# Phylogenetic and evolutionary analysis of influenza A H7N9 virus

**Muhammed Babakir-Mina[1,2], Salvatore Dimonte[3,4], Alessandra Lo Presti[5], Eleonora Cella[5], Carlo Federico Perno[3,4,] Marco Ciotti[3], Massimo Ciccozzi[5,6]**

*[1]Biomedical Research Center, Sulaimani Polytechnic University, Sulaimaniyah, Iraqi Kurdistan Region;*
*[2]Laboratory of Molecular Virology, Polyclinic Tor Vergata Foundation, Rome, Italy;*
*[3]Biomolecular Laboratory, Barletta, Italy;*
*[4]Department of Experimental Medicine and Surgery, "Tor Vergata" University of Rome, Italy;*
*[5]Department of Infectious, Parasitic and Immunomediated Diseases, Istituto Superiore di Sanità, Rome, Italy;*
*[6]University Campus Biomedico, Rome, Italy*

## SUMMARY

Recently, human infections with the novel avian-origin influenza A H7N9 virus have been reported from various provinces in China. Human infections with avian influenza A viruses are rare and may cause a wide spectrum of clinical symptoms. This is the first time that human infection with a low pathogenic avian influenza A virus has been associated with a fatal outcome. Here, a phylogenetic and positive selective pressure analysis of haemagglutin (HA), neuraminidase (NA), and matrix protein (MP) genes of the novel reassortant H7N9 virus was carried out. The analysis showed that both structural genes of this reassortant virus likely originated from Euro-Asiatic birds, while NA was more likely to have originated from South Korean birds. The Bayesian phylogenetic tree of the MP showed a main clade and an outside cluster including four sequences from China. The United States and Guatemala classical H7N9-isolates appeared homogeneous and clustered together, although they are distinct from other classical Euro-Asiatic and novel H7N9 viruses. Selective pressure analysis did not reveal any site under statistically significant positive selective pressure in any of the three genes analyzed. Unknown certain intermediate hosts involved might be implicated, so extensive global surveillance and bird-to-person transmission should be closely considered in the future.

*KEY WORDS:* Novel influenza virus, Pandemic, Avian flu outbreak, Re-assortment, Phylogenetic analysis, Selective pressure analysis.

*Received* January 1, 2014 *Accepted* June 8, 2014

As human infections with the novel influenza A H7N9 virus have been reported from some provinces in China (Chowell *et al.*, 2013), researchers, public health experts and all others concerned including the general public have started to question the pandemic potential of this virus (Butler, 2013). The question largely stems from the fact that a high transmission potential could indicate a high risk of a new pandemic. While the direct link between animal and human has yet to be established (Uyeki and Cox, 2013), preliminary genetic analysis of the virus, virus isolation from poultry, quail, and a contact history with birds among some of the confirmed cases imply that these cases were likely caused by bird-to-human transmission (Fang *et al.*, 2013). Otherwise, the sources of infection have yet to be firmly clarified and confirmed cases from one province have not been linked to those from other provinces. *Eurosurveillance* has published some timely papers related to the emergence of this new influenza A H7N9 virus affecting humans in China (Kageyama *et al.*, 2013; Jonges *et al.*, 2013; Corman *et al.*, 2013). Genetic studies by Kageyama *et al.,* Jongens *et al.*, and Liu *et al.* assessed the virus origin, its adaptation and virulence (Liu *et al.*, 2013). In line

*Corresponding author*
Massimo Ciccozzi
Department of Infectious
Parasitic and Immunomediated Diseases
Istituto Superiore di Sanità, Rome, Italy
E-mail: massimo.ciccozzi@iss.it

with our previous studies on influenza viruses (Babakir-Mina *et al.*, 2010; Babakir-Mina *et al.*, 2009; Babakir-Mina *et al.*, 2007), we investigated the genetic diversity of the H7N9 novel virus. The present study downloaded all available complete genome sequences of hemagglutinin (HA), neuraminidase (NA), and matrix protein (MP) from human and bird H7N9 viral isolates from GenBank and analysed them by phylogenetic analysis. The site-specific positive and negative selection for these genes was also estimated. The sequences from 1988 to 2013 were downloaded from the NCBI influenza virus resource (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html). GenBank accession numbers are reported in Supplementary Table 1. Three different data sets were built for HA, NA, and MP gene segments. A letter code was assigned to the sequences to identify the host. Sequences isolated from humans were identified with an "h" code; sequences from goose with a "g" code; sequences from duck, anas crecca, mallard and blue-winged teal with a "d" code; sequences isolated from chicken were indicated with a "c" code; sequences from guinea fowl were identified by an "f" code; sequences isolated from turkey were identified by a "t" code; sequences identified in the environment were denoted by an "e" letter code; sequences from bird, ruddy turnstone, were identified by a "b" code.

