Differentiation of two distinct clusters among currently circulating influenza A(H1N1)v viruses, March-September 2009

S R Fereidouni (sasan.fereidouni@fli.bund.de)[1], M Beer[1], T Vahlenkamp[1], E Starick[1]
1. Friedrich-Loeffler-Institut, Greifswald-Insel Riems, Germany

This article was published on 19 November 2009.
Citation style for this article: Fereidouni SR, Beer M, Vahlenkamp T, Starick E. Differentiation of two distinct clusters among currently circulating influenza A(H1N1)v viruses, March-September 2009. Euro Surveill. 2009;14(46):pii=19409. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19409

Analysis of all complete genome sequences of the pandemic influenza A(H1N1)v virus available as of 10 September 2009 revealed that two closely related but distinct clusters were circulating in most of the affected countries at the same time. The characteristic differences are located in genes encoding the two surface proteins - haemagglutinin and neuraminidase - and four internal proteins – the polymerase PB2 subunit, nucleoprotein, matrix protein M1 and the non-structural protein NS1. Phylogenetic inference was demonstrated by neighbour joining, maximum likelihood and Bayesian trees analyses of the involved genes and by tree construction of concatenated sequences.

Following the worldwide spread of the pandemic influenza A(H1N1)v virus after its emergence in the United States (US) and Mexico in March 2009, the World Health Organization (WHO) raised the influenza pandemic alert level to phase 6 on 11 June 2009. It is expected that this new influenza virus will continue to circulate and spread due to efficient human to human transmission. Data on the genetic composition of the virus became available very early in the pandemic [1], and until 10 September 2009, more than 3,500 individual gene sequences had been deposited in public databases such as GISAID and GenBank. The influenza A(H1N1)v virus, which is a unique combination of gene segments from both North American and Eurasian swine influenza viruses [2], has a high mean evolutionary rate for individual segments and the whole genome (3.66×10^{-3} substitutions per site per year) [3].

Analysis of all eight gene segments of more than 300 full-length influenza A(H1N1)v sequences available in the Genbank database (Figure 1; this figure is only available in the online version) enabled us to show that two closely related but distinct clusters of the virus were circulating in most of the affected countries at the same time. The two clusters could be differentiated clearly by nine nucleotide signatures. These were located in the genes for the two surface proteins haemagglutinin HA and neuraminidase NA and in the genes for four internal proteins, the polymerase PB2 subunit, the nucleoprotein NP, matrix protein M1 and the non-structural protein NS1. The polymerase genes PB1 and PA were identical in all isolates and no genetic signature was evident in these two segments. Four of the nine nucleotide changes, present on the HA, NA, NP and NS1 segments, were non-synonymous and lead to amino acid replacements (Table). Eight of the mutations were transition substitutions (seven of them A/G substitutions), and one change was a transversion substitution (A/T substitution). None of the changes in the sequences seemed to be located in regions of the genome responsible for known phenotypic differences or biological functions.

The differentiation of circulating influenza A(H1N1)v viruses into two clusters based on their nucleotide sequence differences was also supported by phylogenetic inference. Concatenated sequences were prepared using open reading frames of six viral segments (the ones included in the Table). Distance-based neighbour-joining trees were constructed using the Tamura 3-parameter model available in MEGA 4.0 [4]. Clustering of influenza A(H1N1)v viruses could be demonstrated by individual trees of the involved single genes (not shown) and with higher evidence by tree construction of concatenated sequences (Figure 2), despite the fact that the differences between the two clusters comprised only a few nucleotides. The analyses were supported by maximum likelihood using generalised time reversible substitution model (GTR) and Bayesian inference implemented in TOPALi v2 [5]. All phylogenetic analyses were conducted on all available sequences (Figures 3

