VERSITA

Biologia **67**/1: 14—31, 2012
Section Cellular and Molecular Biology
DOI: 10.2478/s11756-011-0163-6

Phylogenetic analysis of pandemic influenza A/H1N1 virus

Ramaiah Arunachalam, Kanniah Paulkumar & Gurusamy Annadurai*

Environmental Nanotechnology Division, Sri Paramakalyani Centre for Environmental Sciences, Manonmaniam Sundaranar University, Alwarkurichi – 627412, Tamil Nadu, India; e-mail: gannadurai@hotmail.com

Abstract: The principle of the present study was to determine the evolution of pandemic novel influenza A/H1N1 2009 virus (NIV) by phylogenetic, comparative and statistical analyses. The phylogenetic trees of eight genomic segments illustrate that, so far, the sequences of the NIVs (outbreak group A) are relatively homogeneous and derived by the event of multiple genetic reassortment of Eurasian and North American swine, avian and human viruses (group B). It implies that some of the influenza viruses in group B had higher potential to evolve and getting the ability to transmit from human-to-human after animal-to-human cross-species transmission. The second analysis shows that NIV had attempted a little evolutionary change among humans and before introduction into human it had long evolutionary history. Statistical analysis shows that viruses from both outbreak and nearest group have homologous genes in their genomes which might be reflecting the phylogenetic relationship of strains, and also the presence of unique mutations between groups A-B may associate with increased virulence of NIVs. Both phylogenetic and cluster analyses confirm that the gene exchange takes place between viruses originated from different species and it could be generated NIV with unpredictable pandemic potential. Hence, we conclude that an extensive study should be made to recognize, which reassortment groups are closely related to NIVs, and to determine the sites in the genes of NIV under greatest or least selection pressure, which will ultimately be important in the effective design of vaccine and drugs for 'swine flu'.

Key words: pandemic 2009 NIV; genomic sequence; phylogeny; reassortment; evolution.

Abbreviations: GD, genetic distance; HA, haemagglutinin; MCL, maximum composite likelihood MP, matrix protein; NA, neuraminidase; NIV, novel influenza A/H1N1 2009 virus; NJ, neighbour joining; NP, nucleoprotein; NS, non-structural protein; PA, RNA polymerase; PB1, catalytic RNA polymerase; PB2, cap-binding RNA polymerase.

Introduction

Influenza A virus is one of the most significant causes of annual morbidity and mortality in humans (Palese 2004) and has largest databases of sequenced genes (Kryazhimskiy et al. 2008). Influenza A virus is found in human as well as in other animals including swine, horses, sea mammals, and birds, of which waterfowl are considered the natural reservoir (Webster et al. 1992; Tamuri et al. 2009). Subtypes of influenza A are distinguished by two surface glycoproteins: haemagglutinin (HA), the primary target of the immune response, and neuraminidase (NA). There are sixteen known types of HA (H1 to H16) and nine of NA (N1 to N9); all found in waterfowl. Only H1, H2, H3, H5 (Hatta et al. 2007; Gambotto et al. 2008; Li et al. 2010) and N1, N2, however, are known to have caused epidemic disease in humans. The predominant forms of influenza A/H1N1 and H3N2 are currently circulating in humans.

There are two distinct problems represented by influenza. Firstly, the various subtypes currently circulating in humans cause significant morbidity and loss of life. Secondly, periodically a subtype of influenza can make the shift from aquatic birds to humans, possibly through an intermediate host like pigs, resulting in a widespread of pandemic in an immunologically-naive population. These 'antigenic shifts' can occur either through the transfer of an entire virus from one host to another, or through a reassortment process where the avian and swine viruses mixing their genomic segments (Tamuri et al. 2009).

In 1957, three virus segments (HA, NA, and catalytic RNA polymerase PB1) from an avian-like source were combined with the other five segments already circulating in humans to create the H2N2 'Asian flu' pandemic, while in 1968, two segments (HA and PB1) from an avian-like source were combined with the other six from the already exited human H2N2 virus to form the H3N2 'Hong Kong flu' pandemic (Schafer et al. 1993). It has been suggested that the 1918 H1N1 'Spanish flu' virus was the result of a single host-shift event from birds to humans (Reid et al. 2004; Taubenberger et al. 2005; Taubenberger 2006) but this remains controversial (Antonovics et al. 2006; Gibbs & Gibbs 2006; Reis et al. 2009; Smith et al. 2009a; Tamuri et al. 2009). It is noted that the genomes of the aforementioned pan-

* Corresponding author

demic influenza viruses all originated in entire or in part from non-human reservoirs (Garten et al. 2009).

In 21st century, the WHO announced on 29 April 2009 that the rapid global spread of a strain of novel influenza A/H1N1 2009 virus (NIV) was detected and was found to be a global pandemic alert level to phase 5 (http://www.who.int/csr/disease/swineflu/) which indicated sustained human-to-human transmission of NIV in one WHO region of the world, and exported cases detected in other regions (Fraser et al. 2009). As of 18 May 2009, there had been 8,829 laboratory confirmed cases in 40 countries, resulting in 74 deaths (Garten et al. 2009) and WHO announced that the global pandemic alert level to phase 6 on 11 June 2009 (Qu et al. 2011). This NIV has been widely spread over 170 countries from April to June 2009, and has the ability to persist in the human population, potentially with more severe clinical consequences (Munster et al. 2009; Qu et al. 2011).

Most of the influenza A viruses from avian can result in sporadic human infection via animal-to-human transmission, but lack the ability of human-to-human transmission (Ding et al. 2009; Webby & Webster 2001). However, the NIV not only accomplishes the cross-species transmission from its original hosts to humans, but it also gains the ability to spread efficiently among humans causing a human pandemic (Scholtissek 1990; Webby & Webster 2001; CDC 2009a,b; Babakir-Mina et al. 2009; Ding et al. 2009; Van-Reeth & Nicoll 2009). As part of widespread surveillance effort, it is important to understand the genesis of each genomic segment of NIV. Therefore, the purpose of the present study was to discover the evolution of NIV by phylogenetic and comparative analyses.

Material and methods

Phylogenetic analyses

The eight protein-coding genomic segments of the NIV were downloaded from the NCBI Influenza Virus Resource (http://www.ncbi.nlm.nih.gov/genomes/FLU/SwineFlu.html) in April 2010 and the existing sequences of A/H1N1, H2N2, H3N2 and H5N1 viruses sampled from human, avian and swine around the world during the period of 1918–2010 were also retrieved from GenBank (Benson et al. 2011). For each genome segment 74, 75, 73, 95, 74, 93, 86 and 78 sequences haphazardly were used for phylogenetic analysis for cap-binding RNA polymerase (PB2), PB1, RNA polymerase (PA), HA, nucleoprotein (NP), NA, matrix protein (MP) and non-structural protein (NS), respectively. It is noted that all the analyzed sequences were partial (if a complete genome is not available in the databank) as well as complete genomes. MEGA 4.1 software (Tamura et al. 2007) was started with a set of aligned sequences of gene segments of the NIVs with other influenza A viruses using Clustal W (Thompson et al. 1994), and searches for phylogenetic trees (radial view) that are optimal according to neighbour-joining (NJ) algorithms using p distance model under complete deletion option. It should be noted that the p distance is known to produce reliable phylogenetic trees, when a large number of closely related sequences is analyzed (Nei & Kumar 2000; Suzuki 2006).

