

Original Article

## Phylogenetic and nucleotide sequence analysis of influenza A (H1N1) HA and NA genes of strains isolated from Saudi Arabia

Ahmed A Al-Qahtani<sup>1,2</sup>, Muhammad Mubin<sup>3</sup>, Damian M Dela Cruz<sup>1</sup>, Sahar Isa Althawadi<sup>4</sup>, Muhammad Shah Nawaz ul Rehman<sup>3</sup>, Marie Fe F. Bohol<sup>1</sup>, Mohammed N Al-Ahdal<sup>1,2,4</sup>

<sup>1</sup> Department of Infection and Immunity, Research Center, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

<sup>2</sup> Department of Microbiology and Immunology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

<sup>3</sup> Virology Lab, Center of Agricultural Biochemistry and Biotechnology, University of Agriculture, Faisalabad, Pakistan

<sup>4</sup> Department of Pathology and Laboratory Medicine, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia

### Abstract

**Introduction:** In early 2009, a novel influenza A (H1N1) virus appeared in Mexico and rapidly disseminated worldwide. Little is known about the phylogeny and evolutionary dynamics of the H1N1 strain found in Saudi Arabia.

**Methodology:** Nucleotide sequencing and bioinformatics analyses were used to study molecular variation between the virus isolates.

**Results:** In this report, 72 hemagglutinin (HA) and 45 neuraminidase (NA) H1N1 virus gene sequences, isolated in 2009 from various regions of Saudi Arabia, were analyzed. Genetic characterization indicated that viruses from two different clades, 6 and 7, were circulating in the region, with clade 7, the most widely circulating H1N1 clade globally in 2009, being predominant. Sequence analysis of the HA and NA genes revealed a high degree of sequence identity with the corresponding genes from viruses circulating in the South East Asia region and with the A/California/7/2009 strain. New mutations in the HA gene of pandemic H1N1 (pH1N1) viruses, that could alter viral fitness, were identified. Relaxed-clock and Bayesian Skyline Plot analyses, based on the isolates used in this study and closely related globally representative strains, indicated marginally higher substitution rates than the type strain ( $5.14 \times 10^{-3}$  and  $4.18 \times 10^{-3}$  substitutions/nucleotide/year in the HA and NA genes, respectively).

**Conclusions:** The Saudi isolates were antigenically homogeneous and closely related to the prototype vaccine strain A/California/7/2009. The antigenic site of the HA gene had acquired novel mutations in some isolates, making continued monitoring of these viruses vital for the identification of potentially highly virulent and drug resistant variants.

**Key words:** H1N1; virus; molecular; analysis; Saudi Arabia.

*J Infect Dev Ctries* 2017; 11(1):81-88. doi:10.3855/jidc.9259

(Received 03 August 2016 – Accepted 19 September 2016)

Copyright © 2017 Al-Qahtani *et al.* This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Introduction

Influenza virus strains such as H1N1 and H5N1, are a major concern for both public health and the global economy, with outbreaks in recent decades resulting in major socio-economic damage because of lost productivity and the medical costs of infection. The increasing availability of sequence data and the development of new computational and statistical methods for analysis have contributed highly to a better understanding of the emergence, spread, and evolution of influenza viruses [1,2]. Influenza A viruses are members of the *Orthomyxoviridae* family, and have a segmented negative-sense single-stranded ribonucleic acid (ssRNA) genome that is capable of re-assortment, thus producing novel viruses [3,4]. The Influenza A

genome comprises eight segments of viral RNA, two of which encode surface glycoproteins, HA and NA, that are essential for infection. Infection is initiated by HA present on the outer surface of the virus, which recognizes sialic acid on the outer surface of a host cell and promotes receptor-mediated endocytosis [5]. On the other hand, NA removes the sialic acid residues from the surfaces of infected cells, allowing the virus progeny to be released from the cell and to infect other cells [5].

In April 2009, a novel influenza A virus, ‘pandemic H1N1’ (pH1N1), was identified in humans, initiating the first influenza pandemic of the 21st century. As of November 2<sup>nd</sup> 2009, there were nearly 400,000 laboratory-confirmed cases, resulting in over 4,500

deaths worldwide [6]. As most countries have stopped reporting cases, this is thought to represent a minimum of 2–5 million people infected. pH1N1 is a novel reassortant between the North American and Eurasian swine influenza viruses, containing polymerase basic 2 (PB2), polymerase basic 1 (PB1), polymerase acidic (PA), HA, nucleoprotein (NP), and non-structural (NS) segments from the North American triple-reassortant swine viruses, and NA and matrix (M) segments from the Eurasian swine lineage [7,8]. From July 19<sup>th</sup> 2009, the new virus spread across the world, reaching more than 140 countries [9]. The early viral diversification into seven discrete genetic clades [10] was confirmed by several subsequent studies [5,11]. Clade 7 rapidly became the most prevalent worldwide, but other variants continued to circulate, as H1N1pdm affected most countries through multiple introductions of different clade members [5,12-14]. This circulation of multiple variants can be explained by the air-borne transmission of influenza [9] and by heavy international air traffic and exchanges [13].