Table

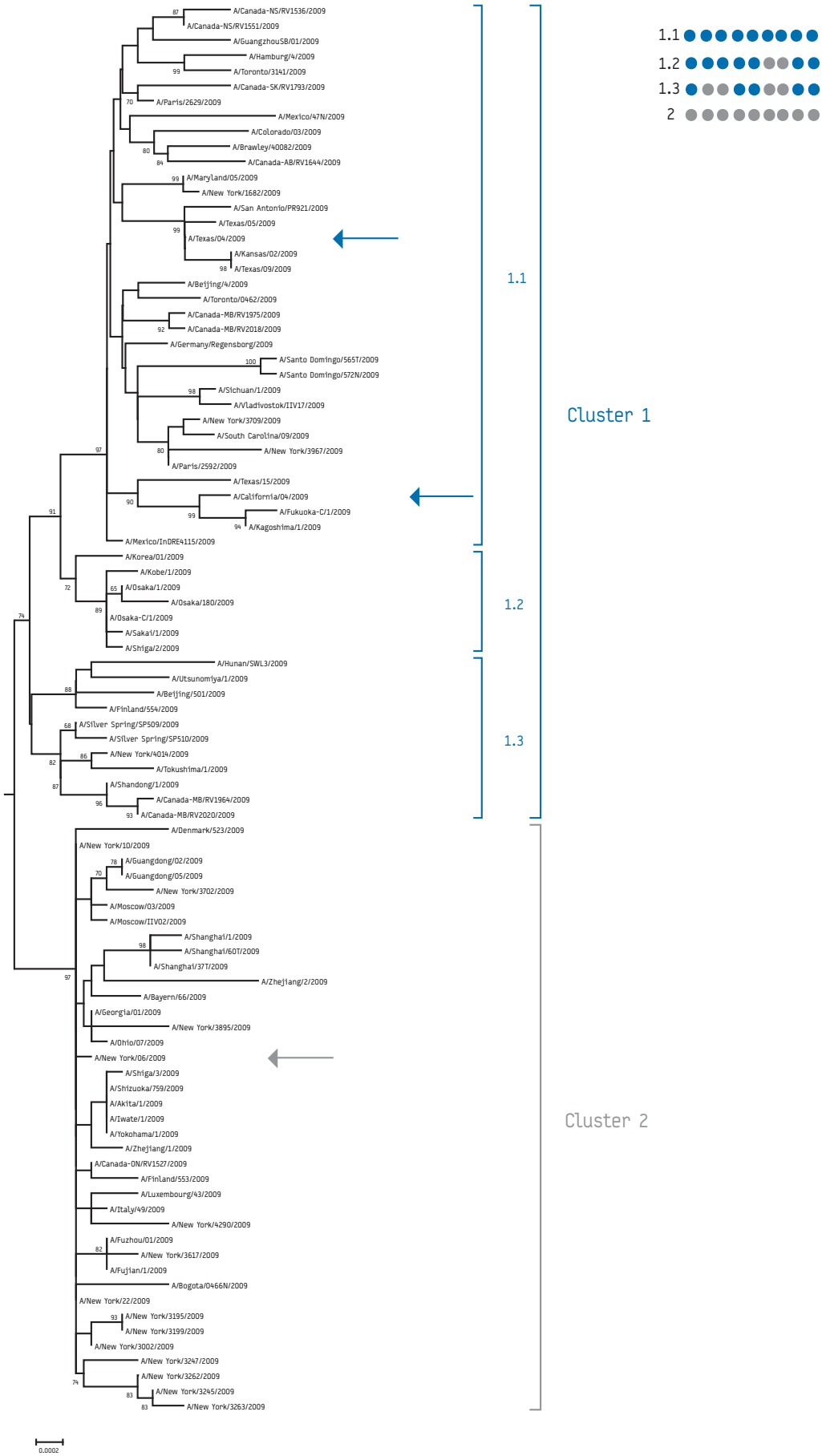
Nucleotide and amino acid residues located in six segments of the new H1N1 influenza viruses specific for the two clusters

	HA		NA	M		NP		NS	PB2
	658 (220)[a]	1,408 (470)	742 (248)	492 (164)	600 (200)	298 (100)	1,143 (381)	367 (123)	2,163 (721)
Cluster 1	T (S)	C (L)	A (N)	G (Q)	G (A)	G (V)	G (A)	A (I)	G (K)
Cluster 2	A (T)	T (L)	G (D)	A (Q)	A (A)	A (I)	A (A)	G (V)	A (K)

Nucleotide positions and amino acid positions (in brackets) for all genes are counted from the start codon.