The first dataset included 28 Influenza A H7N9 haemagglutinin (HA) gene sequences which can be distinguished based on the host as follows: three "h" sequences; four "g" sequences; 11 "d" sequences; one "c" sequence; one "f" sequence; two "t" sequences; one "e" sequence, and five "b" sequences. The sampling locations of the sequences were: China, CN (n=4); Czech Republic, CZ (n=2); United States, US (n=10); Japan, JP (n=1); Korea, KR (n=4); Mongolia, MN (n=2); Spain, ES (n=2); Sweden, SE (n=1); Guatemala, GT (n=2).

The second dataset included 24 Influenza A H7N9 neuraminidase (NA) gene sequences that based on the host of origin were classified into three "h" sequences; four "g" sequences; eight "d" sequences; one "c" sequence; one "f" sequence; two "t" sequences; five "b" sequences. The sampling locations of the sequences were: United States, US (n=9); Guatemala, GT (n=2); Czech Republic, CZ (n=2); Mongolia, MN (n=1); Korea, KR (n= 5); Spain, ES (n=1); China, CN (n=4).

The third dataset included 27 Influenza A H9N2 M gene sequences from China plus 24 Influenza A H7N9 M gene sequences. Based on the host, the Influenza A H7N9 M gene sequences were divided into three "h" sequences; three "g" sequences; eight "d" sequences; one "c" sequence; three "t" sequences; six "b" sequences. The sampling locations of the H7N9 sequences were China, CN (n=4); Czech Republic, CZ (n=2); United States, US (n=9); Guatemala, GT (n=3); Mongolia, MN (n=1); Spain, ES (n=1); Korea, KR (n=4). The H9N2 matrix protein gene influenza A virus sequences from China were reported as accession number - host - location.

The phylogenetic signal of each sequence dataset was investigated by means of likelihood mapping analysis of 10,000 random quartets generated using TreePuzzle (Strimmer and von Haeseler, 1997). Groups of four randomly chosen sequences (quartets) were evaluated. For each quartet the three possible unrooted trees were reconstructed using the maximum likelihood approach under the selected substitution model. The posterior probabilities of each tree were then plotted on a triangular surface so that fully resolved trees fall into the corners and the unresolved quartets in the centre of the triangle (a star tree). When using this strategy, if more than 30% of the dots fall into the centre of the triangle, the data are considered unreliable for the purposes of phylogenetic inference. All the data sets were aligned using CLUSTAL X software as already described (Ciccozzi *et al.*, 2011), then manually edited with the Bioedit software (Ciccozzi *et al.*, 2011). The ModelTest program v. 3.7 was used to select the simplest evolutionary model that fitted the sequence data adequately (Lo Presti *et al.*, 2012). The Bayesian phylogenetic tree was reconstructed by means of Mr Bayes (Zehender *et al.*, 2010) using the GTR + I + G model of nucleotide substitution for the first and third datasets and the HKY + G for the second dataset. For each dataset a Markov Chain Monte Carlo search was made for $10x10^6$ generations using tree sampling every 100th generation and a burn-in fraction of 25%. Statistical support for specific clades was obtained by calculating the posterior probability of each monophyletic clade, and

TABLE 1 - *The accession numbers of all three genes analyzed (HA, NA, and Matrix protein) of the novel (H7N9)-2013 and all the other classical (H7N9) viruses.*

