

## Sequence analysis of the 2009 pandemic influenza A H1N1 virus haemagglutinin gene from 2009-2010 Brazilian clinical samples

João Leandro de Paula Ferreira<sup>+</sup>, Samanta Etel Treiger Borborema, Luis Fernando de Macedo Brígido, Maria Isabel de Oliveira, Terezinha Maria de Paiva, Cecília Luiza Simões dos Santos

Centro de Virologia, Instituto Adolfo Lutz, Av. Dr. Arnaldo 355, 01246-000 São Paulo, SP, Brasil

*In this paper, we analysed the haemagglutinin (HA) gene identified by polymerase chain reaction from 90 influenza A H1N1 virus strains that circulated in Brazil from April 2009-June 2010. A World Health Organization sequencing protocol allowed us to identify amino acid mutations in the HA protein at positions S220T (71%), D239G/N/S (20%), Y247H (4.5%), E252K (3.3%), M274V (2.2%), Q310H (26.7%) and E391K (12%). A fatal outcome was associated with the D239G mutation ( $p < 0.0001$ ). Brazilian HA genetic diversity, in comparison to a reference strain from California, highlights the role of influenza virus surveillance for study of viral evolution, in addition to monitoring the spread of the virus worldwide.*

Key words: influenza - H1N1 - haemagglutinin - polymorphism - Brazil

Influenza is a highly contagious acute respiratory disease. It is responsible for both local and regional outbreaks every year and even worldwide epidemics. Although it causes only mild symptoms in the majority of cases, it can also be associated with severe disease and even death. Influenza viruses are known for their high evolutionary rate. Some mutations can result in amino acid substitutions that occur at key locations, such as antigenic or receptor-binding sites of the haemagglutinin (HA) gene (WHO 2010). Mutations in the HA gene of pandemic influenza A H1N1 viruses have already been described, including the D239G substitution in virus isolated from severe and fatal cases in several countries (Kilander et al. 2010, Melidou et al. 2010).

The Adolfo Lutz Institute has cooperated with the World Health Organization (WHO) since the 1950s for influenza surveillance. In early May 2009, laboratory-confirmed Brazilian cases of pandemic influenza A H1N1 virus were identified at the Institute using real-time reverse transcription-polymerase chain reaction (RT-PCR) based on a protocol from the WHO. Epidemiological analysis of the notified Brazilian cases reported that the case-fatality rate was 11.2% during the first wave of the pandemic period (Oliveira et al. 2009).

To evaluate the amino acid diversity at the HA gene of the influenza A H1N1 virus that circulated in Brazil from April 2009-June 2010, respiratory samples (nasal and/or pharyngeal swabs, nasopharyngeal swabs or aspirates) and necropsy tissues were collected from sus-

pected cases of pandemic influenza A H1N1 virus infection. We analysed 90 confirmed cases, 28 of which were fatal. Clinical samples from five states were examined in this study: Goiás (n = 7), Mato Grosso (n = 1), Mato Grosso do Sul (n = 1), Piauí (n = 4), São Paulo (SP) (n = 76) and Distrito Federal (n = 1).

Viral RNA was extracted using the QIAmp Viral RNA Mini Kit (Qiagen, CA, USA). The HA gene was amplified by RT-PCR and sequenced using a protocol from the WHO ([who.int/csr/resources/publications/swineflu/sequencing\\_primers/en/index.html](http://who.int/csr/resources/publications/swineflu/sequencing_primers/en/index.html)). RT-PCR products were directly sequenced using the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction (PE Applied Biosystems, CA, USA) with an ABI 3130 Genetic Analyser. All nucleotide sequences obtained were edited and aligned using Sequencher 4.7 (Gene Codes Corporation, MI, USA) and Bioedit 7.0, respectively. Phylogenetic analysis was performed with PAUP 4.0b10 (Swofford 1999) using an evolutionary model selected by Modeltest 3.7 (Posada & Crandall 1998). Statistical support for tree branches was estimated by bootstrapping (1,000 replicates) and the likelihood ratio test (0-length). Pairwise genetic distance was calculated using MEGA 4.0.1 and trees were visualised using TreeView (Page 1996). The nucleotide sequences obtained in this study were submitted to the National Centre of Biotechnology Information. The sequences are available at GenBank under accessions GQ247724, GQ356787, GQ368664-GQ368665, GQ368767, GQ414764-GQ414768, GQ915017-GQ915025 and HQ595232-HQ595299.

