asia
Europe
Middle East
North America
North Asia
Oceania
South America
South Asia
South East Asia
Cluster III
Cluster I
Cluster II
6B
6C
7
8
6A
4
3
2
1
A_Malaysia_11_2014
A_England_41420357_2014
A_Florida_01_2014
A_Louisiana_41_2013
A_Bolivia_559_2013
A_Singapore_TT1061_2013
A_Brazil_69221_2015
A_Laos_952_2014
A_California_80_2015
A_India_DRDE_GWL812_2015
A_Bangladesh_01_2015
A_Puerto_Rico_30_2014
A_Egypt_4372_2014
A_Ethiopia_63_2014
A_Bahrain_602_2014
A_California_06_2014
A_Alaska_38_2014
A_Nagasaki_13N057_2014
A_Brisbane_1007_2014
A_Hong_Kong_7561_2014
A_Dominican_Republic_9251_2015
A_Auckland_8_2014
A_South_Australia_22_2015
A_Moscow_144_2015
A_Argentina_11551_2015
A_Martinique_2131_2015
A_Ukraine_292_2015
A_Mali_239_CI_2015
A_Daegu_2001_2014
A_Germany_51_2014
A_Mexico_1374_2015
A_Hawaii_59_2015
A_South_Africa_5325_2015
A_Sri_Lanka_31_2014
A_Sydney_82_2013
A_South_Africa_3826_2013
A_Fukuoka_SDC1_2015
A_California_90_2015
A_Indonesia_Nihrdi-Diy152_2015
A_French_Guiana_6135_2015
A_Bangladesh_9001_2015
A_India_DRDE_GWL672_2015
A_Kazakhstan_57_2015
A_England_364_2015
A_Hong_Kong_15587_2015
A_Wisconsin_88_2015
A_Stockholm_67_2015
A_Florida_99_2015
A_Brisbane_35_2015
A_Nigeria_306_2015
A_Taiwan_1036_2015
A_India_DRDE_GWL84_2015
A_Pakistan_336_2015
A_Nepal_3375_2015
A_India_DRDE_GWL897_2015
A_India_DRDE_GWL989_2015
A_India_6427_2014
A_England_673_2013
A_Lebanon_14L66_2014
A_Myanmar_14M445_2014
A_California_59_2013
A_India_P131027_2013
A_Lisbon_137_2013
A_Vietnam_3060_2013
A_Uganda_1625_2014
A_India_VD122268_2012
A_Malaysia_162_2012
A_St._Petersburg_100_2011
A_Paris_1119_2012
A_Mexico_3723_2011
A_Texas_23_2012
A_Cambodia_11_2011
A_Singapore_12_2012
A_Paris_1230_2011
A_Taiwan_4611_2011
A_Fujian-Xinluo_SWL1141_2011
A_Germany-MV_R26_2011
A_India_P1114854_2011
A_India_P132194_2013
A_Moscow_IIV-83_2011
A_Perth_299_2011
A_Maryland_07_2011
A_India_GWL_DSC_2010
A_Singapore_KK066_2010
A_Guangdong-Xinxing_SWL198_2011
A_India_GWL-01_2011
A_India_Blore_2010
A_Singapore_640_2010
A_Kenya_196_2011
A_Texas_03_2010
A_Bangkok_SIMI511_2010
A_California_17_2011
A_Czech_Republic_112_2011
A_Pune_NIV6196_2009
A_Hyd_NIV51_2009
A_California_07_2009
0.7
2008.0
2009.0
2010.0
2011.0
2012.0
2013.0
2014.0