Tamura et al. (2004) showed that pair-wise distances and the related substitution parameters are accurately estimated by maximizing the composite likelihood. They also found that, unlike the cases of ordinary independent estimation of each pair-wise distance, a complicated model had virtually no disadvantage in the composite likelihood method (CML) for phylogenetic analyses. For that reason, Tamura-Nei model (Tamura & Nei 1993) is available for this method in MEGA 4.1 which was also performed to confirm the NJ trees.

Comparative analyses

The phylogenetic trees of eight genomic segments show that the NIVs (group A) had a distinct shorter branch length to the nearest group B (Fig. 1). These results suggest that the reassortment occurred recently and should appear shorter to (phylo) genetic divergence to their nearest genetic neighbours (Schnitzler & Schnitzler 2009) (group B). For that, we have calculated the genetic distance (GD) between groups A–F by MEGA 4.1 using the NJ algorithm under the p distance model (Tamura et al. 2007). On the basis of the significant results of GD, the groups A and B were considered for further comparative and statistical analysis. Groups A and B were defined as outbreak and non-outbreak group, because of having pandemic 2009 influenza A/H1N1 viruses and classical/seasonal influenza viruses which are not responsible for 2009 pandemic outbreak, respectively.

In the pair-wise comparative analysis, the bottom sequence of outbreak group A (e.g. PB2 – A/California/VRDL73/2009(H1N1)) was used as a query to calculate the sequence similarity using Clustal W2 (Thompson et al. 1994) and GD using MEGA 4.1 with other sequences of group A and B stated in phylogenetic trees. The cluster analysis was carried out for sequence similarity and GD variables of both groups A-B using unweighted pair-group average (UPGMA) algorithm under Euclidean distance model by PAlaeontological STatistics (PAST) 2.05 software (Hammer et al. 2001).

Whether the increased rate of evolution led to the outbreak, or there were long time lags for which the period of ancestors of the current epidemic went unsampled, they were determined by regressing GD and similarity against sampling year for each gene. Therefore, to quantify the period of unsampled diversity, and to estimate the date of origin for the NIV outbreak, we estimated the year mean of groups A and B, and the degree of excess of year could be considered for each gene.

Results and discussion

Phylogenetic analysis

In past few years, both swine influenza and triple reassortant swine influenza viruses have rarely been isolated from human (Webster et al. 1992; Garten et al. 2009), even though these virus infections cause occasionally deaths, only limited human-to-human transmission has previously been documented (Garten et al. 2009). Presently, unaccounted NIV was isolated from humans which contains a mixture of gene segments that previously has not been reported in swine and/or human in worldwide. There is a need of identifying the evolution of each gene segments of NIV that will be important for effective vaccine and drug development.

Numerous circumstances exist about NIVs reassortment and origin; however, it is presently indistinct

where the reassortment events most likely occurred. The detailed study should be made on detecting of the closest ancestral genes for each of eight segments of NIVs that may imply where reassortment events most likely happened. In the present study, in all the phylogenetic trees (Fig. 1a–h), the outbreak group A has well-supported branch with group B, which typically has lineages of swine, avian and human influenza viruses. It implies that some of the influenza viruses in group B has higher potential to evolve and getting the ability to transmit from human-to-human after animal-to-human cross-species transmission (Ding et al. 2009). Our results show that each gene segment of the NIV is nested within well-established swine influenza lineages. Phylogenetic trees indicate that the multiple-reassortant NIV itself consists of genes, which were originated from North-American Avian (PB2 and PA), North-American human H3N2 (PB1), North-American swine (HA, NP and NS) and Eurasian swine lineages (NA and MP) (Fig. 1). In PB2 and PA trees, the group A closely clusters with group B, which contains North-American swine and avian influenza A/H1N1 virus lineages, whereas group B of PB1 had North-American swine and human A/H3N2 virus lineages. Besides, the presence of the North-American swine influenza A/H1N1 virus (highlighted in red colour) and human A/H3N2 viruses within the same cluster (group B) may suggest that the PB2, PB1 and PA genes of North-American avian and human have been circulating among American pigs for decades. Pigs could be infected with wholly avian and/or human viruses, allowing swine viruses to acquire avian and/or human virus gene segments to generate NIVs (Olsen et al. 2003; Ma et al. 2010). Pigs have been considered to be a 'mixing vessel' for human and avian influenza viruses because the epithelial cells in pigs have receptors for both influenza viruses (Ma et al. 2010) and become a reservoir of H1 viruses with the potential to cause major respiratory outbreak or even a possible pandemic in humans (Garten et al. 2009). It undoubtedly indicates that the NIV genesis happened in swine; here the genes PB2 and PA of North-American avian and PB1 of North-American human viruses were transferred from their respective host to pigs for genetic swap that could be occurred competently with already existing genes HA, NP and NS of North American swine and NA and MP of Eurasian swine genetic lineages. Our report demonstrates that all genomic segments of this NIV are very closely related to those of swine influenza viruses. Results of PB2, PB1 and PA genes had swine influenza lineages within the same cluster of group B that makes this study recast. The ultimate outcomes of the present study are more comprehensive that validates the hypotheses which have been postulated by others (Dawood et al. 2009; Ding et al. 2009; Garten et al. 2009; Smith et al. 2009b; Qu et al. 2011). These findings prove the premise that the swine is an important animal reservoir for NIV infection in human populations; the swine flu is believed to be caused by the cross-species transmission (Webster et al. 2006) and the NIV not only accomplishes the cross-species transmission from its original hosts to human but it also gains the ability to spread efficiently among humans (Dawood et al. 2009; Smith et al. 2009b).

The NJ tree of each segments were evaluated by CML trees that show no significance variations in all gene segments except in NP gene tree. NJ method for NP tree shows the strain A/Swine/1931(H1N1)/USA_1931 was positioned in group C, whereas in CML method (figures not shown) the strain was in group B (highlighted in blue colour) with long distance. The most parsimonious interpretation of these results is the progenitor of the NIV epidemic originated in pigs (Smith et al. 2009b).

Potential NIV multiple-reassortment

The PB2 and PA gene segments are in the swine lineage (Fig. 1a,c) that were seeded in swine (swine strain highlighted with red colour) lineage from avian in around 1976 and 1986, respectively (Fig. 2). These avian-to-swine reassortant swine lineages subsequently seeded in North American swine lineages in 2005. The PB1 gene segment is in the swine lineage that was seeded in North-American swine (H1N1) from human (H3N2) lineage in 2003–05, before that this human lineage was derived from avian (H1N1) in around 1976 (Fig. 1b; swine strain highlighted with red colour). The HA, NP and NS gene segments are in the swine lineages that were seeded in North American swine (H1N1) from human (H1N1) lineages (Fig. 1d,e) and avian (H1N1, H2N2 and H3N2) lineages in 1977 and 1931, respectively (Fig. 1h); before that this human (H1N1) lineages was derived from avian (H1N1) lineages in around 1976 and later circulated in classical and triple reassortant swine viruses (Olsen 2002). The NA and MP gene segments in swine lineages were originally derived from an avian (H1N1, H2N2 and H3N2) lineages (Fig. 1f,g) and thought to have entered the Eurasian swine (group B) in around 1976 and 1978, respectively (Fig. 2). The outcome of the present study, in particularly the year and the exact subtypes of seeded lineages for each gene segments validate the assumption comprehensively that have been postulated by others (Garten et al. 2009; Qu et al. 2011). These results strongly suggest that NIVs might have acquired gene segments from different groups of parental influenza viruses, which underwent multiple reassortments and transmissions among swine, avian and human lineages. The reports of Garten et al. (2009), Smith et al. (2009b) and Qu et al. (2011) on HA and NP gene segments are in agreement with the present result. In the context of these findings, with large number of sequences, further study should focus on determining whether HA and NP gene segments could have entered the North American swine through human lineages.