Molecular analyses of the 90 laboratory-confirmed 2009 pandemic influenza A H1N1 virus samples revealed a number of HA sequence variations in Brazilian circulating strains. Amino acid polymorphisms (in comparison with the A/California/7/2009 vaccine strain) are depicted in Table. The substitution of aspartic acid to glycine at position 239 (D239G) showed a significant association with fatality. This substitution was observed in 46% (13 of 28) of fatal cases analysed, whereas it was not

Financial support: IAL/SES/SP, MS, CDC

The authors represent the Work Group of Influenza of the Adolfo Lutz Institute.

+ Corresponding author: [mariliadejoao@yahoo.com.br](mailto:mariliadejoao@yahoo.com.br)

Received 8 December 2010

Accepted 6 June 2011

TABLE

Percentages of amino acid polymorphisms at the haemagglutinin gene observed in fatal and non-fatal cases (n = 90) of reverse transcription-polymerase chain reaction confirmed pandemic influenza A H1N1 virus 2009

Mutation position <sup>a</sup>	Non-fatal cases	Fatal cases	p value	Total % (n)
	(n = 62) % (n)	(n = 28) % (n)		
S220T	67.7 (42)	78.6 (22)	0.430	71.0 (64)
D239G	0 (0)	25 (7)	0.002 <sup>b</sup>	7.8 (7)
D239N	3.2 (2)	10.7 (3)	0.170	5.6 (5)
D239G/N/S	0 (0)	21.4 (6)	0.006 <sup>b</sup>	6.7 (6)
Y247H	4.5 (4)	0 (0)	0.350	3.6 (4)
E252K	4.8 (3)	0 (0)	0.550	3.3 (3)
M274V	3.2 (2)	0 (0)	1.000	2.2 (2)
Q310H	29 (18)	21.4 (6)	0.620	26.7 (24)
E391K	14.5 (9)	7.1 (2)	0.500	12 (11)

a: amino acid position was counted from the start codon of the vaccine strain of pandemic influenza A H1N1 virus 2009 (subtype H1) A/California/7/2009; b:  $p < 0.05$ . GenBank accession: FJ966974.

identified in non-fatal cases ( $p < 0.0001$ , Fisher's exact test, 2-tailed). The substitution to asparagine (D239N) was observed both in non-fatal cases (3%) and fatal ones (11%). A total of 20% (18 of 90) of the samples showed amino acid changes at this position (D239G/N/S).

Another frequent polymorphism in the HA amino acid sequence identified was Q310H, which was present in 27% of cases (24/90). However, we observed no association of this mutation with a fatal outcome (21.4% - 6/28) when compared to recovered cases (29% - 8/62,  $p = 0.6$ ). The majority of the samples (64/90) contained the S220T substitution, but this substitution showed no correlation with clinical outcomes. Moreover, HA sequences from two non-fatal cases, A/Piauí/21402/2010 and A/São Paulo/22205/2010, presented additional amino acid substitutions at positions 247 (Y-H), 252 (E-K) and 274 (M-V). The polymorphisms were observed in samples from different areas of the country, including those from SP, where most of the samples were collected. All sequences were obtained from clinical isolates in order to avoid *in vitro* evolution of HA gene.

The most striking observation from this study was the association of the D239G substitution in the HA nucleotide sequence with fatal cases. This D239G mutation has also been detected in several other countries. These data had been previously summarised in a WHO review, which reported that the overall prevalence of the D239G mutation was  $< 1.8\%$  in contrast to a rate of 7.1% in fatal cases (WHO 2010). This specific mutation in the viral HA has been associated with both severe and fatal cases. In Norway, Kilander et al. (2010) have previously reported that 11 of 61 (18%) patients with severe and fatal disease had the D239G mutation, which was not observed in samples from 205 patients with mild dis-

ease. Furthermore, Miller et al. (2010), from the United Kingdom, have shown that virus samples isolated from two of 23 patients (8.7%) with fatal disease exhibited the D239G mutation and that no mutations were observed among samples from 35 patients with severe and mild disease. Mak et al. (2010) have observed this mutation in nine of 219 (4.1%) severe and fatal cases in Hong Kong and no mutation among the 239 non-severe cases.