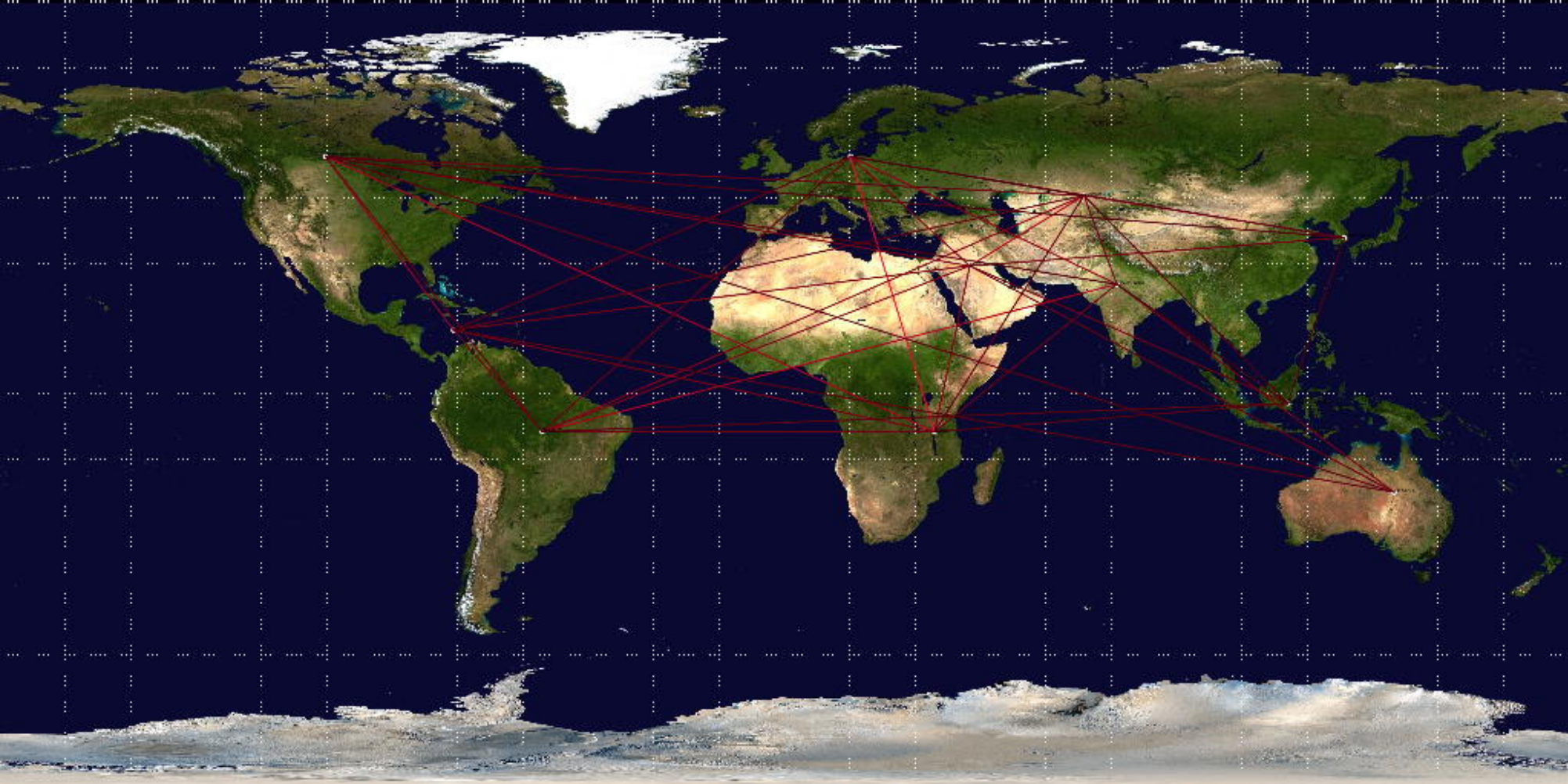


Table 1: Detailed patient profile of pandemic swine flu H1N1 investigated in this study

Virus ID	Sex	Age (Years)	Location	Category	Treatment taken	Ct value	Date of sample collection	Day since onset of symptoms	X – Ray Findings
A/India/DRDE_GWL812/2015	F	55	BHOPAL	C	Yes	24	9 Mar 16	08 Days	B/L Homogenous Opacities
A/India/DRDE_GWL84/2015	M	35	BHOPAL	C	Yes	26	17 Mar 16	08 Days	B/L Pneumonitis
A/India/DRDE_GWL672/2015	F	42	BHOPAL	C	No	24	05 Mar 16	04 Days	B/L Homogenous Opacities
A/India/DRDE_GWL897/2015	M	56	GWALIOR	A	Yes	23	14 Mar 16	04 Days	Nil
A/India/DRDE_GWL989/2015	F	09	SEHORE	B	Yes	23	18 Mar 16	03 Days	B/L Hyperinflation with Pneumonitis changes