On June 2<sup>nd</sup> 2009, Saudi Arabia declared their first H1N1 case, in a woman who travelled to the country from the Philippines. Subsequently, close to fifteen thousand cases were identified, with more than 128 deaths. Significant mortality was observed in younger age groups (0 to 30 years), consistent with observations from other countries during this pandemic. The Influenza A (H1N1) viruses isolated from Saudi Arabia were found to be heterogeneous; however, though viral HA and NA genes belonged to various clades, most belonged to clade 7. The aim of this study was to establish the genetic relatedness, phylogeny, and mutation rate of viruses from Saudi Arabian isolates (Saudi isolates), compared to other isolates derived worldwide. Understanding the diversity and epidemiology of the virus is essential to devising strategies, both for the control of viral spread and in overcoming drug-resistance.

## Methodology

### *Patient samples and H1N1 virus detection*

Respiratory specimens (nasopharyngeal swabs (NP), bronchoalveolar lavages (BAL), or pharyngeal lavages (PL) were collected from 300 patients with suspected H1N1 virus infection, and tested for the presence of H1N1 viral RNA. The study was approved by the institutional review board of the King Faisal Specialist Hospital and Research Center, and conducted in accordance with the Helsinki Declaration of 1975. Informed consent was obtained from all adult patients and from the guardians of minor patients. RNA was

isolated using a MagNA Pure LC Total Nucleic Acid Isolation kit (Roche Diagnostics, Indianapolis, IN, USA) according to the manufacturer's instructions. H1N1 was detected using an artus Influenza LC real-time PCR (RT-PCR) Kit (Roche Diagnostics), and the reactions were performed in 96-well LightCycler 480 plates using the LightCycler 480 instrument (Bio-Rad, Hercules, USA).

### *RNA extraction, cDNA synthesis, and PCR*

Ten microliters of extracted RNA were mixed with 1 $\mu$ L of random hexamers (Invitrogen, Carlsbad, USA) (50 ng/ $\mu$ L) and heated at 65°C for 15 min. Subsequently, 5  $\mu$ L of 5X cDNA synthesis buffer (Invitrogen, Carlsbad, USA), 2.0  $\mu$ L of 10 mM dNTP mix, 1  $\mu$ L of 0.1M DTT, 1  $\mu$ L RNase inhibitor (Invitrogen, Carlsbad, USA) (40U/ $\mu$ L), 5 U of cloned avian myeloblastosis virus (AMV) reverse transcriptase, and DEPC water were added, giving a total volume of 25  $\mu$ L. The mixture was incubated for 1 h at 50°C before the reaction was terminated by heating at 85°C for 5 min, and the samples were stored at -70°C until use. PCR was performed in two rounds, in 50  $\mu$ L reaction volumes. The mixture in round 1 contained 1X PCR buffer [10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 5  $\mu$ L template cDNA, 0.5  $\mu$ M of each primer, 200  $\mu$ M of each deoxyribonucleotide, and 1 Unit of Taq polymerase (Applied Biosystems, Foster City, USA)]. PCR conditions were: 95°C for 7 min (initial denaturation), followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 1 min, before a final extension for 5 min at 72°C. HA and NA genes were amplified using gene specific primers. The PCR products were then separated on a 2% agarose gel, stained with 1  $\mu$ g/mL ethidium bromide, and visualized under ultraviolet illumination.

### *Sequence assembly, manipulation, and analysis*

PCR products were gel purified and sequenced using an automated DNA sequencing system (ABI 3100) and a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, USA) according to the manufacturer's instructions. Sequences were assembled and analyzed using the Lasergene v.8 software package (DNASTAR, Madison, USA), before being compared with publicly available sequences using the Basic Local Alignment Search Tool (BLAST) (NCBI). The coding regions and coding capacities of the HA/NA clones were predicted using the EditSeq module of the Lasergene package.

### Phylogenetic analysis and estimation of nucleotide substitution rates

Sequences were aligned using the MEGA (version 6) tool [15] (<http://www.megasoftware.net/>), applying the ClustalW alignment algorithm, and the resulting alignments were used in pairwise comparisons and the construction of phylogenetic trees using the Bayesian Evolutionary Analysis Sampling Trees (BEAST) software package version 1.7. For both datasets (HA and NA), NEXUS files were generated after aligning sequences in MEGA6, as described above. The general time reversible + empirical (GTR+E) substitution model was chosen for phylogenetic analysis, and the nucleotide dataset was partitioned into 3 subsets, corresponding to codon positions 1, 2, and 3. To set the tree prior, the ‘Coalescent: Bayesian Skyline’ settings were chosen in the tree panel of the BEAST package. Each dataset was run for a chain length of  $4 \times 10^7$ , ensuring an adequate sample size for the Markov chain Monte Carlo (MCMC) function, which was run using the BEAUTi module of the BEAST package [16].