Comparative and statistical analysis

The phylogenetic trees show that the group A of NIV had distinct short branch length to the nearest group B (Fig. 1a–h). Outbreak group A from eight genomic segments could exhibit a short phylogenetic distance to

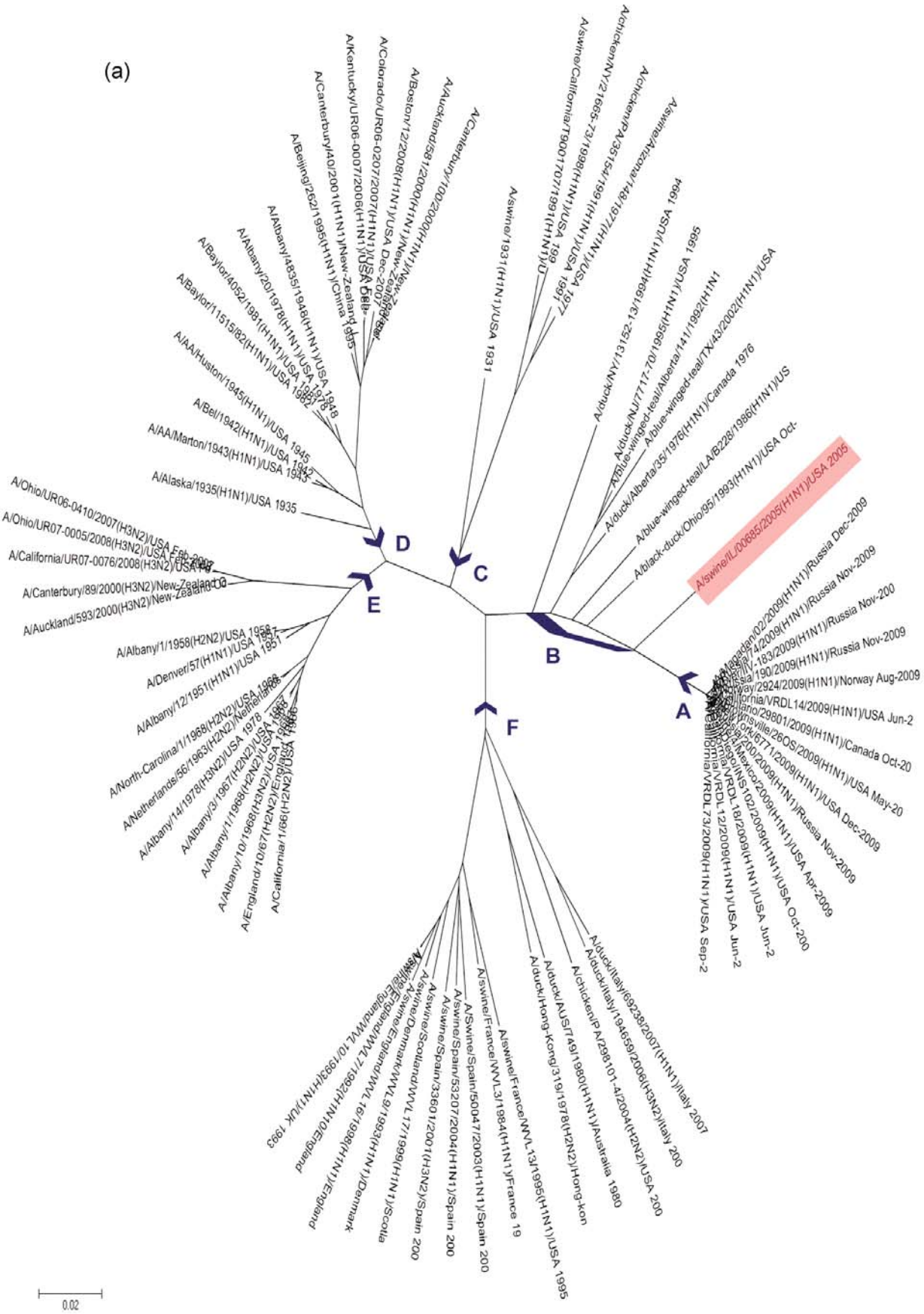

Fig. 1a. Phylogenetic tree of PB2 segment of the novel influenza A/H1N1 virus. The tree was constructed with MEGA 4.1 software by using the NJ method under p-distance model. The reliability of the tree was evaluated by the bootstrap method with 1,000 replications. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). The groups of sequences were labelled from A–F.