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 2: Contig statistics of mapping assemblies

a. Contig Length of Samples

Gene Name	**Contig Length of Samples**				
	A/India/DRDE_GWL84/2015	**A/India/DRDE_GWL672/2015**	**A/India/DRDE_GWL812/2015**	**A/India/DRDE_GWL897/2015**	**A/India/DRDE_GWL989/2015**
NS1	2280	2281	2281	2283	2280
NP	2274	2274	2274	2275	2277
M1/M2	2151	2151	2151	2154	2151
PA	1704	1704	1701	1705	1703
NA	1497	1497	1496	1496	1498
HA	1410	1415	1411	1411	1410
PB2	982	982	982	983	983
PB1	862	863	862	863	863
Total Assembly	**13160**	**13167**	**13158**	**13170**	**13165**

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

b. No. of Reads in Assembly

Gene Name	No. of Reads in Assembly				
	A/India/DRDE_GWL84/2015	A/India/DRDE_GWL672/2015	A/India/DRDE_GWL812/2015	A/India/DRDE_GWL897/2015	A/India/DRDE_GWL989/2015
NS1	5837	1093	4777	2065	235
NP	4219	1104	3215	1343	373
M1/M2	4224	1433	5341	2145	292
PA	3523	1024	2677	1398	145
NA	3776	1832	3623	1342	298
HA	3190	768	3756	1426	173
PB2	1993	981	2016	1097	172
PB1	1537	481	1517	733	138
Total No. of Reads	**28299**	**8716**	**26922**	**11549**	**1826**

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

c. Average Coverage

Gene Name	Average Coverage				
	A/India/DRDE_GWL84/2015	A/India/DRDE_GWL672/2015	A/India/DRDE_GWL812/2015	A/India/DRDE_GWL897/2015	A/India/DRDE_GWL989/2015
NS1	303.83	83.13	256.99	101.36	11.7
NP	236.46	83.2	169.32	65.07	18.84
M1/M2	234.87	117.41	293.32	108.14	14.99
PA	256.24	102.01	187.12	90.19	10.06
NA	310.13	218.53	288.1	90.9	22.82
HA	273.27	98.09	333.66	115.24	13.79
PB2	245.7	174.42	240.48	123.25	18.8
PB1	215.54	90	206.59	93.16	18.66

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

d. Average Quality

Gene Name	Average Quality				
	A/India/DRDE_GWL84/2015	A/India/DRDE_GWL672/2015	A/India/DRDE_GWL812/2015	A/India/DRDE_GWL897/2015	A/India/DRDE_GWL989/2015
NS1	50	49	42	43	50
NP	47	53	42	41	53
M1/M2	49	54	42	42	52
PA	52	55	44	43	54
NA	54	62	45	45	54
HA	53	52	42	44	51
PB2	53	62	48	46	54
PB1	48	50	46	42	55

Note: Gene names refer to Influenza A segments coding for the following proteins –

- NS1 – Non-structural protein 1
- NP – Nucleoprotein,
- M1/M2 – Matrix protein,
- PA – Polymerase acidic protein
- NA – Neuraminidase,
- HA – Hemagglutinin,
- PB2 – Polymerase basic protein2

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

- PB 1 – Polymerase basic protein 1, RNS directed RNA polymerase (minimal, essential submit)

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 3: Amino acid substitutions among the five Indian H1N1pdm virus (sequenced in this study) compared to prototype H1N1pdm strain (A/California/07/2009) and other Indian viruses (2009-2014)

Position	A/ California/ 07/2009	A/India/DRDE_GWL812/2015	A/India/DRDE_GWL84/2015	A/India/DRDE_GWL672/2015	A/India/DRDE_GWL897/2015	A/India/DRDE_GWL989/2015	A/Hyd_NIV51/2009	A/India/GWL_DSC/2010	A/India/Gwl_01/2011	A/India/Nsk12388/2012	A/India/P132194/2013	A/India/Ind_6427/2014
PB2												
54	R	K	K	K	K	K	-	K	K	-	-	K
66	M	I	I	I	I	I	-	-	-	-	-	I
104	P	-	-	-	-	-	-	-	-	-	L	-
125	L	-	-	-	-	-	-	F	-	-	-	-
183	L	-	-	-	-	-	-	-	M	-	-	-