### Antigenic Site Mapping of HA

The amino acid sequences of HA from Saudi Arabian influenza virus isolates were compared with those of vaccine strains [17]. The identified amino acid substitutions were mapped to reported HA antigenic sites [17-19]. Crystal structures for HA were downloaded from the Protein Data Bank (RCSB PDB, <http://www.pdb.org>) (HA structures 3LZG for H1, 1MQL for H3, and 2RFT for type B were used) [20].

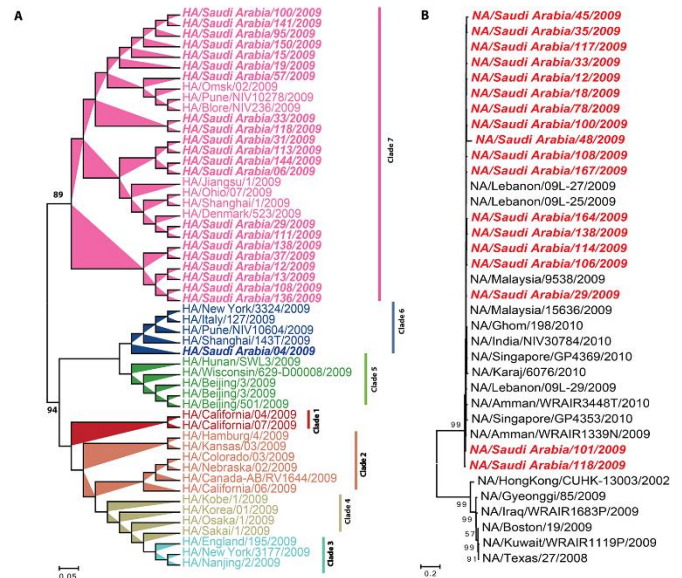
## Results

### Phylogenetic Analysis

The phylogeny of 72 HA and 45 NA genes was estimated using the BEAST software package. The phylogenetic tree for the HA gene was constructed using sequences from both the Saudi isolates and public databases. All but one of the HA genes from the Saudi isolates grouped into a cluster that was identified as clade 7 (Figure 1A). The discrepant isolate, HA/Saudi Arabia/04/2009, clustered with isolates from clade 6 (Figure 1A), and displayed the clade-specific amino acid substitutions K15E, Q293H, and S203T (Figure 2) [10]. The clade 7 isolates were characterized by an S203T substitution, along with other K15E and Q293H substitutions specific to clade 7 (Figure 2).

Phylogenetic analysis showed that the Saudi isolates did not appear to have diverged significantly from other regional viruses. The phylogenetic tree for the NA gene shows greater heterogeneity in the Saudi isolates, which cluster with isolates of influenza virus

**Figure 1:** Phylogenetic analysis of A) HA sequences and B) NA sequences. Neighbor-joining dendrogram based upon a ClustalW alignment of the HA and NA sequences was determined with selected publicly available HA and NA sequences. The database accession number for each sequence is shown. The HA and NA genes from the Saudi isolates are highlighted with bold italic text. The numbers at nodes represent the percentage bootstrap values (1000 replicates).



from Asia or the Middle East (Figure 1B), than does the tree for the HA gene.

### Molecular characterization of HA and NA genes

An analysis of clade-specific markers within the Saudi isolates revealed previously identified HA gene mutations characteristic of clade 7 [8,10], including the HA gene substitution S203T (Table 1). The majority of the Saudi isolates have an aspartate residue at position 222, within the receptor binding region of the HA molecule [21]. The HA D222G substitution is reportedly associated with severe disease [5,22,23], but was not found in the viruses analyzed here (Table 1 and Figure 2). However, the D222E substitution was identified in one isolate, HA/Saudi Arabia/138/2009 (Figure 2). Another mutation, E374K, either singly or with either D97N or V30A, is observed in the HA gene of isolates from several countries with increasing frequency [24]. Interestingly, two of the Saudi isolates, HA/Saudi Arabia/100/2009 and HA/Saudi Arabia/15/2009, carried the E374K mutation. None of the isolates carried the D97N substitution, but one isolate, HA/Saudi Arabia/04/2009, did possess the V30I substitution.

