ORIGINAL ARTICLE

Genetic stability and linkage analysis of the 2009 influenza A(H1N1) virus based on sequence homology

Jun Lu · Dabin Liu · Wei Liu · Taoxing Shi ·
Yigang Tong · Wuchun Cao

Abstract The 2009 swine-origin Influenza A virus subtype H1N1 (S-OIV) is generally believed to be a mixture of human, bird and swine viruses, resulting from multiple reassortments. The evolutionary origin of the S-OIV is of high interest but still remains obscure. In order to understand the evolution of the new virus, we performed sequence homology, segment stability and segment linkage analysis, as well as analysis of the host and geographic distribution of the evolutionarily related viruses. Stability analysis demonstrated that segment 6 (NA) was the most unstable one, followed by segment 4 (HA), while the other 6 segments were relatively stable. Host and geographic distribution analysis indicated that all 8 segments of the new virus were closely related to those of swine influenza viruses circulating either in North America or in Eurasia. Segment linkage analysis showed that segments 1 (PB2), 2 (PB1), 3 (PA), 4 (HA), 5 (NP), and 8 (NS) are in linkage disequilibrium exclusively with North American swine influenza viruses, and segments 6 (NA) and 7 (M) are evolutionarily linked solely with Eurasian swine influenza viruses. Two North American swine strains and 2 Eurasian swine strains were identified as possible ancestors of S-OIV 2009. Based on the most recent linkage analysis with the updated influenza sequences, South Dakota avian strains were found to be the closest known relatives of S-OIV 2009.

J. Lu, D. Liu and W. Liu contributed equally to this work.

J. Lu · D. Liu · W. Liu · T. Shi · Y. Tong (✉) · W. Cao (✉)
State Key Laboratory of Pathogen and Biosecurity, Beijing
Institute of Microbiology and Epidemiology, Beijing, China
e-mail: tong.yigang@gmail.com

W. Cao
e-mail: caowc@nic.bmi.ac.cn

J. Lu
Beijing YouAn Hospital, Capital Medical University,
Beijing, China

Introduction

A new influenza A virus outbreak occurred in North America, originating in Mexico and the southwestern United States in or before April 2009 [2, 6]. The United States Centers for Disease Control and Prevention (CDC) reported the earliest two H1N1 cases in different counties of California on April 21, 2009 [1]. With thousands of suspected and confirmed H1N1 cases and over one hundred H1N1-attributed deaths in multiple countries around the world, the World Health Organization (WHO) formally declared the situation a "public health emergency of international concern" and changed its pandemic alert phase to the highest level, 6 [11]. The United States CDC has released sequence data on the emerging influenza virus through the Global Initiative on Sharing Avian Influenza Data (GISAID) and GenBank since late April 2009, but the source of the pathogen still remains obscure. Various sources of information have indicated that the emerging A(H1N1) virus appears to be a reassortant of four influenza viruses, one human, one bird and two swine strains [4]. Recent publications [2, 3, 9, 10] have suggested that S-OIV could be related to North American and Eurasian swine influenza viruses.

To better understand the evolution of S-OIV, we performed BLAST analysis to study the sequence homology of the genome of the new influenza viruses. Instead of employing a phylogenetic method, which involves collecting and depositing a considerable number of influenza

virus sequences and uses specialized software with elusive algorithms to calculate genetic distances, we tried to use some simple and easy-to-understand methods based on the commonly used GenBank sequence database and BLAST program to look into the evolution of the new virus. Two quick approximate methods were developed based on the NCBI BLAST analysis results: one to study the relative evolutionary stability of individual genome segments and the other to study the evolutionary linkage of each segment with each of the others. To further understand the origin of the new virus, we analyzed the hosts and circulating regions of evolutionarily related influenza viruses. Our study provides some novel data regarding the evolution of the new virus and identifies several source virus strains originally circulating in North America and Eurasia, respectively.

Materials and methods

The S-OIV genomic sequence and BLAST analysis

The sequences of the first nearly complete genome of the emerging influenza virus (A/California/04/2009(H1N1)) were submitted by the United States CDC to GenBank on April 27, 2009, with accession numbers FJ966079 through FJ966086, representing segments 1 through segment 8, which range from 2280 bp to 838 bp in size (Table 1). On April 28, 2009, 1 day after the first S-OIV sequence was published, we performed BLAST analysis against Gen-Bank database "nr" on the website of the National Center for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/) with the online nucleotide BLAST program BLASTN [12]. The eight segments of the S-OIV strain A/California/04/2009(H1N1) were analyzed individually with default parameters, and one hundred targets for each query were retrieved. Of the retrieved targets, the newly submitted 2009 influenza A(H1N1) viruses were excluded.

