

Origin of the 2009 Mexico influenza virus: a comparative phylogenetic analysis of the principal external antigens and matrix protein

Muhammed Babakir-Mina · Salvatore Dimonte ·
Carlo Federico Perno · Marco Ciotti

Received: 1 June 2009 / Accepted: 22 June 2009 / Published online: 7 July 2009
© Springer-Verlag 2009

Abstract Triple-reassortant swine influenza A (H1) viruses, containing genes from avian, human, and swine influenza viruses, emerged and became an outbreak among humans worldwide. Over a 1,000 cases were identified within the first month, chiefly in Mexico and the United States. Here, the phylogenetic analysis of haemagglutinin (HA), neuraminidase (NA), and matrix protein (MP) was carried out. The analysis showed that the H1 of this reassortant originated from American pigs, while NA and MP were more likely from European pigs. All of the 2009 isolates appear homogeneous and cluster together, although they are distinct from classical human A (H1N1) viruses.

Swine have been hypothesized to act as a mixing vessel for the re-assortment of avian, swine, and human influenza viruses and might play an important role in the emergence of novel influenza viruses capable of causing a human pandemic [1, 2, 8]. Recent reports of widespread transmission of swine-origin influenza A (H1N1) viruses among humans in Mexico, the United States, Europe, and elsewhere highlight this ever-present threat to global public health [1, 2]. On 9 April 2009, it became apparent to public

health officials in Mexico City that an outbreak of influenza was in progress late in the influenza season [3]. On 17 April, two cases in children were also reported in California near the Mexican border [1]. Virus samples were obtained, and a novel strain of influenza A of the H1N1 serotype was identified. Preliminary tests conducted by the Center for Disease Control and Prevention (CDC) indicated that the virus was a novel re-assortant, containing genetic elements of influenza viruses found in swine, birds and humans. This report analyzes the possible origin of this novel virus by phylogenetic analysis of the principal external antigens and matrix protein of influenza H1N1 2009 and other H1N1 isolates identified previously in humans and swine.

In this paper, 141 sequences of hemagglutinin (HA), 123 of neuraminidase (NA), and 96 of matrix protein (MP) from human and swine H1N1 viral isolates were randomly downloaded from GenBank and analysed by phylogenetic analysis. The sequences from 1999 to 2009 were downloaded from the NCBI influenza virus resource (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>). GenBank accession numbers are available in Supplementary Table S1. Three different data sets were built for HA, NA, and MP gene segments. Multiple sequence alignments of HA, NA, and MP segments were performed by using CLUSTAL X [11] and manually edited with the Bioedit software [6]. Positions containing gaps were removed from the final alignment. For our data set, the best-fitting nucleotide substitution model was tested with a hierarchical likelihood ratio test following the strategy described by Swofford and Sullivan [10], using a neighbor-joining (NJ)-based tree with LogDet-corrected distances. Maximum-likelihood (ML) trees were then inferred with the selected model and ML-estimated substitution parameters. The heuristic search for the ML

Electronic supplementary material The online version of this article (doi:10.1007/s00705-009-0438-1) contains supplementary material, which is available to authorized users.

M. Babakir-Mina · C. F. Perno · M. Ciotti (✉)
Laboratory of Molecular Virology,
University Hospital Tor Vergata,
Viale Oxford, 81-00133 Rome, Italy
e-mail: marco.ciotti@ptvonline.it

S. Dimonte · C. F. Perno
Virology, Department of Experimental Medicine
and Biochemistry, University of Rome Tor Vergata, Rome, Italy