Twenty isolates had a histidine (H) residue at position 275 of the NA protein, this being associated with sensitivity to neuraminidase inhibitors, while 2



**Figure 2:** Antigenic structure of HA from the Saudi H1N1 virus and the 2009 H1N1 pandemic virus. Sequence alignment of membrane-distal domains from representative H1N1 HAs. Antigenic epitopes are color-coded as in [17-19]. Numerals show amino acid positions.

	70	124	137	153
A/Saudi Arabia/4/2009	CESLSTANSWSYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PYAGAKSFYKNLIWL	VKKGNS
A/Saudi Arabia/144/2009	CESLSTASSWPYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	LNHDSNKGVTAAC	PHAGAKSFYKNLIWL	VKKGNS
A/Saudi Arabia/138/2009	CESLSTASSWSYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PYAGAKSFYKNLIWL	VKKGNS
A/Saudi Arabia/89/2009	CESLSTASSWSYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PYAGAKSFYKNLIWL	VKKGNS
A/Saudi Arabia/142/2009	CESLSTASSWPYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PHAGAKSFYKNLIWL	VKKGNS
A/California/2009	CESLSTASSWSYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PHAGAKSFYKNLIWL	VKKGNS
A/Bahrain/N11661/2009	CESLSTASSWSYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PHAGAKSFYKNLIWL	VKKGNS
A/Karaj/5929/2010	CESLSTASSWSYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PHAGAKSFYKNLIWL	VKKGNS
A/Kuwait/N13093/2009	CESLSTASSWSYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PHAGAKSFYKNLIWL	VKKGNS
A/Iraq/WRAIR1683P/2009	CESLSTASSWSYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PHAGAKSFYKNLIWL	VKKGNS
A/Amman/WRAIR1339N/2009	CESLSTASSWSYIVETSSSDNGDFIDYEELREQLSSVSSFERFEIFPKTSSW	PNHDSNKGVTAAC	PHAGAKSFYKNLIWL	VKKGNS

	159	166	184	203	221	235				
A/Saudi Arabia/4/2009	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGSSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF
A/Saudi Arabia/144/2009	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGTSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	ESGDKITF
A/Saudi Arabia/138/2009	YPKLSKSY	TNDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGTSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF
A/Saudi Arabia/89/2009	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGTSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF
A/Saudi Arabia/142/2009	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGTSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF
A/California/2009	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGSSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF
A/Bahrain/N11661/2009	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGTSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF
A/Karaj/5929/2010	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGTSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF
A/Kuwait/N13093/2009	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGSSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF
A/Iraq/WRAIR1683P/2009	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGSSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF
A/Amman/WRAIR1339N/2009	YPKLSKSY	INDKGKEVLV	LWGIHHP	TSADQQSLYQNA	DAYV	VFGTSRYSK	FKFPEIAIRPKV	RDQEGRM	NYWTLV	EPGDKITF

isolates had a tyrosine (Y) residue at this position. Two other substitutions, V106I and N248D, were present in NA gene product of the majority of the viruses analyzed, with some isolates lacking the N248D substitution. The sequences of the Saudi isolates did not reveal the presence of mutations S143G, S185T, E374K, S451N, and V520A, which reportedly form a genetic signature of the predominant viruses from the 2010–2011 influenza season in the United Kingdom, Australia, Singapore, and New Zealand [25].

*Estimation of nucleotide substitution rates for HA and NA genes*

The mean nucleotide substitution rates for the HA and NA genes were determined using recombination

free datasets, with the Relaxed-Clock and Bayesian Skyline Plot (BSP) methods. For each dataset, the sequences were partitioned into the 3 codon positions (1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup>). The mean substitution rates for the HA and NA genes were  $5.14 \times 10^{-3}$  and  $4.18 \times 10^{-3}$  substitutions/nucleotide/year, respectively (Table 2). Similarly, the mean substitution rates for the HA and NA genes from the Saudi isolates, using the different codon start positions, were measured. The substitution rate at the third codon position of the HA gene (1.34) was higher than at codon positions 1 and 2 (0.725 and 0.928, respectively) (Table 2). However, the substitution rate in the NA gene was more surprising, with the mutation rate at codon position 2 more than double that at codon position 1 (CP1, 0.493 and CP2,

**Table 1.** Comparison of gene sequences of Saudi Arabia H1N1 viruses with regional and global H1N1 isolates at key amino acid positions in HA gene.