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

195	D	N	N	N	N	N	-	-	-	N	-	N
216	R	-	-	-	-	-	-	-	G	-	-	-
241	E	-	-	-	-	-	-	G	-	-	-	-
278	A	-	-	-	-	-	-	V	V	-	-	-
288	Q	-	-	-	-	-	-	P	-	-	-	-
293	R	K	K	K	K	K	-	-	-	K	-	K
299	R	-	K	K	K	K	-	-	-	-	K	-
340	K	-	-	-	-	-	-	T	T	-	-	-
344	V	M	M	M	M	M	-	-	-	M	M	M
354	I	L	L	L	L	L	-	-	-	L	L	L
439	Q	-	-	-	-	-	-	H	-	-	-	-
441	D	-	-	-	-	-	-	H	-	-	-	-
453	S	P	T	T	T	T	-	-	-	-	-	-
456	N	-	-	-	-	-	-	-	-	-	S	-
588	T	-	-	-	-	-	-	-	I	-	-	-
731	V	I	I	I	I	I	-	-	-	I	-	I

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

758	I	-	-	-	-	-	-	L	-	-	-	-
PB1												
154	G	D	D	D	D	D	-	-	-	D	-	D
213	N	S	-	-	-	-	-	-	-	-	-	-
226	T	-	-	-	-	-	-	S	-	-	-	-
229	K	-	-	-	-	-	-	I	-	-	-	-
236	L	-	-	-	-	-	-	V	-	-	-	-
254	F	-	-	-	-	-	L	-	-	-	-	-
323	T	-	-	-	-	-	-	K	-	-	-	-
397	I	M	M	M	M	M	-	-	M	M	-	M
633	S	-	-	-	-	-	-	-	-	T	-	-
652	A	-	-	-	-	-	-	-	-	-	V	-
667	I	-	-	-	-	-	-	-	-	S	-	-
679	K	-	-	-	-	-	-	-	-	T	-	-
NS1												
55	E	K	K	K	K	K	-	-	-	-	-	K

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

71	E	-	-	-	-	-	-	K	-	-	-	-
86	T	-	-	-	I	-	-	-	-	-	-	-
90	L	I	I	I	I	I	-	-	-	I	-	I
123	I	V	V	V	V	V	-	V	V	V	V	V
129	V	-	-	-	-	-	-	-	-	-	I	-
131	K	E	E	E	E	E	-	-	-	-	-	E
205	N	S	S	S	S	S	-	-	-	S	S	S
206	C	-	-	-	-	-	-	-	-	Y	-	-
NS2												
29	N	S	S	S	S	S	-	-	-	-	-	S
48	T	A	A	A	A	A	-	-	-	A	A	A
49	V	-	-	-	I	-	-	-	-	M	-	-
83	M	-	I	I	I	I	-	-	-	-	-	-
115	A	-	-	-	-	-	T	-	-	-	-	-
M1												
28	L	-	-	-	-	-	-	M	-	-	-	-

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

30	S	-	-	-	-	-	-	-	-	-	-	N
80	V	I	I	I	I	I	-	I	-	I	I	I
108	T	-	-	-	K	-	-	-	-	-	-	-
169	T	-	-	-	A	-	-	-	-	-	-	-
170	N	-	-	-	E	-	-	-	-	-	-	-
192	M	V	V	V	V	V	-	-	-	-	-	V
208	Q	-	K	K	-	K	-	-	-	-	-	-
227	A	-	-	T	-	-	-	-	-	-	-	-
230	K	R	R	R	R	R	-	-	-	R	R	R
M2												
21	D	G	G	G	G	G	-	-	-	G	G	G
27	V	-	A	-	-	-	-	-	-	-	-	-
HA												
8	L	-	-	-	-	-	-	-	-	M	-	-
13	A	-	T	T	T	T	-	-	-	-	-	T
15	A	-	-	-	-	-	T	-	-	-	-	-