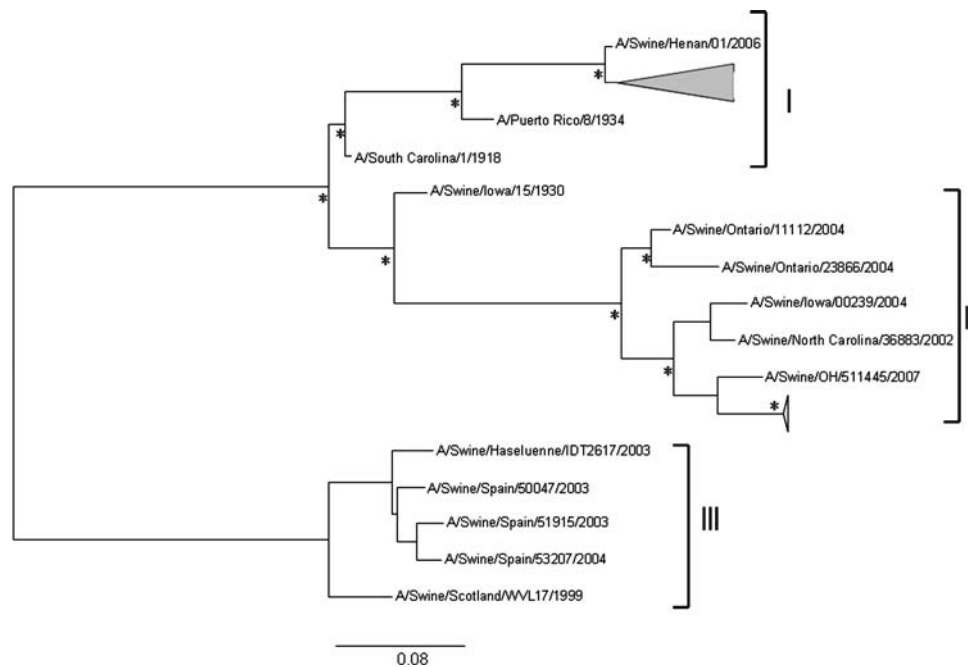


Fig. 1 Phylogenetic analysis of the haemagglutinin (HA) gene (1,701 nt) of the H1N1 influenza A viruses was carried out using HKY + I + C (alpha-parameter = 0.0763) as the best evolutionary model. Branch lengths were estimated with the best-fitting nucleotide substitution model according to a hierarchical likelihood ratio test [7] and were drawn to scale, with the bar at the bottom indicating 0.08 nucleotide substitutions per site. One asterisk along a branch

represents significant statistical support for the clade subtending that branch ($P < 0.001$ in the zero-branch-length test) and bootstrap support $>95\%$. The tree was unrooted. Three main clades are indicated (I, II, and III). All H1N1 viruses isolated from 2000 to 2008 are presented as a gray triangle, and all novel 2009 H1N1 viruses as a white triangle