Relative evolutionary stability analysis of influenza virus genome segments

The relative evolutionary stability analysis is based on the BLAST analysis results. In the BLAST analysis, every matched target was returned with an identity score representing the similarity of the target to the query sequence. For every BLAST analysis with a particular segment query sequence, the retrieved isolates (excluding the S-OIV 2009) were grouped. The highest identity score minus the lowest identity score results in an "identity score discrepancy", which is normalized by dividing it by the total number of the entries in the group and then multiplying by 10,000. The resulting value is used to calculate the "instability index" by multiplying by a "frequency coefficient" that is derived by dividing the total number of a segment by the total number of PB1 (full-length, non-S-OIV 2009) deposited in NCBI's influenza virus resource (http://www.ncbi.nih.gov/genomes/FLU/FLU.html, see Table 1). For a particular segment, if the instability index is higher than that of another segment, it means the evolutionary stability of the former is relatively lower.

Segment linkage analysis of the influenza virus genome

Like relative stability analysis, segment linkage analysis is also based on the BLAST analysis results. The previous isolates other than the 2009 A(H1N1) viruses retrieved from BLAST analysis of each segment were grouped, and each pair of groups was compared to check for shared entries. The linkage index of each pair of segments was calculated as the number of the shared entries divided by the total entry number of the smaller group, and then multiplied by the average "frequency coefficient" of the two segments (for the calculation of "frequency coefficient", see the stability analysis section above). For any two segments, if the linkage index is higher, it means the two segments are more closely linked in the progress of evolution.

Table 1 BLAST analysis of influenza A(H1N1) virus genome segments

Seg. no. Gene	Query	Length	2009 strains	Previous strains	Identity		Total no. f.l. non-2009 S-OIV[a]
					Lowest (%)	Highest (%)	
1 PB2	FJ966079	2280	7	93	95.03	96.58	1062
2 PB1	FJ966080	2274	7	93	95.60	96.88	1090
3 PA	FJ966081	2151	9	91	93.63	96.10	1083
4 HA	FJ966082	1701	9	91	89.78	95.31	1764
5 NP	FJ966083	1497	7	93	94.46	96.80	1162
6 NA	FJ966084	1410	9	91	87.54	94.19	2028
7 M	FJ966085	972	14	86	94.86	97.72	1530
8 NS	FJ966086	838	7	93	94.64	96.55	1126

[a] Total number of full-length sequences of non-2009 S-OIV

Results

BLAST results and relative stability of S-OIV genome segments

After the S-OIV was identified as a new pandemic influenza strain, a large amount of sequence data on the new virus was deposited into the public sequence databases, making these databases very biased toward S-OIV 2009. In order to study the genetics of the new virus with less effect of the bias, we analyzed the BLAST results obtained at the earliest possible time. For every BLAST analysis with a genome segment of the S-OIV 2009 strain, the retrieved 100 targets included several entries of other S-OIV 2009 strains (Table 1, see Supplementary Materials for details). These S-OIV 2009 sequences were almost identical to the query sequences (with 99.6–100% identity) and obviously represent the same pandemic virus. Thus, these strains were excluded from the subsequent analyses. The previously isolated target strains (non-S-OIV 2009) shared 87.54–97.72% sequence identity with the query sequences (Table 1). Specifically, segment 6 (NA) targets exhibited the lowest sequence similarity (87.54–94.19%), and segment 4 (HA) showed the second-lowest similarity (89.78–95.31%), while the other segments had much higher sequence identity scores (Table 1). Segment stability analysis (see Materials and Methods) also demonstrated that segments 6 (NA) and 4 (HA) had the highest relative instability indexes, while the other segments had much lower instability indexes, which was in accordance with the abovementioned sequence similarity data (Fig. 1; Table 1). These results indicate that segments 6 (NA) and 4 (HA) are more variable than the other segments in influenza A(H1N1) virus evolution.