| *Virus strain (H7N9)* | *HA accession no.* | *NA accession no.* | *M protein accession no.* |
|---|---|---|---|
| A/Hangzhou/1/2013 | KC853766 | KC853765 | KC853764 |
| A/Shanghai/4664T/2013 | KC853228 | KC853231 | KC853227 |
| A/goose/CzechRepublic/1848K9/2009 | GU060482 | GU060484 | GQ404575 |
| A/goose/Czech Republic/1848T14/2009 | HQ244415 | HQ244417 | HQ244418 |
| A/northernshoverl/Mississippi/11OS145/2011 | CY133649 | CY133651 | CY133650 |
| A/ruddy turnstone/Delaware Bay/220/1995 | CY127253 | CY127255 | CY127254 |
| A/duck/Mongolia/119/2008 | AB481212 | AB481213 | N |
| A/duck/Gunma/466/2011 | AB813056 | N | N |
| A/chicken/Zhejiang/DTID-ZJU01/2013 | KC899669 | KC899671 | KC899672 |
| A/Zhejiang/DTID-ZJU01/2013 | KC885956 | KC885958 | KC885959 |
| A/environment/Colorado/NWRC186223-18/2007 | CY122531 | N | N |
| A/emperor goose/Alaska/44063-061/2006 | JX080746 | JX081137 | JX081204 |
| A/goose/Nebraska/17097-4/2011 | JX899805 | JX899806 | N |
| A/guinea fowl/Nebraska/17096-1/2011 | JX899803 | JX899804 | N |
| A/wild duck/Mongolia/1-241/2008 | JN029686 | N | JN029688 |
| A/wild bird/Korea/A14/2011 | JN244231 | JN244222 | JN244134 |
| A/spot-billed duck/Korea/447/2011 | JN244236 | JN244224 | JN244142 |
| A/wild bird/Korea/A9/2011 | JN244234 | JN244225 | JN244136 |
| A/wild bird/Korea/A3/2011 | JN244232. | JN244223 | JN244135 |
| A/ruddy turnstone/DE/1538/2000 | EU684261 | HQ541731 | DQ021735 |
| A/mallard/Spain/08.00991.3/2005 | GU354035 | GU354036 | N |
| A/Anas crecca/Spain/1460/2008 | HQ244407 | HQ244409 | HQ244410 |
| A/turkey/Minnesota/38429/1988 | GU053163 | GU053165 | GU053164 |
| A/Mallard/Sweden/91/02 | AY999981 | N | N |
| A/blue-winged teal/Ohio/566/2006 | CY024818 | CY024820 | CY024819 |
| A/turkey/Minnesota/1/1988 | CY014786 | CY014788 | CY014787 |
| A/blue-winged teal/Guatemala/CIP049-02/2008 | CY067678 | CY067680 | CY067681 |
| A/blue-winged teal/Guatemala/CIP049-01/2008 | CY067670 | CY067672 | CY067673 |

N= not available sequences in GebBank.

a posterior consensus tree was generated after a 25% burn-in. Clusters were recognized on the basis of the same tree, and posterior probability was used as a statistical support for each clade (considering significant cluster supported by a posterior probability >0.95). Selective pressure analysis was performed on the first, second and the H7N9 sequences of the third dataset. The dN/dS rate (ω) was estimated by the ML approach implemented in the programme HyPhy (Pond and Muse, 2005). Site-specific positive and negative selection were estimated by two different algorithms: the fixed-effects likelihood (FEL), which fits an (ω) rate to every site and uses the likelihood ratio to test if dN≠dS, and random effect likelihood (REL), a variant of the Nielsen–Yang approach (Nielsen&Yang, 1998), which assumes that a discrete distribution of rates exists across sites and allows both dS and dN to vary independently site by site. The two methods have been described in more detail elsewhere (Kosakovsky Pond and Frost 2005). In order to select sites under selective pressure and keep our test conservative, a P value of ≤0.1 or a posterior probability of ≥0.9 as relaxed critical values (Kosakovsky Pond and Frost 2005) was assumed. Part of the analysis was conducted using the web-based interface Datamonkey (http://www.datamonkey.org/).

The phylogenetic noise of each data set was investigated by means of likelihood mapping. The percentage of dots falling in the central area of the triangles was 0.8%, 2.2% and 9.8% for the first, second and third dataset, respectively. As none of the datasets showed more than 30% noise, all of them contained a sufficient phylogenetic signal (Figure 1 panel a, b and c).

The Bayesian phylogenetic tree of the first dataset (Figure 2) revealed two main statistically supported clades. The first clade (clade I) included two statistically supported clusters: the largest one was characterized by sequences isolated from goose (g); duck, anas crecca, mallard, blue-winged teal (d) and bird, ruddy turnstone (b); with different sampling locations (Czech Republic, Japan, Korea, Mongolia and Spain); the smallest cluster included only four sequences from China, three of them isolated from humans and one from chicken. Within this smallest cluster the novel H7N9 2013 viruses appeared to be more closely related to a Mongolian isolate (A/duck/Mongolia/119/2008 or 6dMn-08) whose H7 gene may probably haveoriginated from Mongolian duck H7N9 isolates as well as the other Euro-Asiatic birds (Figure 2). In addition, the clustering of the novel H7N9 isolates with the viruses of South Korean, Mongolia, and some other European countries within the same clade (clade 1) may suggest that the H7 gene has been circulating