Gene position isolates	HA					NA*		
	83	203	222	293	321	106	248	275
HA/California/07/09	P	S	D	Q	I	V	N	Y
HA/Saudi Arabia/89/2009	S	T	D	Q	I	I	D	H
HA/Saudi Arabia/142/2009	S	T	D	Q	I	I	D	H
HA/Saudi Arabia/57/2009	S	T	D	Q	I	I	D	Y
HA/Hong Kong/ 2885/09	S	T	D	Q	V	I	D	H
HA/Bahrain/N11661/09	S	T	D	Q	V	I	D	H
HA/Amman/WRAIR1339N/09	S	T	D	Q	I	I	D	H
HA/Karaj/5924/10	S	T	D	Q	V	I	D	H
HA/Kuwait/N13093/09	S	S	D	Q	V	I	D	H
HA/Iraq/WRAIR1683P/09	P	S	D	Q	I	I	N	Y
HA/Sulaimani/05/09	S	T	D	Q	V	I	D	H
HA/Egypt/N14648/09	S	T	N	Q	I	I	D	H
HA/Ghom/169/10	S	T	D	Q	V	I	D	H

\* Amino acids shown in this table for NA protein are of Saudi isolates NA/Saudi Arabia/22/2009, NA/Saudi Arabia/31/2009, NA/Saudi Arabia/08/2009.

**Table 2.** Mean substitution rates of HA and NA proteins.

Model used	Gene	Number of sequences used	Mutation rate			Mean rate (substitution/nucleotide/year)
			Codon position 1	Codon position 2	Codon position 3	
Relaxed clock + BSP	HA	49	0.725	0.928	1.34	$5.14 \times 10^{-3}$
Relaxed clock + BSP	NA	39	0.493	1.036	1.473	$4.18 \times 10^{-3}$

BSP: Bayesian skyline plot.

1.036) (Table 2), and the mutation rate at codon position 3 similar to that of the HA gene (1.473).

#### *Predicted antigenic sites*

The antigenic sites of the HA gene can be identified using an amino acid alignment (Figure 2). Four antigenic sites were defined from the Saudi isolates, with two strain-specific sites (Sa and Sb) and two common antigenic sites (Ca and Cb) identified in the virus HA gene [17-19]. The antigenic pattern of the Saudi isolates was similar to that of the A/California/07/2009 strain [17-19], with some novel mutations. Mapping the substitutions to known HA antigenic sites revealed the presence of mutation P124L within the Sa antigenic motif of the HA/Saudi Arabia/144/2009 strain, mutation H138Y in all of the Saudi isolates except for HA/Saudi Arabia/138/2009, and mutation S203T within the Ca antigenic site of all of the Saudi isolates except for HA/Saudi Arabia/4/2009 (Figure 2). Similarly, the mutation I166T was observed within the Ca antigenic site of the HA/Saudi Arabia/138/2009 strain. These novel substitutions within the HA antigenic sites of the Saudi isolates could potentially affect HA antigenicity.

#### *Glycosylation sites of HA and NA molecules*

Changes to potential HA glycosylation sites in these virus isolates, as compared with the vaccine strain, were analyzed [26]. Eight potential glycosylation sites were reported in the HA molecule of the A/California/07/2009 virus, at positions 27, 28, 40, 104, 293, 304, 498, and 557. Conversely, the HA protein of pH1N1 viruses is known to possess eight glycosylation sites, at positions 10, 11, 23, 87, 276, 287, 481, and 540. No amino acid substitutions were observed at the above positions in the Saudi isolates, emphasizing the strength of conservation at these positions.

## **Discussion**

H1N1 viruses have diversified sufficiently to form 7 distinct clades, although the epidemiological behavior of these viruses is largely uniform, with certain risk

groups more prominently vulnerable than others [10]. These viruses were detected in Saudi Arabia as early as June 2009, being first imported through international travel (June to August 2009), before an increasing number of domestic infections was observed (September 2009 to March 2010).

The phylogenetic analysis presented here demonstrates that the Saudi isolates clearly belonged to clades 6 and 7, with the majority of the isolates clustering in clade 7. Amino acid sequence analysis of HA genes showed that the viruses belonging to clade 7 displayed the signature mutation S203T, and that no isolates belonged to clades 5 or 3 [5,8]. One Saudi isolate, HA/Saudi Arabia/4/2009, showed the characteristic clade 6 substitutions K15E and Q293H, along with S203T, indicating that two different clades were circulating in Saudi Arabia. The predominance of clade 7 isolates, usually prominent in Asia, in the current analysis, can be explained by the relatively late confirmation of the first Saudi Arabian cases, and the limited number of viruses that were sequenced in this study.

Epidemiological data confirm that the first H1N1 cases in Saudi Arabia were imported by foreign travelers; however, soon afterwards, indigenous virus evolution and transmission resulted in widespread infection. The relatively low number of reported and confirmed H1N1 cases in Saudi Arabia can be explained by the predominance of mild or sub-clinical cases, combined with a limited surveillance network for case detection.