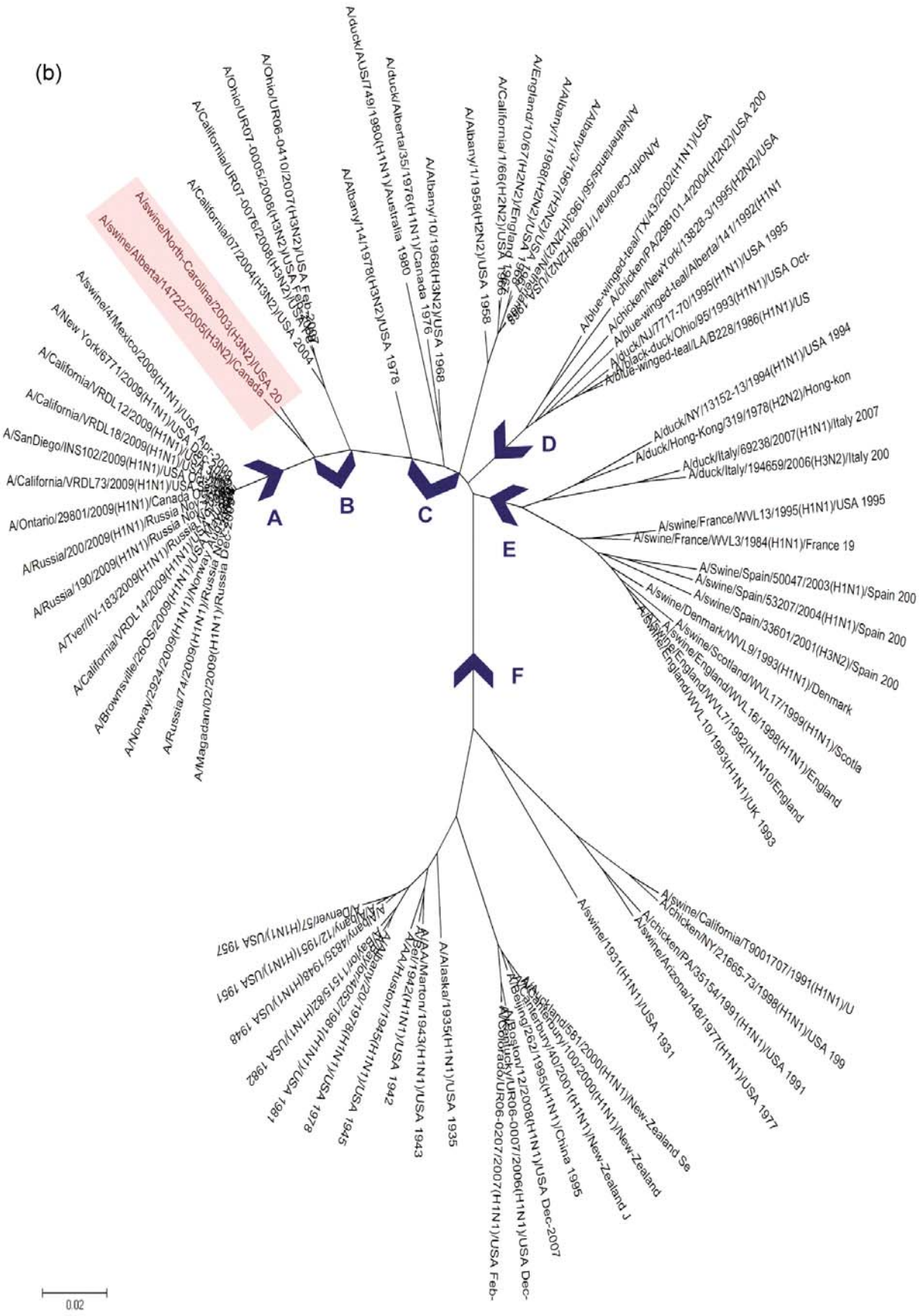

Fig. 1b. Phylogenetic tree of PB1 segment of the novel influenza A/H1N1 virus. The tree was constructed with MEGA 4.1 software by using the NJ method under *p*-distance model. The reliability of the tree was evaluated by the bootstrap method with 1,000 replications. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). The groups of sequences were labelled from A–F.

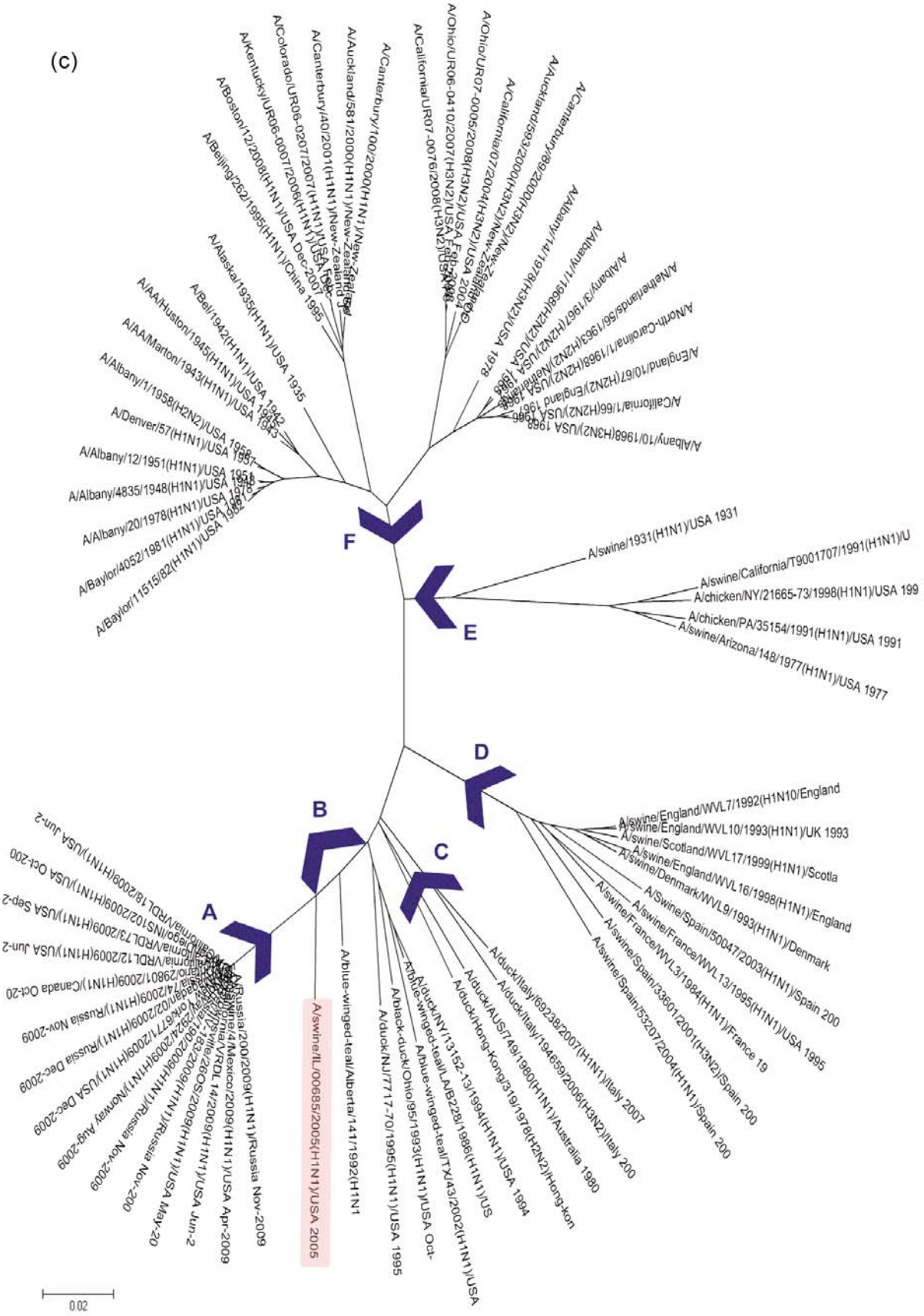

Fig. 1c. Phylogenetic tree of PA segment of the novel influenza A/H1N1 virus. The tree was constructed with MEGA 4.1 software by using the NJ method under *p*-distance model. The reliability of the tree was evaluated by the bootstrap method with 1,000 replications. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). The groups of sequences were labelled from A–F.