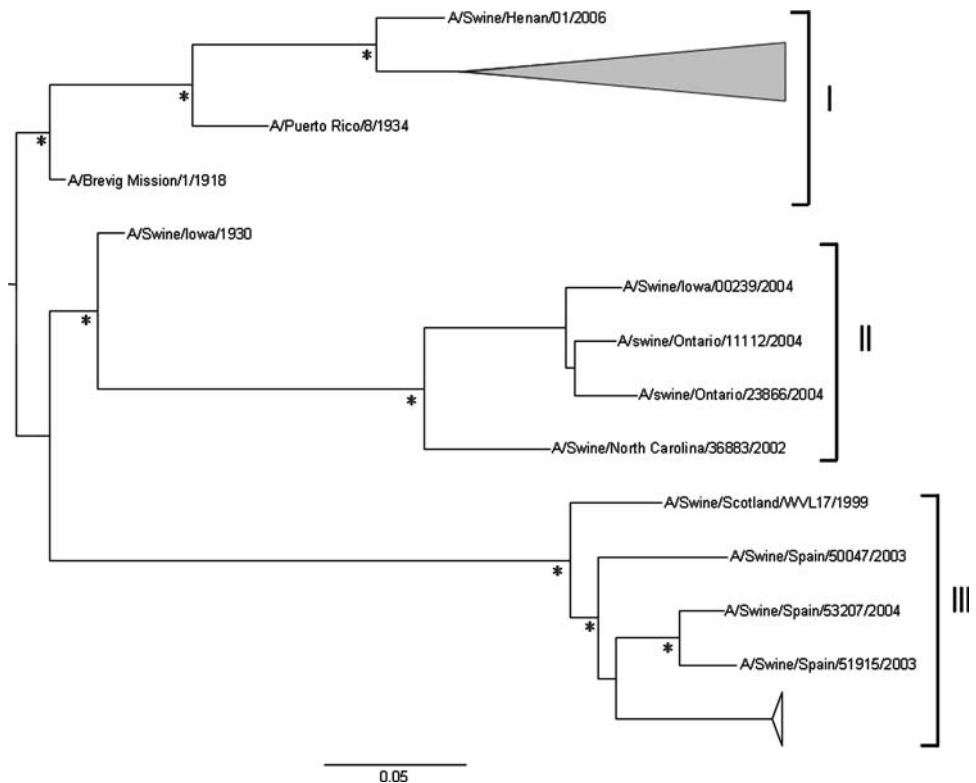
tree was performed using an NJ tree as the starting tree and the TBR branch-swapping algorithm. NJ trees were also estimated using pair-wise distances inferred by ML with the best-fitting nucleotide substitution model. Calculations were performed with PAUP* 4.0b10 according to Swofford and Sullivan [10]. Statistical support for internal branches in the NJ trees was obtained by bootstrapping (1,000 replicates) and with the ML-based zero-branch-length test for the ML trees [10]. In the phylogenetic analysis, the trees were unrooted.

Phylogenetic analysis of the HA gene showed that viruses cluster into three main groups (Fig. 1). The first group contains all classical H1N1 viruses isolated from humans from 2000 to 2008 (gray triangle) as well as A/Swine/Henan/01/2006, A/Puerto Rico/8/1934, and A/South Carolina/1/1918, which correspond to Chinese swine, classical human, and 1918 pandemic isolates, respectively. The second group contains all novel H1N1 2009 viruses isolated from humans (white triangle). They did not cluster with the other human viruses isolated in the previous years (2000–2008, gray triangle), but clustered with the H1N1 influenza isolated from American pigs (A/Swine/Iowa/15/1930, A/Swine/Ontario/11112/2004, A/Swine/Ontario/23866/2004, A/Swine/Iowa/00239/2004, A/Swine/North Carolina/

36883/2002, and A/Swine/OH/511445/2007). This result suggests that the H1 of this novel virus may come from American swine. In addition, the presence of the novel H1N1 isolates and A/Swine/Iowa/15/1930 within the same cluster may suggest that the H1 gene has been circulating among American swine for decades. The third group contains only the European H1N1 isolated in swine (A/Swine/Haseluenne/IDT2617/2003, A/Swine/Spain/50047/2003, A/Swine/Spain/51915/2003, A/Swine/Spain/53207/2004, and A/Swine/Scotland/WVL17/1999), confirming that the H1 of this novel virus was not from European pigs, but came from American swine.

Phylogenetic trees of the other two gene segments (NA and MP) showed a similar pattern, as shown in Fig. 2; Fig. S1, respectively. Viruses with N1 NA and MP clustered into three main groups. The first group contains all classical H1N1 isolated from humans from 2000 to 2008 (gray triangle) as well as A/Swine/Henan/01/2006, A/Puerto Rico/8/1934, and A/South Carolina/1/1918, which correspond to Chinese swine, classical human, and 1918 pandemic isolates, respectively. The second group comprises H1N1 isolated from American pigs as shown in Fig. 2 (A/Swine/Iowa/15/1930, A/Swine/Ontario/11112/2004, A/Swine/Ontario/23866/2004, A/Swine/Iowa/00239/