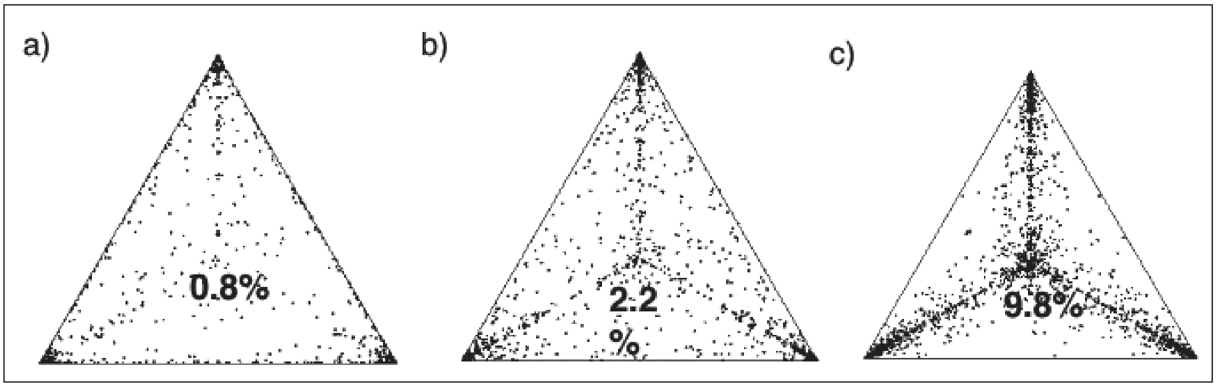

FIGURE 1 - *The phylogenetic signal of each sequence dataset was investigated by means of likelihood mapping analysis of 10,000 random quartets generated using TreePuzzle (Strimmer and von Haeseler, 1997). Groups of four randomly chosen sequences (quartets) were evaluated. For each quartet the three possible unrooted trees were reconstructed using the maximum likelihood approach under the selected substitution model. The posterior probabilities of each tree were then plotted on a triangular surface so that fully resolved trees fall into the corners and the unresolved quartets in the centre of the triangle (a startree). When using this strategy, if more than 30% of the dots fall into the centre of the triangle, the data are considered unreliable for the purposes of phylogenetic inference.*

among the birds of those countries for the past few years. The second statistically supported clade (clade II) included sequences from only two locations (United States and Guatemala) but with different isolation hosts (d, g , f, e, b and t). So, the phylogenetic analysis revealed that these isolates are highly similar to viruses from the American lineage suggesting that bird migration dictates the ecology of these viruses in the Guatemalan bird population (González-Reiche *et al.,* 2012). The Bayesian phylogenetic tree of the second dataset (Figure 3) also revealed two main statistically supported clades (named as before). The first was composed of sequences from the United States and Guatemala, with different isolation hosts (d, g, f, b, t). The second statistically supported clade was characterized by sequences from different locations (Czech Republic, Mongolia, Korea, Spain and China) and with different isolation hosts (g, d, b, h, c). Inside the second clade sequences from Korea and China segregated mostly into