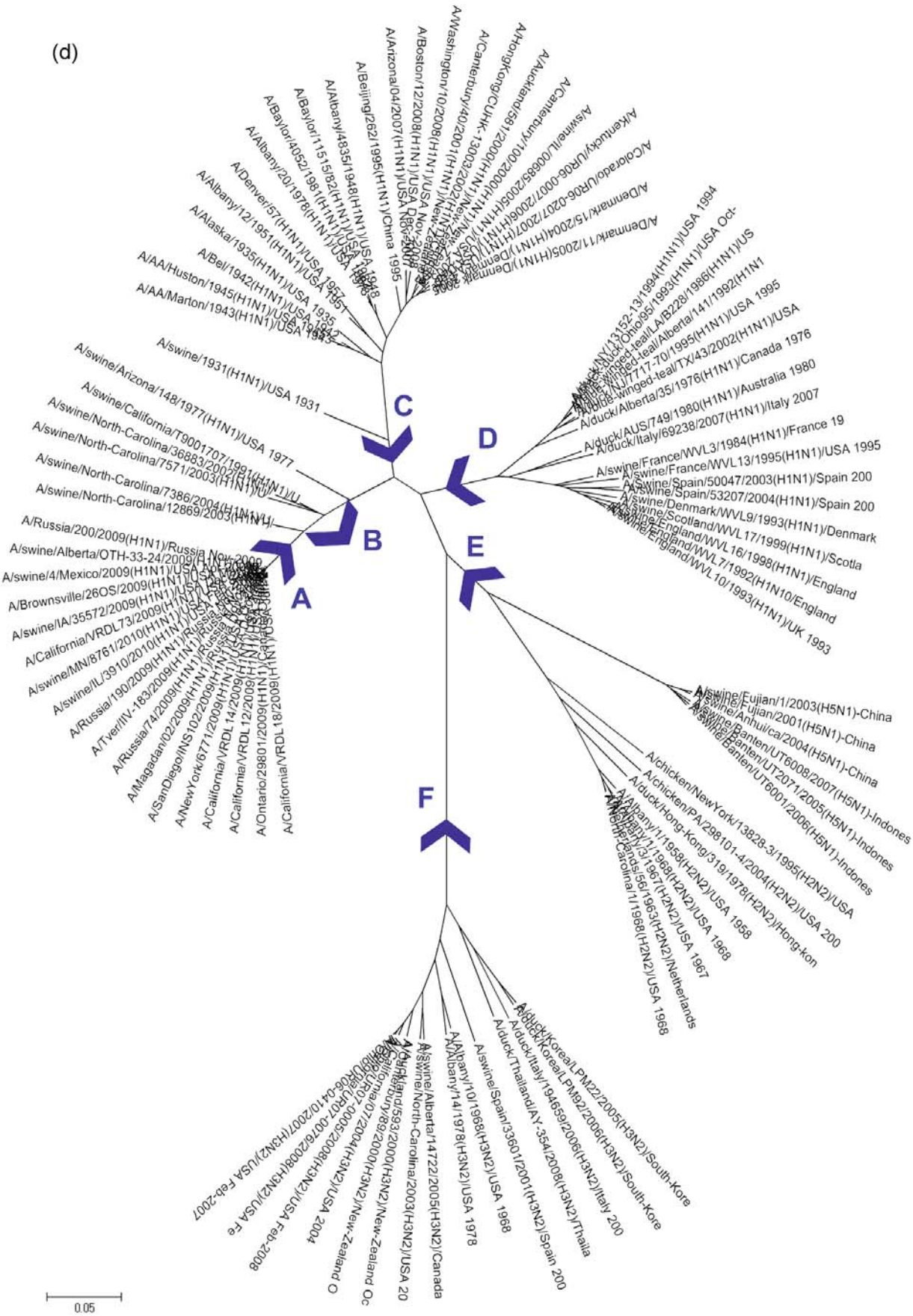

Fig. 1d. Phylogenetic tree of HA segment of the novel influenza A/H1N1 virus. The tree was constructed with MEGA 4.1 software by using the NJ method under *p*-distance model. The reliability of the tree was evaluated by the bootstrap method with 1,000 replications. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). The groups of sequences were labelled from A–F.

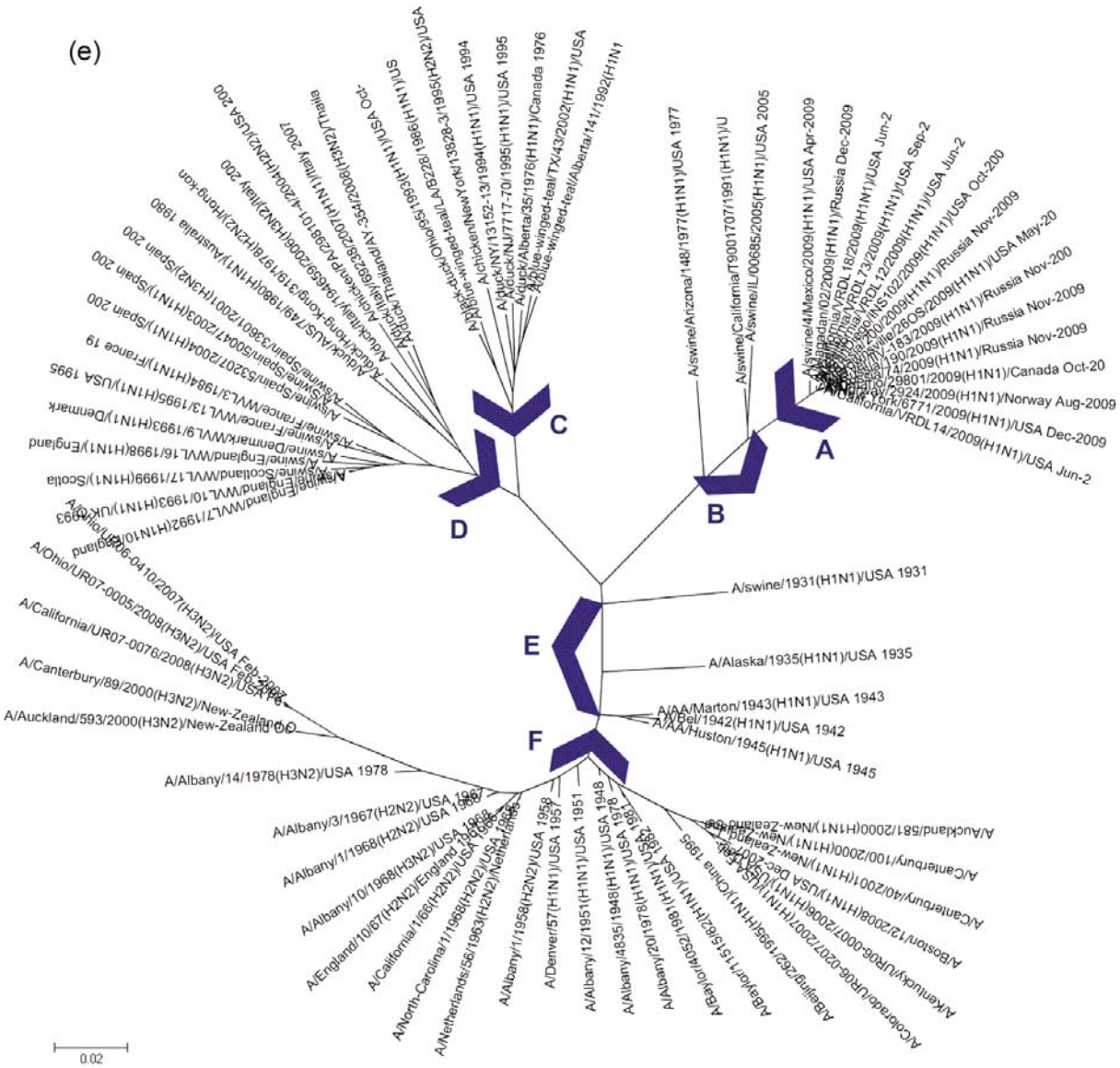

Fig. 1e. Phylogenetic tree of NP segment of the novel influenza A/H1N1 virus. The tree was constructed with MEGA 4.1 software by using the NJ method under *p*-distance model. The reliability of the tree was evaluated by the bootstrap method with 1,000 replications. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). The groups of sequences were labelled from A–F.