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

100	P	S	S	S	S	S	S	S	S	S	S	S
101	S	-	N	N	N	N	-	-	-	G	-	N
114	D	N	N	N	N	N	-	-	-	-	N	N
146	N	D	-	-	-	-	-	-	-	-	-	-
160	S	-	-	-	-	-	-	-	-	G	-	-
180	K	Q	Q	Q	Q	Q	-	-	-	-	Q	Q
202	S	T	T	T	T	T	-	T	-	T	T	T
214	A	-	-	-	-	-	-	-	-	T	-	-
216	V	-	-	-	-	A	-	-	-	-	-	-
220	S	T	T	T	T	T	-	T	T	T	T	T
233	I	-	-	-	-	-	-	-	-	V	-	-
239	D	-	-	-	-	-	-	-	-	-	-	N
244	M	-	-	-	-	-	-	-	-	I	-	-
245	N	-	-	-	-	-	-	-	I	-	-	-
246	Y	-	-	-	-	-	-	-	N	-	-	-
273	A	T	T	T	T	T	-	-	-	-	-	T

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

300	K	E	E	E	E	E	-	-	-	-	E	E
338	I	V	V	V	V	V	V	V	V	V	V	V
391	E	K	K	K	K	K	-	K	K	K	K	K
435	I	-	-	-	-	-	-	T	-	-	-	-
436	W	-	-	-	-	-	-	P	-	-	-	-
442	L	-	-	-	-	-	-	-	R	-	-	-
468	S	N	N	N	N	N	-	N	-	N	N	N
508	E	G	-	-	-	-	-	-	-	-	-	-
516	E	K	K	K	K	K	-	-	-	-	K	K
PA												
28	P	-	-	-	-	-	Q	-	-	-	-	-
30	I	-	-	-	-	-	-	-	N	-	-	-
100	V	I	I	I	I	I	-	-	-	-	I	I
127	V	-	I	-	-	-	-	-	-	-	-	-
142	K	-	-	-	-	-	-	-	N	-	-	-
167	S	-	-	-	-	-	-	T	-	-	-	-

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

168	R	-	-	-	-	-	-	T	-	-	-	-
224	P	S	S	S	S	S	S	S	S	S	S	S
311	M	-	-	V	-	-	-	-	-	-	-	-
321	N	K	K	K	K	K	-	S	S	K	K	K
330	I	V	V	V	V	V	-	-	-	-	V	V
343	A	-	-	-	-	-	-	-	-	T	-	-
361	K	R	-	-	-	-	-	-	-	-	-	-
362	R	K	K	K	K	K	-	-	-	-	-	K
395	S	-	-	-	-	-	-	-	-	C	-	-
473	S	-	A	-	-	-	-	-	-	-	-	-
514	D	-	-	-	-	-	-	Y	-	-	-	-
534	P	-	-	-	-	S	-	-	-	-	-	-
542	V	-	-	-	-	-	-	-	-	A	-	-
551	R	-	-	-	-	-	-	-	-	K	-	-
640	L	-	-	-	-	-	-	F	-	-	-	-
643	K	-	-	-	-	-	-	R	-	-	-	-

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

NP												
22	A	T	T	T	T	T	-	-	-	-	-	-
67	V	-	-	-	-	-	-	G	-	-	-	-
100	V	I	I	I	I	I	I	I	I	I	I	I
103	K	-	-	-	-	-	-	Q	-	-	-	-
104	W	-	-	-	-	-	-	C	-	-	-	-
105	M	T	T	T	T	T	-	-	-	-	-	-
119	V	-	-	-	-	-	-	-	-	-	I	-
232	T	-	-	-	-	-	P	-	-	-	-	-
446	R	-	-	-	-	-	-	-	-	-	-	K
483	N	-	-	-	G	-	-	-	-	-	-	-
498	S	N	N	N	N	N	-	-	-	-	-	N
NA												
13	V	-	I	I	I	I	-	-	-	-	I	-
29	I	-	-	M	-	-	-	-	-	-	-	-
30	I	-	-	L	-	-	V	-	-	-	-	-

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

31	S	-	-	Q	-	-	-	-	-	-	-	-
34	I	V	V	V	V	V	-	-	-	-	-	V
40	L	I	I	I	I	I	-	-	-	-	-	-
44	N	S	S	S	S	S	-	-	-	S	-	S
51	Q	-	-	-	-	-	-	-	-	-	-	R
67	V	I	-	-	-	-	-	-	-	-	-	-
72	T	-	I	-	-	-	-		-	-	-	-
83	V	-	-	-	-	-	-	-	-	-	L	-
95	S	-	-	-	-		-	-	-	-	-	G
106	V	-	-	-	-	-	I	I	I	-	I	-
114	V	-	-	-	-	-	-	-	-	-	I	-
149	I	-	-	-	-	L	-	-	-	-	-	-
182	S	-	-	-	-	-	-	-	-	-	T	-
188	I	-	-	-	-	-	-	-	-	-	-	T
200	N	S	S	S	S	S	-	-	-	S	-	S
216	I	-	-	-	V	-	-	-	-	-	-	-