Host analysis of influenza A(H1N1) virus

In order to obtain more information about the host of the new virus in the context of evolution, we looked into the
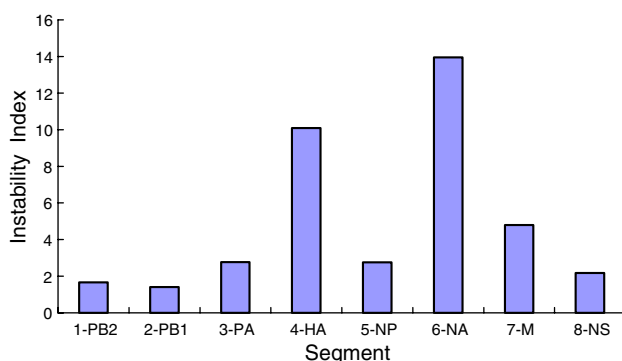

Fig. 1 Relative instability of influenza virus segments

host distributions of the BLAST targets. For all 8 segments except segment 2 (PB1), the best-matched target was isolated from a swine host (Table 2, see Supplementary Materials for details). In the case of segment 2 (PB1), the best-matched target was isolated from "a Wisconsin man infected by a swine-like influenza virus" (GenBank accession number AF342823), which means this strain is actually a swine strain that accidentally infected a human being. Furthermore, the second-best match of segment 2 (PB1) is also a swine virus, and there is only a small difference between the identity scores of the human-hosted best target (2203/2274 = 96.88%) and the swine-hosted second-best target (2195/2275 = 96.48%). Finally, of the top 25 BLAST targets of segment 2, 20 sequences (80%) were isolated from swine hosts. These data suggest that all of the segments, including segment 2, actually have the highest homology with swine influenza viruses, which means the most likely source of the new virus may be swine viruses.

As in the case of segment 2, the best BLAST match may not necessarily be the closest evolutionary source of the segment, due to the tiny differences in identity scores of the top BLAST targets. In order to obtain a relatively safe estimate, we further analyzed the host distributions of the top 25 BLAST targets (non-S-OIV 2009) of each segment and examined the host species frequencies. As shown in Table 2 (see Supplementary Materials for details), for the majority of the segments (segments 4, 5, 6, 7, and 8), more than 90% of the target sequences come from swine viruses. The lowest frequency of swine host (of segment 1 and 2) is 80%, which is still much higher than the frequencies of avian host and human host, which are not greater than 20%. These results further support the hypothesis that all of the genome segments of the novel influenza virus may be derived from swine influenza viruses.

In attempting to find further clues to the origin of the new virus, we analyzed the geographic distributions of the top 25 BLAST targets. The analysis revealed two clusters of segments. The first cluster contains segment 1, 2, 3, 4, 5, and 8, of which the majority of targets are from North America, and the rest from Asia, with no targets from Europe or other continents except for segment 2, of which only one target comes from Europe (Table 2, see Supplementary Materials for details). In this cluster, segments 1 and 8 have the highest percentages of North America distribution, which are 76% (19/25) and 72% (18/25), respectively. Segment 5 has the lowest North America distribution percentage, which is 52%. The second cluster contains segments 6 and 7, for which all the BLAST targets are from Europe and Asia. Specifically, 16 (64%) of the top 25 targets of segment 7 are from Europe and 9 from Asia, while for segment 6, all of the top 25 targets are from Europe. The circulating region analysis combined with the

Table 2 Host and geographic distribution of the top 25 BLAST targets

Seg. no. gene	Best-matched target	Host	North America	Europe	Asia	Total number
1 PB2	A/swine/Korea/JNS06/2004(H3N2)	Swine	14		6	20 (80%)
		Avian	5			5 (20%)
		Human				
		Subtotal	19 (76%)		6 (24%)	
2 pB1	A/Wisconsin/10/98(H1N1)	Swine	13		7	20 (80%)
		Avian	2			2 (8%)
		Human	1	1	1	3 (12%)
		Subtotal	16 (64%)	1	8	
3 PA	A/Swine/Illinois/100084/01(H1N2)	Swine	11		10	21 (84%)
		Avian	4			4 (16%)
		Human				
		Subtotal	15 (60%)		10 (40%)	
4 HA	A/Swine/Indiana/P12439/00(H1N2)	Swine	17		7	24 (96%)
		Avian	1			1 (4%)
		Human				
		Subtotal	18 (72)		28 (28%)	
5 NP	A/Swine/Iowa/533/99(H3N2)	Swine	11		12	23 (92%)
		Avian	1			1 (4%)
		Human	1			1 (4%)
		Subtotal	13 (52%)		12 (48%)	
6 NA	A/Swine/England/195852/92(H1N1)	Swine		24		24 (96%)
		Avian		1		1 (4%)
		Human				
		Subtotal		25 (100%)		
7 M	A/swine/Hong Kong/5212/99(H3N2)	Swine		16	8	24 (96%)
		Avian				
		Human			1	1 (4%)
		Subtotal		16 (64%)	9 (36%)	
8 NS	A/Swine/Minnesota/9088-2/98(H3N2)	Swine	17		6	23 (92%)
		Avian	2			2 (8%)
		Human				
		Subtotal	19 (76%)		6 (24%)	