Profiling of all of the tested Saudi isolates revealed that these viruses were antigenically similar to the A/California/7/2009 vaccine strain. The antigenic sites of the influenza virus HA gene are broadly distributed across four conformational epitopes [17-19]; the Sa and Sb sites are proximal to the receptor-binding pocket, the Ca site (Ca1 and Ca2) is located at the subunit interface, and the Cb site is found within the vestigial esterase domain. One Saudi isolate was found to contain a unique HA mutation, S206T, located within the receptor-binding domain. This domain is a major

determinant of Influenza A virus host specificity, and therefore, the serine to threonine mutation at position 206 may directly affect the infectivity and transmissibility of the virus in humans [27]. This suggests that appropriate vaccination strategies should be developed, using serological studies to identify groups that are particularly at risk.

Several novel substitutions were observed within HA antigenic sites, including the mutations P124L, observed within the Sa antigenic motif of the HA/Saudi Arabia/144/2009 isolate, H138Y, present in all of the Saudi isolates except for HA/Saudi Arabia/138/2009, and S203T, present in all of the Saudi isolates where serine was present at position 203, except for HA/Saudi Arabia/4/2009. Both S203T and the mutation I166T, observed in the HA/Saudi Arabia/138/2009 strain, were located within the Ca antigenic site. These novel substitutions, from 2007 Saudi isolates, could potentially impact upon HA antigenicity, although this needs to be confirmed experimentally. Alterations in glycosylation are used by Influenza and other viruses to interfere with surveillance by the host immune system. The acquisition of a glycosylation site masks the protein surface from antibody recognition, as the viral glycans are derived from the host and are thus considered “self” by the host immune system [28]. The Saudi isolates displayed no additional glycosylation sites that were not present in the type strain.

The HA and NA genes of the Saudi isolates showed a high degree of homology with the equivalent genes from H1N1 viruses, both from neighboring countries (mostly Middle Eastern), and the A/California/07/2009 strain (nucleotide identity ranged from 99–100%). Phylogenetic analysis showed that the isolates did not cluster separately from other viruses. There is therefore no evidence of gene re-assortment, between the pandemic strain and co-circulating seasonal Influenza H1N1 strains, during this time period [29].

Certain HA amino acid substitutions such as D222G are associated with severe disease and poor outcome [5,30,31]. As none of the Saudi Arabian strains analyzed here displayed this mutation, or induced a fatal outcome, further study, especially in more serious cases, is warranted. Most of the Saudi isolates retained isoleucine (I) at position 321 of HA, but very few had the I321V mutation seen in certain European viruses [32]. In contrast to the HA D222G mutation [22], the effect of retaining isoleucine at this position on disease severity has not been clearly demonstrated. Substitutions such as E374K and S451N, reported more recently in isolates from Iran, the

Netherlands, and India [5], were absent in the analyzed Saudi isolates.

Within the NA gene, the clade 7-specific substitutions V106I and N248D were seen in many of the analyzed viruses, including NA/Saudi Arabia/31/2009 and NA/Saudi Arabia/08/2009, while substitutions reportedly associated with drug resistance were absent. However, the NA genes from two Saudi isolates, NA/Saudi Arabia/101/2009 and NA/Saudi Arabia/118/2009, did not possess the N248D mutation characteristic of clade 7, indicating that these viruses could belong to either clades 1 or 3. Other substitutions such as the S95N and R257K mutations identified in isolates from Finland, were not present in the Saudi isolates.

Phylogenetic analysis shows that the Saudi isolates were interspersed with isolates of global and middle-eastern origin, which is probably indicative of multiple introductions of clade 7 viruses into the country. To place the Saudi isolates in a global perspective, the substitution rate of the HA genes was determined. The mean substitution rates for the HA and NA genes were  $5.14 \times 10^{-3}$  and  $4.18 \times 10^{-3}$  substitutions/nucleotide/year, respectively. This is slightly higher than a previously reported estimated substitution rate for the influenza virus ( $3.92 \times 10^{-3}$  and  $3.61 \times 10^{-3}$  substitutions/nucleotide/year for the HA and NA genes, respectively) [33,34]. The higher nucleotide substitution rate indicates a more rapid virus evolution in Saudi Arabia, most likely because of the diversity and high number of travelers to the country. The mean substitution rate at different codon positions was measured for both genes using the Saudi isolates. The mean substitution rate at the third codon position (1.34) was higher than that at positions 1 and 2 (0.725 and 0.928 respectively). This higher mutation rate at the third codon position of the HA gene is plausible, because of its wobble position in the genetic code. The mean substitution rate in the NA gene was, however, more surprising, with the mutation rate at position 2 more than double that of position 1 (CP1, 0.493 and CP2, 1.036). Additionally, the mutation rate at codon position 3 (1.473) was similar to that of the HA gene. This indicates that there is higher selective pressure on the NA gene, as any change at codon position 2 will result in a change in protein sequence. The present study suggests that global evolution of the H1N1 virus has led to higher substitution rates. Our analysis indicates that the predominant strains circulating in Saudi Arabia are very similar to the type strain A/California/07/2009.