Fig. 2 Phylogenetic analysis of the neuraminidase (NA) gene (1,410 nt) of the H1N1 influenza A viruses was carried out using HKY + I + C (alpha-parameter = 0.0763) as the best evolutionary model. Branch lengths were estimated with the best-fitting nucleotide substitution model according to a hierarchical likelihood ratio test [7] and were drawn to scale, with the bar at the bottom indicating 0.05 nucleotide substitutions per site. One asterisk along a branch represents significant statistical support for the clade subtending that branch ($P < 0.001$ in the zero-branch-length test) and bootstrap support $>95\%$. The tree was unrooted. Three main clades are indicated (I, II, and III). All H1N1 viruses isolated from 2000 to 2008 are presented as a gray triangle, and all novel 2009 H1N1 viruses as a white triangle



2004, and A/Swine/NorthCarolina/36883/2002) and A/Swine/Ontario/11112/2004, A/Swine/Ontario/23866/2004, and A/Swine/Iowa/00239/2004, as presented in Fig. S1. In the third group, the novel H1N1 2009 viruses (white triangle) clustered with the H1N1 isolated from European pigs (A/Swine/Haseluenne/IDT2617/2003, A/Swine/Spain/50047/2003, A/Swine/Spain/51915/2003, A/Swine/Spain/53207/2004, and A/Swine/Scotland/WVL17/1999). This suggests that the novel H1N1 viruses acquired both NA and MP genes from European swine rather than American ones. Furthermore, within each gene segment (HA, NA, MP), there is high ($>99\%$) nucleotide identity among the outbreak viruses (white triangle) sequenced to date, suggesting that cross-species introduction into humans was a single event or multiple events of genetically very similar viruses [5]. The novel 2009 H1N1 human isolates did not cluster with other classical human isolates, confirming that we are in the presence of novel viruses. Phylogenetic analysis supports that the H1 of 2009 influenza virus was more probably re-assorted in American swine than those of other countries. Therefore, circulation of an influenza A (H1N1) swine-origin virus in humans with a genetically divergent HA and a previously unrecognized genetic composition [4, 5, 9] is of concern to public health officials worldwide. Likewise, this virus appears to be readily transmissible between humans. Thus, monitoring the genetic properties of the 2009 H1N1 viruses is necessary for selecting new vaccine candidates and antiviral agents.

Acknowledgment We are thankful for Francesco Greco, Amalia Mastrofrancesco and Valeria Cento for their excellent technical assistance.

References

- Centers for Disease Control, Prevention (CDC) (2009) Swine influenza A (H1N1) infection in two children—southern California, March–April 2009. *MMWR Morb Mortal Wkly Rep* 58(15):400–402
- Centers for Disease Control, Prevention (CDC) (2009) Update: swine influenza A (H1N1) infections—California and Texas, April 2009. *MMWR Morb Mortal Wkly Rep* 58(16):1–3
- Centers for Disease Control, Prevention (CDC) (2009) Outbreak of swine-origin influenza A (H1N1) virus infection—Mexico, March–April 2009. *MMWR Morb Mortal Wkly Rep* 58(17):467–470
- Novel swine-origin influenza A (H1N1) virus investigation team, Dawood FS, Jain S, Finelli L et al (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 et al (2009) Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* [Epub ahead of print]
- Hall TA (1999) Bioedit: A user-friendly biological sequence alignment, editor and analysis program for Windows 95/98 NT. *Nucl Acids Symp Ser* 41:95–98
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Scholtissek C (1990) Pigs as ‘mixing vessels’ for the creation of new pandemic influenza A viruses. *Med Princ Pract* 2:65–71
- Shinde V, Bridges CB, Uyeki TM et al (2009) Triple-reassortant swine influenza A (H1) in humans in the United States, 2005–2009. *N Engl J Med* 360:2616–2625

10. Swofford D, Sullivan J (2003) Phylogeny inference based on parsimony and other methods with PAUP*. In: Salemi M, Vandamme AM (eds) *The phylogenetic handbook—a practical approach to DNA and protein phylogeny*. Cambridge University Press, New York, pp 160–206
11. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1994) CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882