above host analysis implies that the new virus may originate from at least one North American swine strain and one Eurasian swine strain.

S-OIV are linked with North American and Eurasian swine viruses

Influenza virus genome segments reassortment is a process of homologous segment exchange between two influenza viruses simultaneously infecting a single host [7]. The combination of a set of 8 segments in a certain influenza virus is the result of mutation and selection and comprises numerous adaptive mutations distributed on different segments. In order to understand the natural history of influenza A(H1N1) evolution, we analyzed the linkage of

different segments by looking for shared virus entries among different segment groups. As shown in Table 1, every segment group contains 86–93 non-S-OIV 2009 entries. It can be imagined that the more entries are shared by two segment groups, the more closely the two segments are evolutionarily linked, since these two segments have a greater chance or propensity to co-exist after a reassortment event. Our linkage analysis of all the retrieved non-S-OIV 2009 strains revealed two linkage clusters of segment groups. The first cluster includes groups of segments 1, 2, 3, 4, 5, and 8, while the second cluster is composed of segment 6 and 7 groups. In the first cluster, segments 1, 3, 5, and 8 have very close mutual linkage (linkage indexes >0.50; Table 3). Segments 2 and 4 are moderately linked with each other and with segments 1, 3, 5 and 8, with the

Table 3 Pairwise linkage of segments of influenza A(H1N1) virus

Seg. no. Gene	1-PB2		2-PB1		3-PA		4-HA		5-NP		6-NA		7-M	
	Count	Index	Count	Index	Count	Index	Count	Index	Count	Index	Count	Index	Count	Index
2 PB1	*25*	*0.27*												
3 PA	*63*	*0.70*	*26*	*0.30*										
4 HA	*27*	*0.40*	*13*	*0.16*	*26*	*0.39*								
5 NP	*56*	*0.63*	*28*	*0.31*	*57*	*0.67*	*43*	*0.65*						
6 NA	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00				
7 M	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	*25*	*0.49*		
8 NS	*47*	*0.53*	*24*	*0.27*	*51*	*0.58*	*34*	*0.50*	*58*	*0.67*	0	0.00	0	0.00

Italicized value indicates the existence of evolutionary linkage

linkage index varying from 0.20 to 0.50. In the second cluster, the only two segments, segments 6 and 7, are also moderately linked, with a linkage index of 0.49. Interestingly, there is no entry shared by the two clusters, so the linkage indexes between any one group in the first cluster and any one group in the second cluster are zero. That means the segments of the two clusters evolved separately. These linkage analysis results are consistent with the results of geographic distribution analysis regarding the clustering of the segments and again suggest the possibility of S-OIV coming from two virus strains, one from North America, and the other from Eurasia.

Influenza A(H1N1) virus evolution and possible progenitor viruses

The linkage analysis described above suggested that S-OIV could arise from reassortment of two swine strains: one North American strain contributing segments 1, 2, 3, 4, 5, and 8 and one Eurasian strain providing segments 6 and 7. If this is the case, there should be strains containing segments that are highly homologous to S-OIV across all target groups in a linkage cluster, especially for the first cluster, which consists of 6 segments. To test this hypothesis, we searched for target strains containing all 6 linked segments in the first linkage cluster. Analysis of all the returned targets revealed that 12 target strains were shared across all 6 segment groups, and all strains were of swine origin (Table 4). If a certain strain is going to evolve into S-OIV, all of its segments should have identity scores very close to that of the best match. To find if such strains exist, we checked the sequence identity scores of individual segments of the shared strains. Again, we found that 6 of the 12 shared target strains in the first cluster had very high identity scores for all 6 segments (Tables 4, 5), and all 6 of the highly homologous strains were from North America, with two strains, A/Swine/Illinois/100084/01(H1N2) and A/Swine/North Carolina/93523/01(H1N2), included in all the top 25 targets across the first cluster. Similarly, analysis