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

234	V	-	-	-	-	-	-	-	-	I	-	-
241	V	I	I	I	I	I	-	-	I	I	I	I
248	N	D	D	D	D	D	D	-	D	D	-	D
257	R	-	K	-	-	-	-	-	-	-	K	-
264	V	I	I	I	I	I	-	-	-	-	-	-
270	N	K	K	K	K	K	-	-	-	-	-	-
273	N	-	-	-	-	-	-	-	-	-	K	-
275	H	-	Y	-	-	-	-	-	-	-	-	-
314	I	-	M	T	M	M	-	-	-	-	-	-
321	I	V	V	V	V	V	-	-	-	-	-	V
362	T	-	-	-	-	-	-	-	-	I	-	-
369	N	K	K	K	K	K	-	-	-	K	K	K
386	N	K	K	K	K	K	-	-	-	-	-	-
432	K	E	E	E	E	E	-	-	-	-	-	E
453	V	-	-	-	-	-	-	-		-	M	-
467	N	I	I	I	I	I	-	I	-	I	I	I

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 4 : Selection pressure analysis of HA protein (566 codons); NA protein (469 codons), NP Protein (498 codons) and M2 Protein (97 codons) of H1N1pdm virus using SLAC, FEL, MEME and FUBAR methods.

Protein	Codon	SLAC		FEL		MEME		FUBAR	
		dN-dS	p-value	dN-dS	p-value	ω^+	p-value	dN-dS	Post. Pr.
HA	16	5.313	0.273	10.659	0.121	>100	0.057	0.831	0.879
	180	5.375	0.276	11.199	0.129	>100	0.063	0.722	0.822
	207	2.659	0.522	4.035	0.327	>100	0.010	0.066	0.598
	239	**9.308**	**0.100**	**17.436**	**0.045**	**>100**	**0.063**	**2.395**	**0.987**

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	251	2.847	0.455	4.827	0.270	>100	0.046	0.096	0.620
	290	-0.813	0.813	-0.202	0.978	>100	0.003	-0.114	0.391
	436	4.757	0.987	11.187	0.829	>100	0.014	-0.563	0.594
NA	**13**	**3.688**	**0.200**	**10.586**	**0.043**	**>100**	**0.085**	**0.778**	**0.910**
	40	3.670	0.204	10.996	0.069	>100	0.097	0.710	0.890
	86	3.699	0.198	9.352	0.096	>100	0.098	0.622	0.881
	248	**4.401**	**0.168**	**12.606**	**0.059**	**>100**	**0.086**	**1.091**	**0.944**
	275	**5.632**	**0.172**	**18.916**	**0.055**	**>100**	**0.029**	**2.388**	**0.978**

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	386	**4.362**	**0.189**	**13.689**	**0.055**	**>100**	**0.094**	**1.270**	**0.953**
NP	180	0.353	0.667	0.351	0.974	>100	0.048	-0.091	0.427
	197	2.033	0.338	8.839	0.142	>100	0.002	0.353	0.790
	456	1.186	0.414	5.422	0.172	>100	0.197	0.101	0.603
	472	1.417	0.444	5.536	0.174	>100	0.199	0.126	0.668
	483	1.285	0.540	5.278	0.307	>100	0.002	0.085	0.624
	498	4.174	0.136	13.160	0.100	>100	0.135	1.079	0.934
M2	21	13.906	0.305	41.507	0.059	>100	0.187	1.116	0.937

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

27	9.357	0.444	24.689	0.089	>100	0.174	0.509	0.848
43	9.322	0.448	16.852	0.199	>100	0.278	0.173	0.749
89	9.322	0.448	27.353	0.121	>100	0.272	0.408	0.810

Note: Criteria to consider sites with significant evidence of positive selection: p-value <0.25 in SLAC, <0.1 in FEL and <0.25 in MEME, and posterior probability >0.9 in FUBAR. The sites selected significantly positive by all the four methods are written in bold font.

bioRxiv preprint doi: https://doi.org/10.1101/693358; this version posted July 5, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.