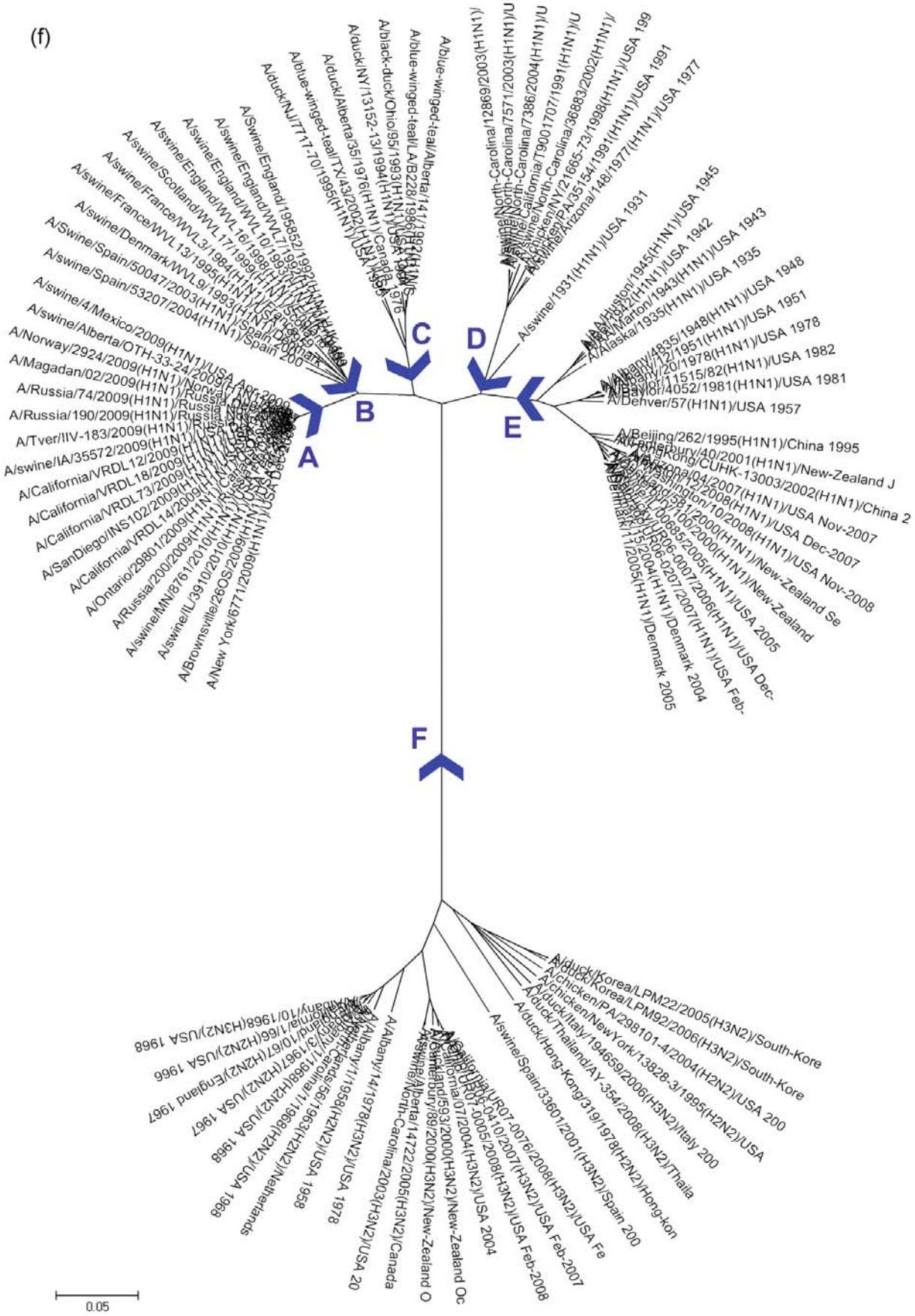

Fig. 1f. Phylogenetic tree of NA segment of the novel influenza A/H1N1 virus. The tree was constructed with MEGA 4.1 software by using the NJ method under *p*-distance model. The reliability of the tree was evaluated by the bootstrap method with 1,000 replications. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). The groups of sequences were labelled from A–F.

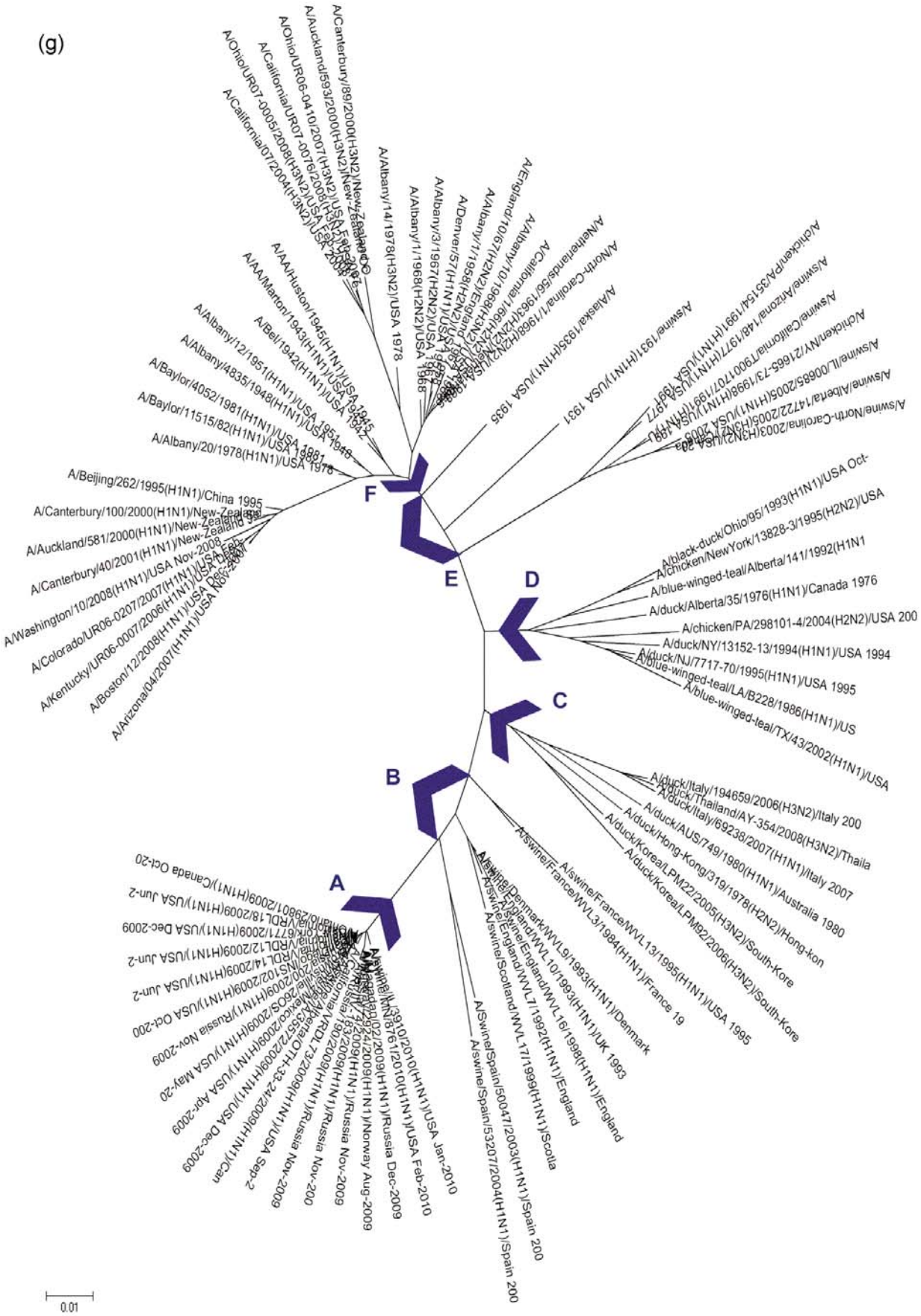

Fig. 1g. Phylogenetic tree of MP segment of the novel influenza A/H1N1 virus. The tree was constructed with MEGA 4.1 software by using the NJ method under *p*-distance model. The reliability of the tree was evaluated by the bootstrap method with 1,000 replications. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). The groups of sequences were labelled from A–F.