of the second cluster revealed 25 strains shared by the segment 6 and 7 groups, and 5 of them demonstrated high sequence identity scores for both segments (Tables 4, 5). All 5 of the highly homologous strains were from Eurasia and were among the top 25 targets of both segment 6 and 7, with two strains A/swine/England/WVL7/1992(H1N1) and A/swine/Spain/WVL6/1991(H1N1), as the best matches according to the sequence similarity. These results suggest that the two abovementioned North American strains and the two European strains are likely the ancestor strains of the new influenza A(H1N1) virus emerging in 2009.

Naturally occurring reassortments resulting in new circulating strains are not very frequent, and only a few reassortments have successfully generated new major circulating strains over a period of nearly one hundred years [5]. The chance of a single reassortment event involving only two viruses would be much greater than a multiple-reassortment event, which would require that more than two viruses infect a single host animal and contribute genome segments. Based on this reasoning and on our findings, we propose that the novel S-OIV is a reassortant of two strains that evolved from the abovementioned candidate North American and European strains. Since these North American candidate strains were already triple-reassortant viruses [13], S-OIV is thus composed of at least four sources of genetic material. This may be the reason why the current epidemic virus is regarded as a multiple reassortant virus.

North American wild duck viruses are the closest known relatives of S-OIV 2009

As the new S-OIV is rapidly spreading around the globe, interest in this virus has stimulated intensive surveillance and study of influenza viruses in human and other animals. Recently, a large amount of influenza virus sequence data has been submitted to public databases. Using our linkage analysis of recent BLAST results, we were surprised to find that, of the six North American segments of the emerging

Table 4 BLAST targets shared across the segment linkage clusters

Cluster 1 (Segment 1, 2, 3, 4, 5, and 8)	Cluster 2 (Segment 6 and 7)	
[b]A/Swine/Illinois/100084/01/H1N2/	A/swine/Belgium/1/1998/H1N1/	A/swine/England/WVL16/1998/H1N1/
[a]A/Swine/Illinois/100085A/01/H1N2/	A/swine/Belgium/WVL2/1983/H1N1/	[b]A/swine/England/WVL7/1992/H1N1/
[a]A/Swine/Indiana/9K035/99/H1N2/	A/swine/Belgium/WVL5/1989/H1N1/	A/swine/France/WVL3/1984/H1N1/
[a]A/Swine/Indiana/P12439/00/H1N2/	A/swine/Chachoengsao/NIAH587/2005/H1N1/	A/swine/France/WVL4/1985/H1N1/
A/Swine/Iowa/930/01/H1N2/	A/swine/Chonburi/NIAH589/2005/H1N1/	A/swine/France/WVL8/1992/H1N1/
A/swine/Korea/CAN01/2004/H1N1/	A/swine/Chonburi/NIAH9469/2004/H1N1/	A/swine/Ratchaburi/NIAH1481/2000/H1N1/
A/swine/Korea/CAS08/2005/H1N1/	A/swine/Chonburi/NIAH977/2004/H1N1/	A/swine/Ratchaburi/NIAH550/2003/H1N1/
A/Swine/Korea/CY02/02/H1N2/	A/swine/CotesdArmor/1482/99/H1N1/	A/swine/Scotland/WVL17/1999/H1N1/
A/Swine/Minnesota/55551/00/H1N2/	[a]A/swine/Denmark/WVL9/1993/H1N1/	[b]A/swine/Spain/WVL6/1991/H1N1/
[b]A/Swine/North Carolina/93523/01/H1N2/	[a]A/swine/England/WVL10/1993/H1N1/	A/swine/Virginia/670/1987/H1N1/
A/Swine/North Carolina/98225/01/H1N2/	[a]A/swine/England/WVL11/1994/H1N1/	A/swine/Virginia/671/1987/H1N1/
[a]A/Swine/Ohio/891/01/H1N2/	A/swine/England/WVL14/1996/H1N1/	A/swine/Zhejiang/1/2007/H1N1/
	A/swine/England/WVL15/1997/H1N1/	