In contrast to results published by Kilander et al. (2010) and Mak et al. (2010), our findings from the phylogenetic analysis and the zero length test indicated that the fatal cases with the D239G mutation clustered together. This cluster was not preserved if the mutated codon was removed (data not shown). Phylogenetic groupings may be observed for other amino acid substitutions, as shown in Fig. 1.

Kilander et al. (2010) have studied paired HA amino acid sequences and did not find discordant isolates when lung material was compared to isolates from the upper respiratory tract. In our study, we did not perform paired analysis to further characterise the association of mutations at position 239 with a fatal outcome. In addition, because we only studied lung tissue, we cannot rule out pneumotropism as a decisive factor in the increased percentage of deaths in all fatal cases. Nonetheless, if these isolates behave as those studied by Kilander et al. (2010), we can assume that a similar proportion would be observed if samples were obtained from the upper respiratory tract. Potential bias due to compartmentalisation should, however, be taken into account when evaluating results from this study and others. It should be noted that the majority of severe and fatal cases analysed did not carry the D239G substitution and, clearly, this mutation is not required for a fatal outcome. It would be important to better contextualise these findings according to the clinical setting (e.g., patient's age, risk factors, disease presentation and progression) to better understand the role, if any, of this mutation on the patient's clinical outcome.

In our study of the six fatal cases resulting from influenza A H1N1 infection, we found the presence of the mixtures RRT nucleotides (R = A or G) in one case and RAT in five cases at position 239 of the HA amino acid sequence. These polymorphisms lead us to deduce the presence of different possible amino acids (D239G/N/S), suggesting that these patients had a mixed population of virus. These patients were not included in this study. The co-existence of mutant and wild-type viruses was verified by Kilander et al. (2010) and Miller et al. (2010). Although other amino acid changes in codon 239 were observed in fatal cases, the D239N (aspartic acid to asparagine) mutation was also present in non-fatal cases. The presence of D239E (aspartic acid to glutamic acid) was not observed in this study.

In contrast to results from Glinsky (2010), our analysis does not indicate any association between mutation Q310H and fatal cases. It is true that non-fatal cases have a higher proportion of this mutation; however, the difference is not statistically significant ( $p = 0.6$ ). These findings corroborate those recently reported by Lee et al. (2010), which showed the location of the Q310H mutation to be far from the receptor-binding pocket, with no