Figure 2

Neighbour-joining phylogenetic tree of concatenated open reading frames of six viral segments of selected influenza A(H1N1)v viruses



The tree was rooted to influenza A/Michigan/01/09 and was calculated using 1,000 bootstrap values. The blue line marks cluster 1, including three sub-clusters, and the grey line marks cluster 2. The blue arrows show the two first isolates of cluster 1 and the grey arrow show the first isolate of cluster 2. The blue and grey circles stand (from left to right) for nucleotide replacements at HA T658A and C1408T, NA A742G, M G492A and G600A, NP G298A and G1143A, NS A367G, and PB2 G2163A.

and 4; these figures are only available in the online version) and representatives of each monophyletic group (Figure 2).

Taking into account the complete sequence data available from Mexico and the US, it is noteworthy that viruses of cluster 1 occurred earlier than those of cluster 2, with a time difference of about two weeks. Most sequences from Mexico, Texas and California belonged to cluster 1, whereas most sequences from New York belonged to cluster 2. Whether these differences were due to the geographical region, the date of isolation or other reasons needs to be elucidated in further epidemiological investigations. Virus sequences of both clusters have been reported from most countries on different continents. In Germany, influenza virus A/Regensburg/2009 was one of the first influenza A(H1N1)v isolates and belonged to cluster 1 [6]. This virus has been investigated by whole genome sequencing (GenBank accession numbers: FN401574–FN401581) and animal experiments in pigs and chickens [7]. Interestingly, viruses of both clusters could be detected in Germany although complete sequences of all eight segments were available only for four viruses at the time of this analysis (Figure 2).

All available full-length sequences for the six segments with cluster specific signatures were selected and duplicate sequences from identical isolates were removed. Of 305 viruses included in the analyses, 150 belonged to cluster 1 and 155 to cluster 2. All viruses in cluster 2 shared nine genetic signatures specific for this cluster. In cluster 1, three sub-clusters were identified. Most viruses in cluster 1 share all nine genetic signatures specific for this cluster (sub-cluster 1.1). In contrast, most viruses from Japan belonged to cluster 1 but had a cluster 2-like nucleoprotein sequence. These viruses constitute sub-cluster 1.2 (Japanese sub-cluster). A small group of sequences fit into cluster 1 when the concatenated sequences were analysed but shared the same four sequence features with cluster 2 (sub-cluster 1.3) (Figure 2), which may point to a reassortment event between the two clusters. The importance of these findings and epidemiological links between different clusters remains to be analysed.

Our findings allow the differentiation of the influenza A(H1N1)v viruses into distinct clusters among the currently circulating influenza A(H1N1)v viruses, contributing additional knowledge of the new pandemic virus and encouraging further research on this topic.

References

1. Novel Swine-Origin Influenza A (H1N1) Investigation Team, Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, et al. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med. 2009;360(25):2605-15.
2. Centers for Disease Control and Prevention (CDC). Swine influenza A (H1N1) Infection in two children – Southern California, March-April 2009. MMWR Morb Mortal Wkly Rep. 2009;58(15):400-2.
3. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009;459(7250):1122-5.
4. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 2007;24(8):1596-9.
5. Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, Marshall DF, et al. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics. 2009;25(1):126-7.
6. Melzl H, Wenzel JJ, Kochanowski B, Feierabend K, Kreuzpaintner B, et al. First sequence-confirmed case of infection with the new influenza A(H1N1) strain in Germany. Euro Surveill. 2009;14(18):pii=19203. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19203
7. Lange E, Kalthoff D, Blohm U, Teifke JP, Breithaupt A, Maresch C, et al. Pathogenesis and transmission of the novel swine-origin influenza virus A/H1N1 after experimental infection of pigs. J Gen Virol. 2009;90(Pt 9):2119-23.