[a] Strains with relatively high sequence identity scores

[b] Candidate progenitor strains with the highest sequence identity scores

Table 5 Candidate progenitor strains with high sequence identity scores

	Sequence identity score (%)							
	1-PB2	2-PB1	3-PA	4-HA	5-NA	6-NA	7-MP	8-NS
Highest identity	96.58	96.88	96.10	95.31	96.80	94.24	97.72	96.55
A/Swine/Illinois/100084/01/H1N2/	96.01	96.18	96.10	94.49	96.26			95.84
A/Swine/Illinois/100085A/01/H1N2/	96.10	96.31	95.64	94.66	96.26			95.48
A/Swine/Indiana/9K035/99/H1N2/	96.19	96.48	95.31	95.19	96.40			96.19
A/Swine/Indiana/P12439/00/H1N2/	96.49	96.18	95.82	95.31	96.66			95.12
A/Swine/North Carolina/93523/01/H1N2/	96.19	95.96	95.87	93.61	95.99			95.84
A/Swine/Ohio/891/01/H1N2/	96.10	96.26	95.18	94.89	96.53			95.84
A/swine/Denmark/WVL9/1993/H1N1/						93.47	96.51	
A/swine/England/WVL10/1993/H1N1/						93.89	96.51	
A/swine/England/WVL11/1994/H1N1/						93.61	96.21	
A/swine/England/WVL7/1992/H1N1/						94.24	96.21	
A/swine/Spain/WVL6/1991/H1N1/						94.17	96.40	

S-OIV, five (segment 1, 2, 3, 5, and 8) show closest identity to North American avian strains (A/pintail duck/South Dakota/Sg-00126/2007, A/mallard duck/South Dakota/Sg-00125/2007, A/mallard duck/South Dakota/Sg-00128/2007, A/mallard duck/South Dakota/Sg-00127/2007) (Table 6). These strains were all isolated from wild birds in South Dakota in 2007 and were presumed to result from cross-species transmission from pigs [8]. Since the four strains were isolated from different duck species in South Dakota, it is most probable that these swine-origin H3N2 viruses had settled in water birds in North America. The linkage of the best hits of five segments in the same circulating strains cannot be a coincidence and suggests that these strains circulating in birds are possible ancestors of S-OIV 2009. Based on the currently available sequence data, these wild duck virus strains should be the closest known relatives of the S-OIV emerging in 2009, which indicates that S-OIV 2009 could have come from wild birds. This hypothesis differs from the current mainstream opinion that the emerging virus are from pigs. Further surveillance of influenza in birds, especially in North American water birds, will probably reveal more interesting results regarding the evolutionary history of the 2009 pandemic virus.

Discussion

The source of the 2009 influenza A(H1N1) virus is of concern internationally, and the answer is still uncertain. Various media report that the new virus is a very unusual

Table 6 South Dakota avian strains with the highest sequence identity scores

Strain	Sequence identity scores (%)				
	1-PB2	2-PB1	3-PA	5-NA	8-NS
A/mallard duck/South Dakota/Sg-00125/2007(H3N2)	96.97	96.59	96.36	96.99	96.55
A/pintail duck/South Dakota/Sg-00126/2007(H3N2)	96.97	96.62	96.38	96.99	96.43
A/mallard duck/South Dakota/Sg-00127/2007(H3N2)	96.93	96.57	96.28	96.99	96.55
A/mallard duck/South Dakota/Sg-00128/2007(H3N2)	96.93	96.59	96.38	97.06	96.55

mixture resulting from multiple reassortments by a human virus, a bird virus and two swine viruses. Our study demonstrates that all 8 genome segments of this newly emerging influenza virus are very closely related to those of swine influenza viruses. Linkage analysis revealed two major linkage groups: one group containing segments 1, 2, 3, 4, 5, and 8, dominated by North American swine flu viruses, and the other containing segments 6 and 7, dominated by Eurasian swine flu viruses. Furthermore, two North American swine virus strains were identified with highly homologous segments 1, 2, 3, 4, 5, and 8, and similarly, two Eurasian swine virus strains were found with highly homologous segments 6 and 7. These data suggest that the newly emerging influenza virus could have resulted from a single reassortment of two swine influenza viruses, one North American strain and one Eurasian strain, which is in accordance with recent publications [2, 3, 9, 10]. The reason for the belief that the 2009 influenza A(H1N1) resulted from multiple reassortment of human, bird and swine viruses is probably that the North American source strain is already a triple-reassortant virus that contains genetic material from human, bird and swine influenza viruses [13].