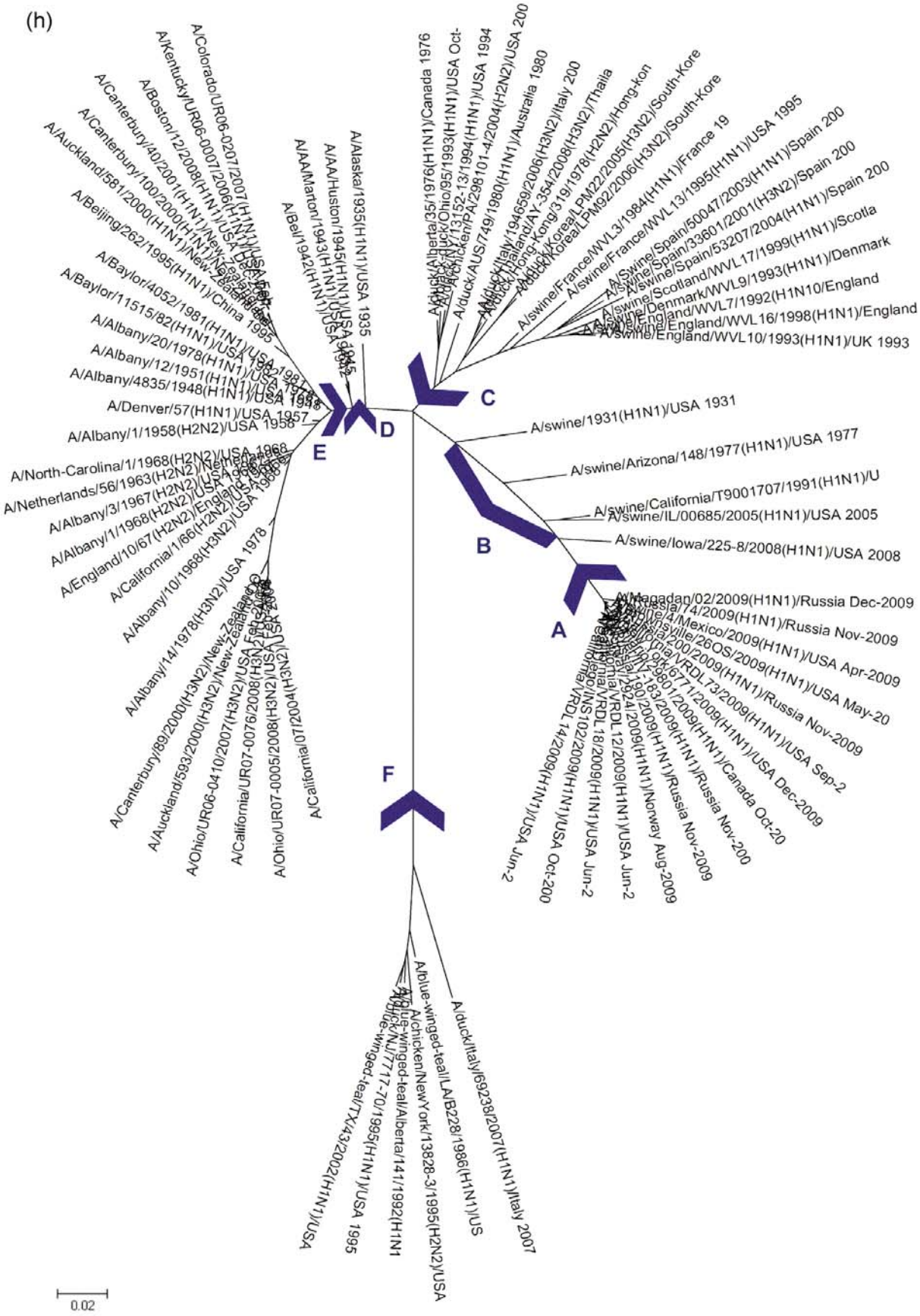

Fig. 1h. Phylogenetic tree of NS segment of the novel influenza A/H1N1 virus. The tree was constructed with MEGA 4.1 software by using the NJ method under *p*-distance model. The reliability of the tree was evaluated by the bootstrap method with 1,000 replications. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). The groups of sequences were labelled from A–F.

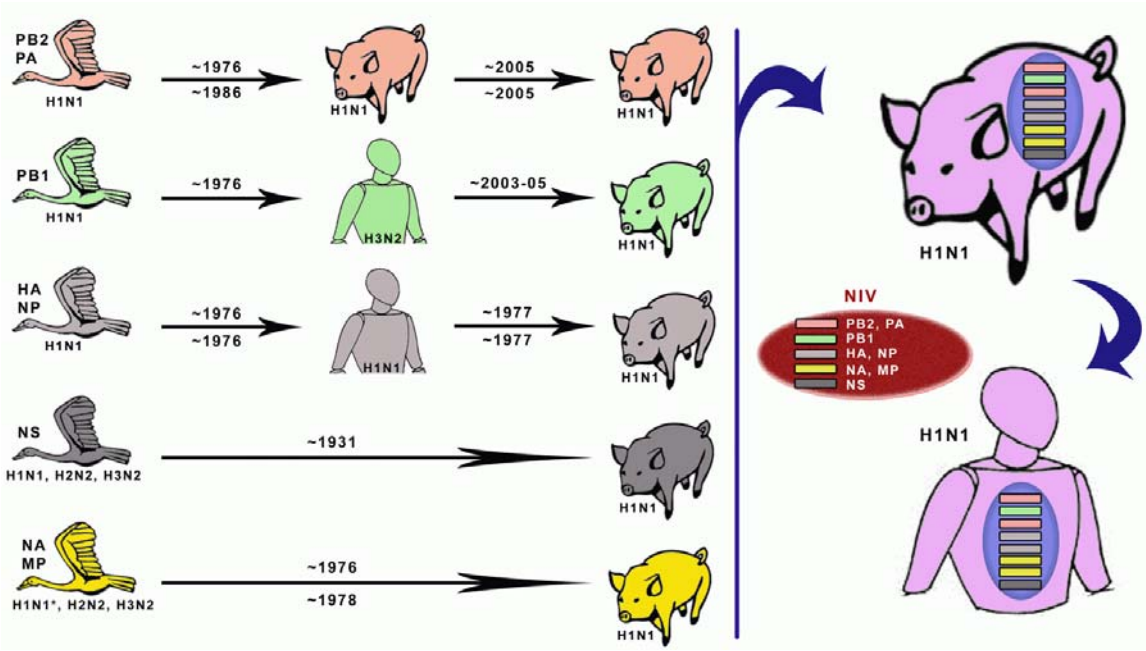

Fig. 2. Occurrence of host and lineage origins for the each gene segments PB2, PB1, PA, HA, NP, NA, MP and NS of the NIV. The gene segments with colour indicate host and lineage origin (see text for details). Each gene segments with respective colours are specified in the red-coloured NIV. *H1N1 is a common subtype appeared in NA and MP segments.

their nearest neighbour group B (Ding et al. 2009) comparable to groups C–F. It indicates the NIVs multiple reassortments occurred very recently; further comparison of GD within and between the NIV group A and other groups B-F shows that the mean of GD (0.002–0.005) interior of the group A is obviously smaller than that of its nearest group B (0.036–0.073) as well as other groups like C–F (0.045–0.080; 0.027–0.109; 0.044–0.197; 0.040–0.118) (Fig. 3a). Here, the GD of interior of the group A is visibly smaller than other groups B–F. These results indicate that NIV has undergone a little evolutionary change among humans, whereas group B–F is indicated comparatively to have a long evolutionary history in swine and avian before the NIV seeded into human.