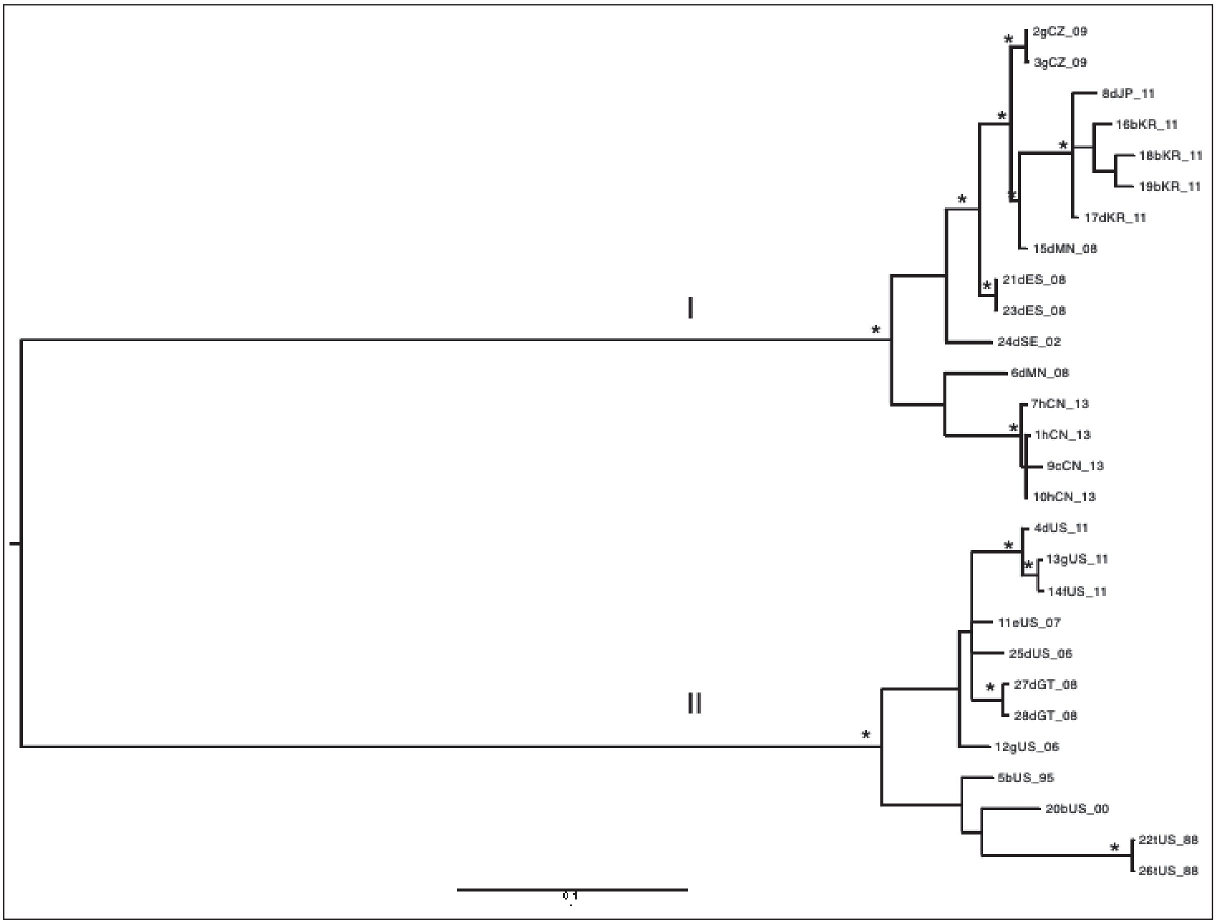

FIGURE 2 - *Phylogenetic analysis of the haemagglutin (HA) gene of the H7N9 influenza A viruses was carried out using GTR + I + G as the best evolutionary model. Branch lengths were estimated with the best-fitting nucleotide substitution model according to a hierarchical likelihood ratio test and were drawn to scale, with the bar at the bottom indicating 0.1 nucleotide substitutions per site. One asterisk along a branch represents significant statistical support for the clade subtending that branch (posterior probability >0.95). The tree was midpoint rooted. The two main clades are indicated (I and II). From 28 sequences, three sequences were assigned to the "h" code; four sequences to the "g" code; 11 sequences to the "d" code; one sequence to "c" code and one to the "f" code; two sequences to "t" code; one to the "e" code and five sequences to the "b" code. The sampling locations of the sequences were: China, CN (n=4); Czech Republic, CZ (n=2); United States, US (n=10); Japan, JP (n=1); Korea, KR (n=4); Mongolia, MN (n=2); Spain, ES (n=2); Sweden, SE (n=1); Guatemala, GT (n=2). All letter codes were mentioned in the text.*

two different statistically supported clusters. This clustering indicated that N9 gene may have nearly originated from the Korean wild bird-H7N9 isolates as well as the other Euro-Asiatic birds (Figure 3). This result suggests that the novel H7N9 virus probably circulated in the past two years in South Korea and then these viruses were re-assorted in China on February 2013. The Bayesian phylogenetic tree of the third dataset (Figure 4) showed a statistically supported cluster and a main clade (M). The statistically supported cluster included two H9N2 sequences isolated from chicken (both from China) and four H7N9 sequences from China, three of them isolated from humans and one from chicken. The main clade was composed of two subclades M1 and M2. M1 included the majority of the H9N2 sequences isolated from poultry from China; M2 included one H9N2 sequence from quail from Hong Kong as a sort of outgroup for this clade, and the Influenza A H7N9 M gene sequences. Inside