The highest sequence identity scores of the best-matched strains vary from 94 to 97%, which suggests that intermediate strains with higher sequence identity scores may exist. These intermediate strains are evolutionary links to the emerging virus. The reason for not finding these links may be insufficient surveillance or the fact that the reassortment is relatively new and the resultant virus evolved so fast that surveillance could not be conducted before it caused an outbreak. It is also possible that pigs passed the virus to another animal host that was not under current surveillance or that the new flu virus resulted from reassortments in a laboratory.

Highly pathogenic influenza viruses pose a great threat to human health. Due to the existence of numerous influenza virus subtypes and their wide host range, as well as the possibility of cross-species transmission, the evolution of new reassortants is usually very complicated and difficult to track. Sophisticated analysis tools with elusive algorithms can be used by researchers with specialized knowledge. As an alternative, our simple and quick approach with an easy-to-understand strategy can be applied to most research settings and adopted by researchers to draw approximate conclusions. With our approach, it is also possible to reveal some interesting characteristics of virus evolution, such as the relative segment stability and linkage propensity, which are not obtained with other established approaches.

References

1. CDC (2009) Swine Influenza A (H1N1) infection in two children—Southern California, March-April 2009. MMWR Morb Mortal Wkly Rep 58:400–402
2. Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, Garten RJ, Gubareva LV, Xu X, Bridges CB, Uyeki TM (2009) Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med 360:2605–2615
3. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, Okomo-Adhiambo M, Gubareva L, Barnes J, Smith CB, Emery SL, Hillman MJ, Rivailler P, Smagala J, de Graaf M, Burke DF, Fouchier RA, Pappas C, Alpuche-Aranda CM, Lopez-Gatell H, Olivera H, Lopez I, Myers CA, Faix D, Blair PJ, Yu C, Keene KM, Dotson PD Jr, Boxrud D, Sambol AR, Abid SH, St George K, Bannerman T, Moore AL, Stringer DJ, Blevins P, Demmler-Harrison GJ, Ginsberg M, Kriner P, Waterman S, Smole S, Guevara HF, Belongia EA, Clark PA, Beatrice ST, Donis R, Katz J, Finelli L, Bridges CB, Shaw M, Jernigan DB, Uyeki TM, Smith DJ, Klimov AI, Cox NJ (2009) Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 325:197–201
4. MacKenzie D (2009) Deadly new flu virus in US and Mexico may go pandemic. New Scientist. http://www.webcitation.org/5gNd9HTT9
5. Nelson MI, Viboud C, Simonsen L, Bennett RT, Griesemer SB, St George K, Taylor J, Spiro DJ, Sengamalay NA, Ghedin E, Taubenberger JK, Holmes EC (2008) Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. PLoS Pathog 4:e1000012

6. Peiris JS, Poon LL, Guan Y (2009) Emergence of a novel swine-origin influenza A virus (S-OIV) H1N1 virus in humans. J Clin Virol 45:169–173

7. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC (2008) The genomic and epidemiological dynamics of human influenza A virus. Nature 453:615–619

8. Recombinomics (2008) Swine Flu Sequences in Wild Birds in South Dakota. Recombinomics Commentary 23:42. http://www.recombinomics.com/News/06170803/Swine_WB_SD.html

9. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghwani J, Bhatt S, Peiris JS, Guan Y, Rambaut A (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 459:1122–1125

10. Trifonov V, Khiabanian H, Greenbaum B, Rabadan R (2009) The origin of the recent swine influenza A(H1N1) virus infecting humans. Euro Surveill 14. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=&dopt=Citation&list_uids=19422769

11. WHO (2009) Current WHO phase of pandemic alert. http://www.who.int/csr/disease/avian_influenza/phase/en/index.html

12. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214

13. Zhou NN, Senne DA, Landgraf JS, Swenson SL, Erickson G, Rossow K, Liu L, Yoon K, Krauss S, Webster RG (1999) Genetic reassortment of avian, swine, and human influenza A viruses in American pigs. J Virol 73:8851–8856