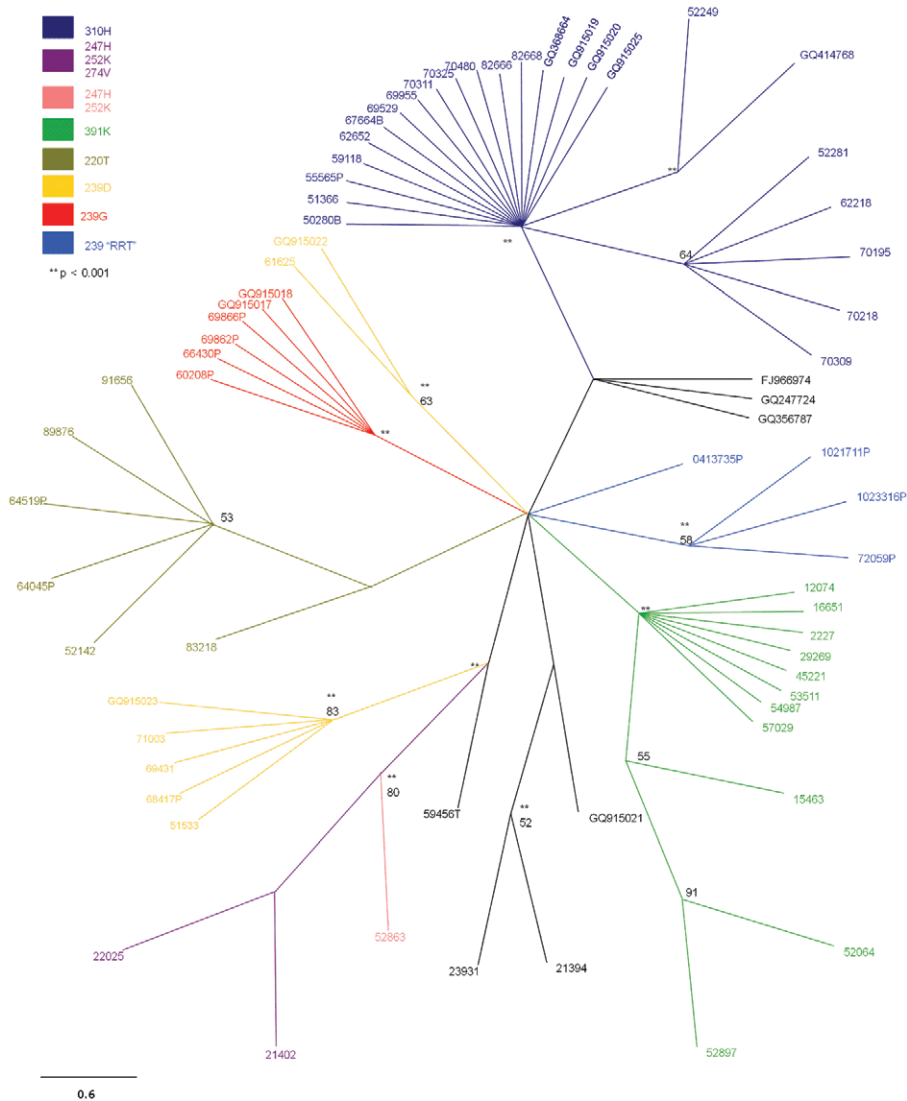


Fig. 1: maximum-likelihood phylogenetic tree of nucleotide sequences of the haemagglutinin gene of pandemic influenza A H1N1 virus 2009. A representative set from available sequences from Brazil and references from California were aligned. Tree was constructed utilizing PAUP 4.0b10, using HKY+G model, as determined by Modeltest 3.7. The robustness of tree nodes was assessed with bootstrap (1,000 replicates), shown as percentages, and zero-length test, shown as \*\* for  $p < 0.001$ . The scale bar indicated the number of nucleotide substitutions per site.

previously known direct biomolecular mechanism that could support an association between a mutation in this position and change in severity.

In the current study, a prominent new E391K mutation was also identified in the HA protein. Maurer-Stroh et al. (2010) modelled this mutation in the context of an available HA crystal structure and concluded that it could alter the salt bridge pattern and stability in a region of the HA oligomerisation interface. This interface is important for membrane fusion and it is known to be an antigenic site. However, more clinical and experimental data are needed to determine whether this mutation changes the biology and fitness of the virus.

A maximum-likelihood phylogenetic tree to examine relationships between HA amino acid sequence

from different samples was constructed using the evolutionary model HKY+G (Fig. 1). Phylogenetic analysis showed clusters with zero-length and/or bootstrap support. Sequences with the D239G mutation clustered together with significant zero-length ( $p < 0.001$ ). Brazilian samples from different time periods showed a tendency to increase the genetic distance when compared to the reference strain (Fig. 2). These results are in agreement with the findings reported by Souza et al. (2010), suggesting continuous viral evolution.

The declaration of the end of pandemic period by WHO is reassuring, but these molecular characteristics should be considered in future strategies to understand viral pathogenesis. Surveillance is pivotal to monitor the evolution of influenza virus worldwide.