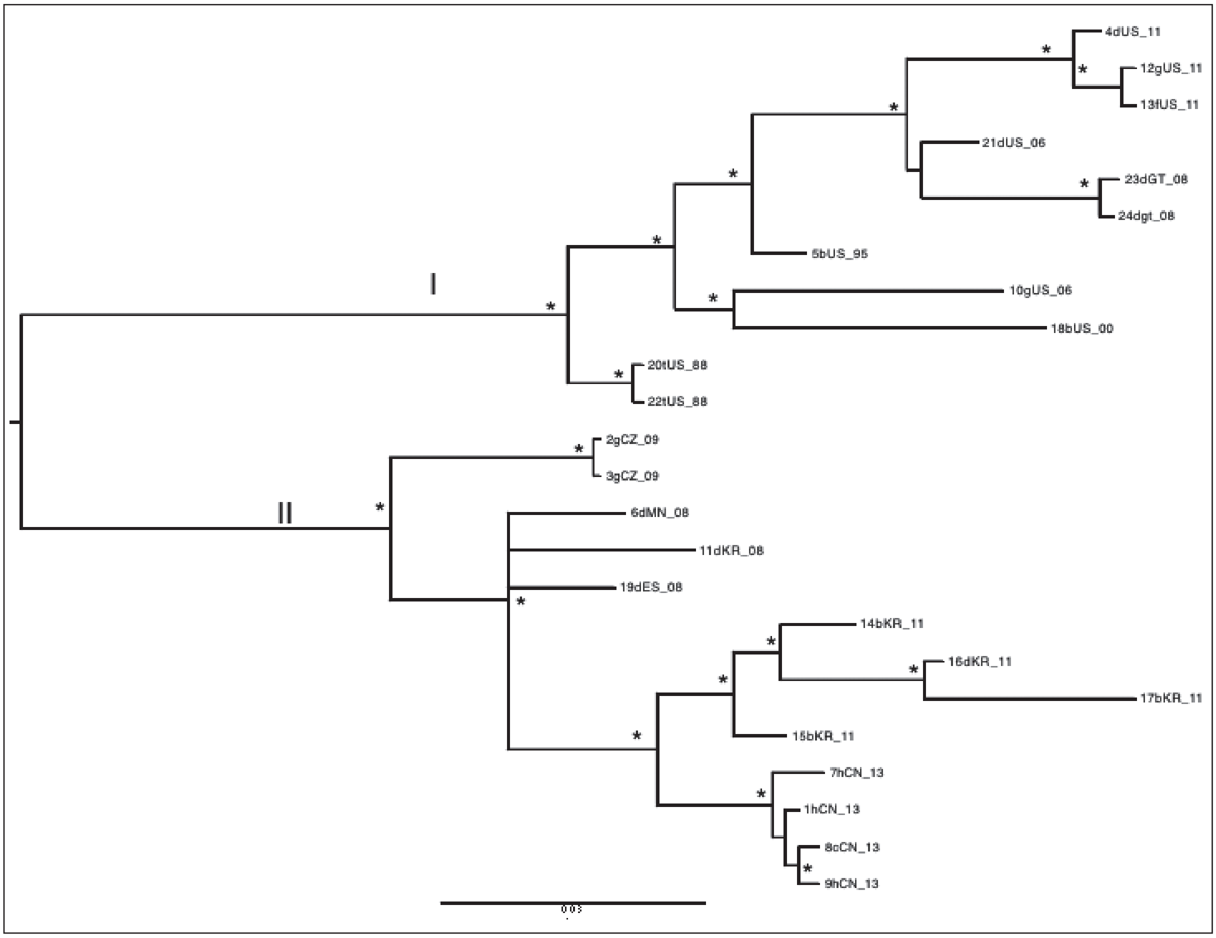

FIGURE 3 - *Phylogenetic analysis of the neuraminidase (NA) gene of the H7N9 influenza A viruses was carried out using HKY + G as the best evolutionary model. Branch lengths were estimated with the best-fitting nucleotide substitution model according to a hierarchical likelihood ratio test and were drawn to scale, with the bar at the bottom indicating 0.03 nucleotide substitutions per site. One asterisk along a branch represents significant statistical support for the clade subtending that branch (posterior probability>0.95). The tree was midpoint rooted. The two main clades are indicated (I and II). From 24 sequences, the three sequences were assigned to the "h" code; four to the "g" code; eight to the "d" code; one sequence to the "c" code; one to the "f" code; two to the "t" code; five to the "b" code. The sampling locations of the sequences were: United States, US (n=9); Guatemala, GT (n=2); Czech Republic, CZ (n=2); Mongolia, MN (n=1); Korea, KR (n=5); Spain, ES (n=1); China, CN (n=4). All letter codes were mentioned in the text.*

this clade, the H7N9 MP gene sequences were arranged in a group of two sequences (isolated from goose and sampled in the Czech Republic) and two clusters, all statistically supported. The largest one was composed of nine sequences from the United States and three from Guatemala. Inside this cluster sequences from different virus hosts were mixed together except the sequences isolated from turkey; the smallest cluster included a group of two sequences (one from Mongolia and the other from Spain, both of them with the "d" host code) and four sequences from Korea, three of them with the "b" sampling host code (bird, ruddy turnstone) and one with the "d" sampling host code (duck, amass crecca, mallard, blue-winged teal). Overall, both HA and NA genes probably originated from Eurasian avian influenza viruses while the remaining MP gene is closely related to the other avian influenza virus genotypes (Kageyama *et al.*, 2013). Selective pressure analysis did not reveal any site under statistically significant