The study thus emphasizes the need for continued surveillance and genetic characterization of whole viral genome sequences, in order to detect re-assortment events that could enhance H1N1 virus fitness. One of the impediments to understanding the severity of the Saudi Arabian Influenza epidemic is the paucity of available patient data. The viruses analyzed in this study were derived from a small proportion of the total infected patient population. These viruses were randomly selected for sequencing, and isolates from cases with adverse or fatal outcomes are currently being analyzed in great detail for mutations associated with high virulence. In the months following the end of pandemic, no new re-assortment events or mutations have been observed in Influenza viruses. As more epidemiological and sequencing data on the Saudi Influenza viruses becomes available, a better understanding of their continuing evolution will be achieved, particularly the occurrence of any re-assortment events with local seasonal Influenza viruses. Continued surveillance of H1N1 viruses will ensure the early detection of new antigenic or drug resistant variants. This will facilitate better pandemic management planning and response capacity, at both national and global levels.

## Conclusions

In conclusion, Saudi Arabian H1N1 virus isolates were antigenically homogeneous, and closely related to the prototype vaccine strain A/California/7/2009. These isolates showed no amino acid changes in the glycosylation sites of HA that would indicate high pathogenicity and virulence. However, in some isolates, some antigenic sites of the HA gene bore novel mutations whose effects require further investigation. The continuous monitoring of these viruses to identify variants that are potentially highly virulent or drug resistant is essential.

## Acknowledgements

This study was funded by King Faisal Specialist Hospital and Research Centre, Project # 2100009 and approved by the Ethics Committee of the hospital. We are grateful to Hanan Shaarawi and Maureen Delos Reyes for secretarial and logistic assistance.

## Authors' contribution

A.A.A, M.M, D.M.D.C, M.F.F.B, M.S.N.R and S.I.A performed experiments and prepared the manuscript. M.N.A.A provided overall directions regarding the designing of all experiments, writing and supervised the work. The final manuscript was read and approved by all authors.

## References

- Cattoli G, Monne I, Fusaro A, Joannis TM, Lombin LH (2009) Highly pathogenic avian influenza virus subtype H5N1 in Africa: a comprehensive phylogenetic analysis and molecular characterization of isolates. *PLoS ONE* 4: e4842.
- Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *PLoS Comput Biol* 5: e1000520.
- Horimoto T, Kawaoka Y (2005) Influenza: lessons from past pandemics, warnings from current incidents. *Nat Rev Microbiol* 3: 591-600.
- Palese P, Shaw ML (2001) Orthomyxoviridae: The Viruses and Their Replication. In Knipe D.M, editors. *Fields Virology* 4<sup>th</sup> Edition. Philadelphia (USA): Lippincott Williams and Wilkins 1487-1531.
- Potdar VA, Chadha MS, Jadhav SM, Mullick J, Cherian SS (2010) Genetic characterization of the influenza A pandemic (H1N1) 2009 virus isolates from India. *PLoS One* 5: e9693.
- World Health Organization (2009) Emergencies preparedness, response. Pandemic (H1N1) 2009-update 72. Available: [http://www.who.int/csr/don/2009\\_10\\_30/en/](http://www.who.int/csr/don/2009_10_30/en/). Accessed 18 July 2016.
- Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S (2009) Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med* 360: 2605–2615.
- Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S (2009) Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* 325: 197–201.
- Balcan D, Hu H, Goncalves B, Bajardi P, Poletto C (2009) Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Med* 7: 45.
- Nelson M, Spiro D, Wentworth D, Fan J, Beck E, St. George K, Ghedin E, Halpin R, Bera J, Hine E, Proudfoot K, Stockwell T, Lin X, Griesemer S, Bose M, Jurgens L, Kumar S, Viboud C, Holmes E, Henrickson K (2009) The early diversification of influenza A/H1N1pdm. *PLoS Curr Influenza* RRN1126.
- Chan KH, To KK, Hung IF, Zhang AJ, Chan JF (2011) Differences in antibody responses of individuals with natural infection and those vaccinated against pandemic H1N1 2009 influenza. *Clin Vaccine Immunol* 18: 867–873.
- Furuse Y, Suzuki A, Oshitani H (2010) Evolutionary analyses on the HA gene of pandemic H1N1/09: early findings. *Bioinformation* 5: 7-10.
- Parks D, Macdonald N, Beiko R (2009) Tracking the evolution and geographic spread of Influenza A. *PLoS currents* 1: RRN1014.
- Shiino T, Okabe N, Yasui Y, Sunagawa T, Ujike M (2010) Molecular evolutionary analysis of the influenza A(H1N1)pdm, May-September, 2009: temporal and spatial spreading profile of the viruses in Japan. *PLoS One* 5: e11057.
- Tamura K, Stecher G, Peterson D, Filipksi A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30: 2725-2729.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.
- Xu R, Ekiert DC, Krause JC, Hai R, Crowe JE (2010) Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science* 328: 357-360.
- Wiley DC, Wilson IA, Skehel JJ (1981) Structural identification of the antibody-binding sites of Hong Kong