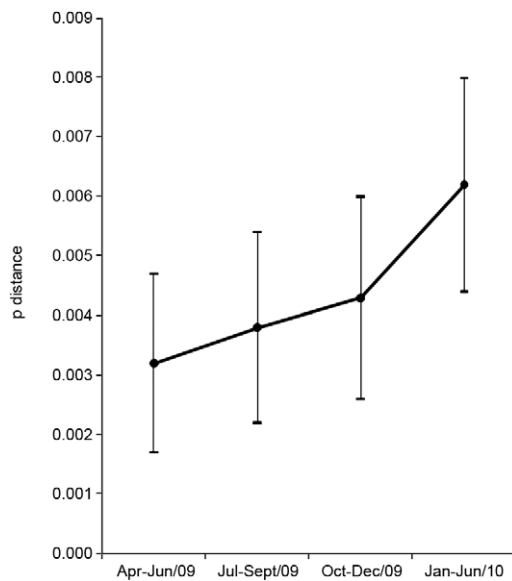


Fig. 2: genetic distances (mean and standard error) according to period of collection of the studied sequence in relation to a reference California strain (accession FJ966974).

#### ACKNOWLEDGEMENTS

To André Minhoto Lança and João Paulo Gervásio Batista, for their technical assistance in bioinformatic analyses.

#### REFERENCES

- Glinsky GV 2010. Genomic analysis of pandemic (H1N1) 2009 reveals association of increasing disease severity with emergence of novel hemagglutinin mutation 2010. *Cell Cycle* 9: 958-970.
- Kilander A, Rykkvin R, Dudman S, 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: pii=19498.
- Lee RTC, Santos CLS, Paiva TM de, Cui L, Sirota FL, Eisenhaber F, Maurer-Stroh S 2010. All that glitters is not gold - founder

effects complicate associations of flu mutations to disease severity. *Virol J* 7: 297.

- Mak GC, Au KW, Tai LS, Chuang KC, Cheng KC, Shiu TC, Lim W 2010. Association of D222G substitution in the haemagglutinin of 2009 pandemic influenza A (H1N1) with severe disease. *Euro Surveill* 15 (14): pii=19534.
- Maurer-Stroh S, Lee RTC, Eisenhaber F, Cui L, Phuap SP, Lin RTP 2010. A new common mutation in the hemagglutinin of the 2009 (H1N1) influenza A virus. *PLoS Curr* 2: RRN1162.
- Melidou A, Gioula G, Exindari M, Chatzidimitriou D, Diza E, Malisiovas N 2010. Molecular and phylogenetic analysis of haemagglutinin gene of pandemic influenza H1N1 2009 viruses associated with severe and fatal outcomes. *Virus Res* 151: 192-199.
- Miller RS, MacLean AR, Gunson RN, Carman WF 2010. Occurrence of haemagglutinin mutation D222G in pandemic influenza A (H1N1) infected patients in the West of Scotland, United Kingdom, 2009-10. *Euro Surveill* 15: pii=19546.
- Oliveira W, Carmo E, Penna G, Kuchenbecker R, Santos H, Araujo W, Malaguti R, Duncan B, Schmidt M, Surveillance Team for the pandemic influenza A (H1N1) 2009 in the Ministry of Health 2009. Pandemic H1N1 influenza in Brazil: analysis of the first 34,506 notified cases of influenza-like illness with severe acute respiratory infection (SARI). *Euro Surveill* 14: pii=19382.
- Page RDM 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12: 357-358.
- Posada D, Crandall KA 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- Souza TM, Salluh JI, Bozza FA, Mesquita M, Soares M, Motta FC, Pitrowsky MT, de Lourdes Oliveira M, Mishin VP, Gubareva LV, Whitney A, Rocco SA, Gonçalves VM, Marques VP, Velasco E, Siqueira MM 2010. H1N1pdm influenza infection in hospitalized cancer patients: clinical evolution and viral analysis. *PLoS ONE* 5: e14158.
- Swofford D 1999. PAUP 4.0: Phylogenetic analysis with parsimony (and other methods), version 4.0b2a. Sunderland, Mass: Sinauer Associates Inc.
- WHO - World Health Organization 2010. Preliminary review of D222G amino acid substitution in the hemagglutinin of pandemic influenza A (H1N1) 2009 viruses. *Wkly Epidemiol Rec* 85: 21-28.