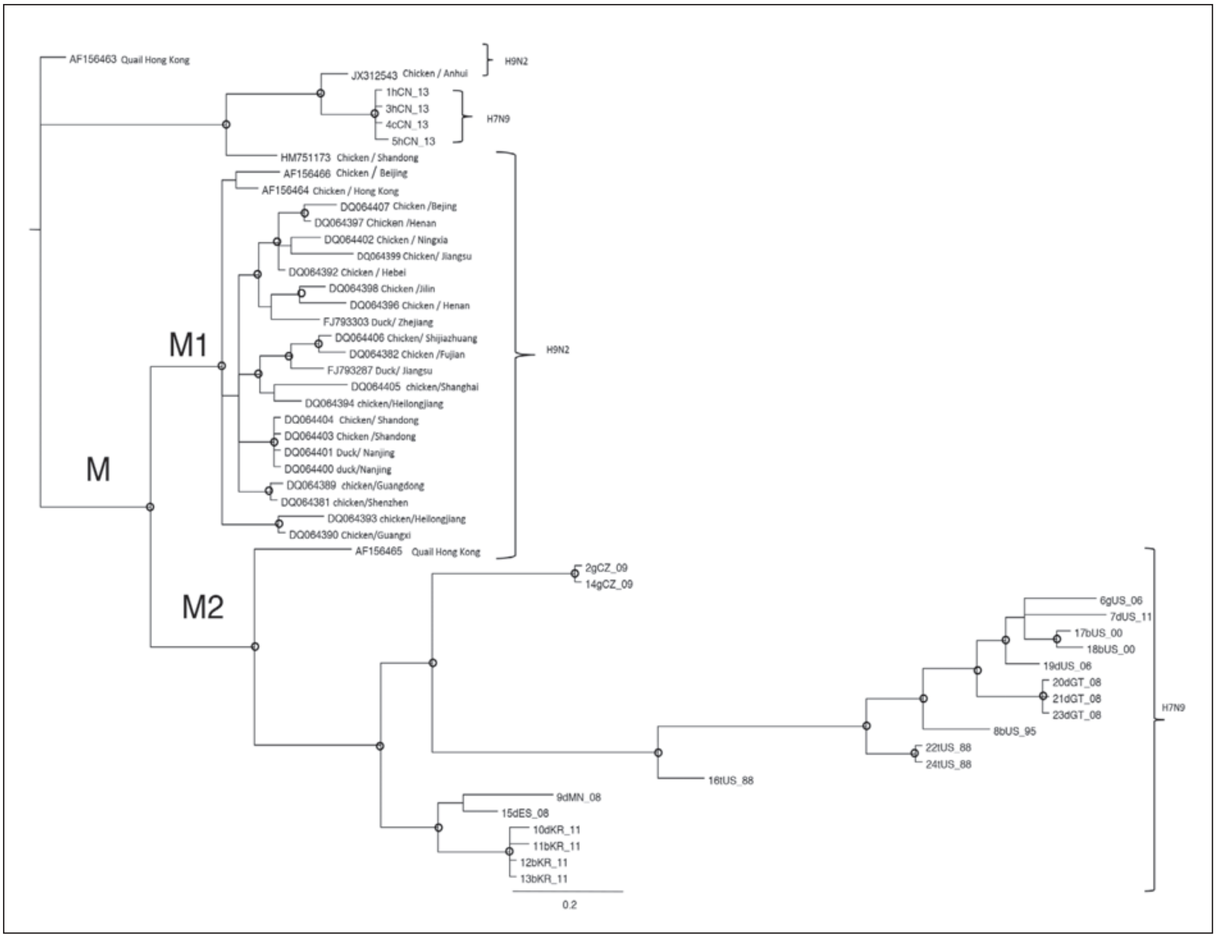

FIGURE 4 - *Phylogenetic analysis of the matrix protein (MP) gene of the H7N9 influenza A viruses and the matrix protein (MP) gene of the H9N2 viruses in China. Branch lengths were estimated with the best-fitting nucleotide substitution model (GTR+I+G) according to a hierarchical likelihood ratio test and were drawn to scale, with the bar at the bottom indicating 0.2 nucleotide substitutions per site. One circle along a branch represents significant statistical support for the clade subtending that branch (posterior probability>0.95). The tree was unrooted. For the 24 H7N9 sequences: three were assigned to the "h" code; three to the "g" code; eight to the "d" code; one to the "c" code; three sequences to the "t" code; six to the "b" code. The sampling locations of the H7N9 sequences were China, CN (n=4); Czech Republic, CZ (n=2); United States, US (n=9); Guatemala, GT (n=3); Mongolia, MN (n=1); Spain, ES (n=1); Korea, KR (n=4). All letter codes were mentioned in the text. The H9N2 matrix protein gene influenza A virus sequences from China were reported as accession number, host and location.*

positive selective pressure in any dataset analyzed. Hyphy analysis revealed 191/553 (34.5%) codons under statistically significant negative pressure in the first dataset (HA gene); 127/454 (27.97%) codons under significant negative selective pressure in the second dataset; and 83/319 (26%) codons under significant negative selective pressure in the third dataset (M gene). Finally, some characteristic amino acid changes in H7N9 novel viruses (Kageyama *et al.,* 2013) probably facilitate binding to human-type receptors and efficient replication in mammals (Xiong *et al.*, 2013; Lui *et al.*, 2013). Therefore, it is prudent to monitor the evolution of influenza A H7N9 virus, as well as to develop strategies to combat any potential pandemic.