- influenza haemagglutinin and their involvement in antigenic variation. *Nature* 289: 373-378.
19. Berton MT, Naeve CW, Webster RG (1984) Antigenic structure of the influenza B virus hemagglutinin: nucleotide sequence analysis of antigenic variants selected with monoclonal antibodies. *J Virol* 52: 919-927.
  20. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN (2000) The protein data bank. *Nucleic Acids Res* 28: 235-242.
  21. World Health Organization (WHO) (2010) Preliminary review of D222G amino acid substitution in the haemagglutinin of pandemic influenza A (H1N1) 2009 viruses. *Wkly. Epidemiol. Rec* 4: 21–28.
  22. Kilander A, Rykkvin R, Dudman SG, Hungnes O (2010) Observed association between the HA1 mutation D222G in the 2009 pandemic influenza A(H1N1) virus and severe clinical outcome, Norway 2009–2010. *Euro Surveill* 15: 19498.
  23. Scalera NM, Mossad SB (2009) The first pandemic of the 21st century: a review of the 2009 pandemic variant influenza A (H1N1) virus. *Post grad Med* 121: 43-47.
  24. Maurer-Stroh S, Lee RTC, Eisenhaber F, Cui L, Phuah SP, Lin RTP (2010) A new common mutation in the hemagglutinin of the 2009 (H1N1) influenza A virus. *PLoS Curr.* 2, RRN1162.
  25. Barr IG, Cui L, Komadina N, Lee RT, Lin RT (2010) A new pandemic influenza A(H1N1) genetic variant predominated in the winter 2010 influenza season in Australia, New Zealand and Singapore. *Euro Surveill* 15: 19692.
  26. Centers for Disease Control and Prevention (2009) Swine influenza A (H1N1) infection in two children-Southern California, March-April 2009. *MMWR Morb Mortal Wkly Rep* 58: 400-402. Available: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5815a5.htm>. Accessed 30 November 2016.
  27. Pan C, Cheung B, Tan S, Li C, Li L, Liu S (2010) Genomic signature and mutation trend analysis of pandemic (H1N1) 2009 influenza A virus. *PLoS One* 5: e9549.
  28. Schulze IT (1997) Effects of glycosylation on the properties and functions of influenza virus hemagglutinin. *J Infect Dis* 176: 24-28.
  29. Perez DR, Sorrell E, Angel M, Jianqiang Y, Hickman D (2009) Fitness of pandemic H1N1 and seasonal influenza A viruses during co-infection: evidence of competitive advantage of pandemic H1N1 influenza versus seasonal influenza *PLoS Current* 1: RRN1011.
  30. Glinisky GV (2010) Genomic analysis of pandemic (H1N1) 2009 reveals association of increasing disease severity with emergence of novel hemagglutinin mutations. *Cell Cycle* 9: 958-970.
  31. Jamieson DJ, Honein MA, Rasmussen SA, Williams JL, Swerdlow DL (2009) H1N1 2009 influenza virus infection during pregnancy in the USA. *Lancet* 374: 451-458.
  32. Ikonen N, Haanpa AM, Ro NE, Lyytika IO, Kuusi M (2010) Genetic of the 2009 pandemic influenza A(H1N1) viruses in Finland. *PLoS ONE* 5: e13329.
  33. Gavin J, Smith G, Vijaykrishna D, Bahl J, Lycett S, Worobey M, Pybus O, Ma S, Cheung C, Raghvani J, Bhatt S, Malik Peiris J, Guan Y, Rambaut A (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459: 1122-1125.
  34. Zaraket H, Saito R, Sato I, Suzuki Y, Li D, Dapat C, Caperig-Dapat I, Oguma T, Sasaki A, Suzuki H (2009) Molecular evolution of human influenza A viruses in a local area during eight influenza epidemics from 2000 to 2007. *Arch Virol* 154: 285-295.

#### Corresponding author

Mohammed N. Al-Ahdal, PhD  
 Department of Infection and Immunity,  
 Research Center, King Faisal Specialist Hospital & Research Center,  
 PO Box 3354 (MBC-03)  
 Riyadh, 11211, Saudi Arabia  
 Phone: +966 011 4427867  
 Fax: +966 011 4424519  
 Email: ahdal@kfshrc.edu.sa

**Conflict of interests:** No conflict of interests is declared.