*ACKNOWLEDGEMENT*
*We thank Bahra Rzgar Abdullah, Francesco Greco, and Stefania Carmen Spinazzola for their excellent technical assistance.*

## REFERENCES

Babakir-Mina M., Dimonte S., Ciccozzi M., Perno C.F., Ciotti M. (2010). The novel swine-origin H1N1 influenza A virus riddle: is it a domestic bird H1N1-derived virus? *New Microbiol.* **33**, 77-81.

Babakir-Mina M., Ciccozzi M., Ciotti M., Marcuccilli F., Balestra E., Dimonte S., Perno C.F., Aquaro S. (2009). Phylogenetic analysis of the surface proteins of influenza A (H5N1) viruses isolated in Asian and African populations. *New Microbiol.* **32**, 397-403.

Babakir-Mina M., Balestra E., Perno C.F., Aquaro S. (2007). Influenza virus A (H5N1): a pandemic risk? *New Microbiol.* **30**, 65-78.

Butler D. (2013). Novel bird flu kills two in China. *Nature* 2013 (Available from: [http://www.nature.com/news/novel-bird-flu-kills-two-in-china-1.12728].

Chowell G., Simonsen L., Towers S., Miller M.A., Viboud C. (2013). Transmission potential of influenza A/H7N9, February to May 2013, China. *BMC. Med*. **11**, 214.

Ciccozzi M., Babakir-Mina M., Lo Presti A., Marcuccilli F., Perno C.F., Ciotti M. (2011). Phylogenesis and Clinical Aspects of Pandemic 2009 Influenza A (H1N1) Virus Infection. *Open Virol. J.* **5**, 22-26.

Corman V.M., Eickmann M., Landt O., Bleicker T., Brunink S., Eschbach-Bludau M., et al. (2013). Specific detection by real-time reverse-transcription PCR assays of a novel avian influenza A(H7N9) strain associated with human spillover infections in China. *Euro Surveill.* **18**, pii=20461.

Fang L.Q., Li X.L., Liu K., Li Y.J., Yao H.W., Liang S., Yang Y., Feng Z.J., Gray G.C., Cao W.C. (2013). Mapping Spread and Risk of Avian Influenza A (H7N9) in China. *Sci. Rep*. **3**, 2722.

González-Reiche A.S., Morales-Betoulle M.E., Alvarez D., Betoulle J.L., Müller M.L., Sosa S.M., Perez D.R. (2012). Influenza a viruses from wild birds in Guatemala belong to the North American lineage. *PLoS One*. **7**, e32873.

Jonges M., Meijer A., Fouchier R.A., Koch G., Li J., Pan J.C., et al. (2013). Guiding outbreak management by the use of influenza A(H7Nx) virus sequence analysis. *Euro Surveill.***18**, pii=20460.

Kageyama T., Fujisaki S., Takashita E., Xu H., Yamada S., Uchida Y., et al. (2013). Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. *Euro Surveill.* **11**, pii: 20453.

Kosakovsky Pond S.L., Frost S.D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208-1222.

Liu Q., Lu L., Sun Z., Chen G.W., Wen Y., Jiang S. (2013). Genomic signature and protein sequence analysis of a novel influenza A (H7N9) virus that causes an outbreak in humans in China. *Microbes. Infect.* pii: S1286-4579 (13) 00085-3.

Lo Presti A., Ciccozzi M., Cella E., Lai A., Simonetti F.R., Galli M., Zehender G., Rezza G. (2012). Origin, evolution, and phylogeography of recent epidemic CHIKV strains. *Infect. Genet. Evol.* **12**, 392-398.

Nielsen R., Yang Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics.* 148, 929-936.

Pond S.L., Frost S.D, Muse S.V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* **21**, 676-679.

Strimmer K., von Haeseler A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci USA.* **94**, 6815-6819.

Uyeki T.M., Cox N.J. (2013). Global concerns regarding novel influenza A (H7N9) virus infections. *N. Engl. J. Med*. doi:10.1056/NEJMp1304661. [Epub ahead of print].

Xiong X., Martin S.R., Haire L.F., Wharton S.A., Daniels R.S., Bennett M.S., McCauley J.W., Collins P.J., Walker P.A., Skehel J.J., Gamblin S.J. (2013). Receptor binding by an H7N9 influenza virus from humans. *Nature.* **499**, 496-499.

Zehender G., De Maddalena C., Canuti M., Zappa A., Amendola A., Lai A., Galli M., Tanzi E. (2010). Rapid molecular evolution of human bocavirus revealed by Bayesian coalescent inference. *Infect. Genet. Evol.* **10**, 215-220.