GD calculated for group A with other groups indicates the number of base differences per site from averaging over all sequence pairs between groups, whereas net GD indicates the number of base differences per site from estimation of net average between groups of sequences (Fig. 3b,c). The results shown in Figure 3b indicate that in eight genomic segments, distance between group A and B ranged between 0.051–0.087, whereas group A with other groups C–F shows 0.088–0.226, 0.102–0.250, 0.124–0.358 and 0.119–0.493, respectively. The net distance calculation was also carried out showing that group A with B ranged between 0.031–0.049, whereas group A with other groups C–F shows 0.056–0.187, 0.077–0.193, 0.095–0.258 and 0.097–0.435, respectively (Fig. 3c). These two results clearly indicate that genetic analyses of NIVs have shown low diversity with group B when compared to other groups C–F, suggesting that NIV introduction in humans is recent and through either a single event or multiple events involving genetically similar viruses. The evolutionary distances between the NIV gene segments and its neighbour relatives indicate a lack of surveillance in swine populations that may harbour influenza viruses with pandemic potential (Garten et al. 2009). These crucial findings may be an additional comprehensive report of Ding et al. (2009) that shows genetic and net genetic distance-based relationships of outbreak virus (group A) with all other influenza viruses (group B–F). Both of these results indicate that the NIV has a short evolutionary history among human and, before introduction into human, it has long evolutionary history, implying that this virus might have been circulating undetected among animal reservoirs somewhere in the world for a relatively long period of time. The genetic reassortment occurs in swine that was demonstrated to play a role as a mixing vessel in the co-infection and the reassortment of various influenza viruses (Ding et al. 2009; Van-Reeth & Nicoll 2009). The ultimate findings of the NIV with swine influenza virus lineages in all eight genomic segments (Fig. 1a–h) strongly suggest that pigs are the most possible animal reservoir, and as a consequence, the NIV 2009 outbreak was caused by the cross-species transmission of NIV from pigs to human after reassortment between multiple swine-origin lineages (Dawood et al. 2009). The estimation of overall mean indicates the number of base differences per site from averaging over all sequence pairs. In eight genomic segment trees, the overall mean was greatly varied (Fig. 3d) due to the variation in number of sequences used.

The groups A and B were considered for further statistical analyses due to presence of short branch length and low GD between these two groups (Fig. 1a–h; Fig. 3b,c). In eight genomic segments tree, the bottom sequence from outbreak group A (PB2: A/California/VRDL73/2009(H1N1); PB1: A/Magadan

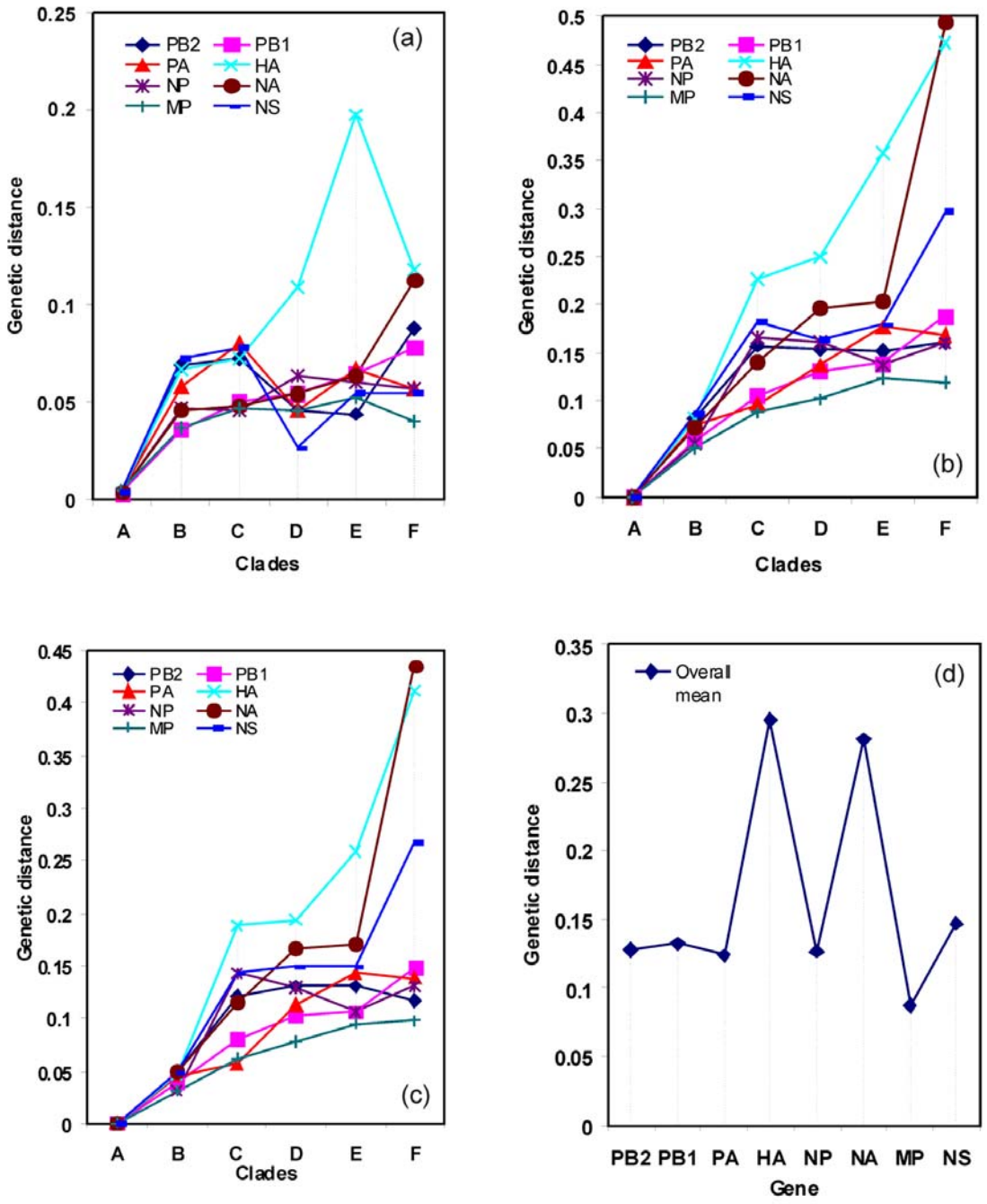

Fig. 3. The groups A–F of each gene segment was used for calculating GD by MEGA 4.1 software using the NJ method under the *p*-distance model and a complete deletion option. (a) GD calculated within a group indicates the number of base differences per site from averaging over all sequence pairs within each group. (b) GD calculated between the groups indicates the number of base differences per site from averaging over all sequence pairs between groups. (c) Net evolutionary divergence calculated between the groups indicates the number of base differences per site from estimation of net average between groups of sequences. (d) Overall mean was calculated for each tree.

/02/2009(H1N1); PA: A/California/VRDL18/2009 (H1N1); HA: A/California/VRDL18/2009(H1N1); NP: A/California/VRDL14/2009(H1N1); NA: A/New York/6771/2009(H1N1); MP: A/Ontario/29801/2009 (H1N1), and NS: A/California/VRDL14/2009(H1N1)) was used as a query to estimate the pair-wise GD and similarity within the sequences of group A and neighbour group B. Based on Figure 4 it should be pointed out that query sequence had very low GD and high similarity with outbreak group A sequences, whereas group B viruses had more GD and less similarity. The mean of pair-wise GD (0.000–0.008) of interior of the NIV group A is evidently smaller than its nearest group B (0.038–0.139) and pair-wise genetic similarity (99–100%) interior of the group A is evidently higher than its nearest group B (86–96%) (Fig. 4). The result of genetic similarity (99–100%) among the NIVs (group A) is parallel to the report of Garten et al. (2009), who found 99.9% genetic identity. Furthermore, NIV gene segment had a sequence identity of 86–99% with closely related

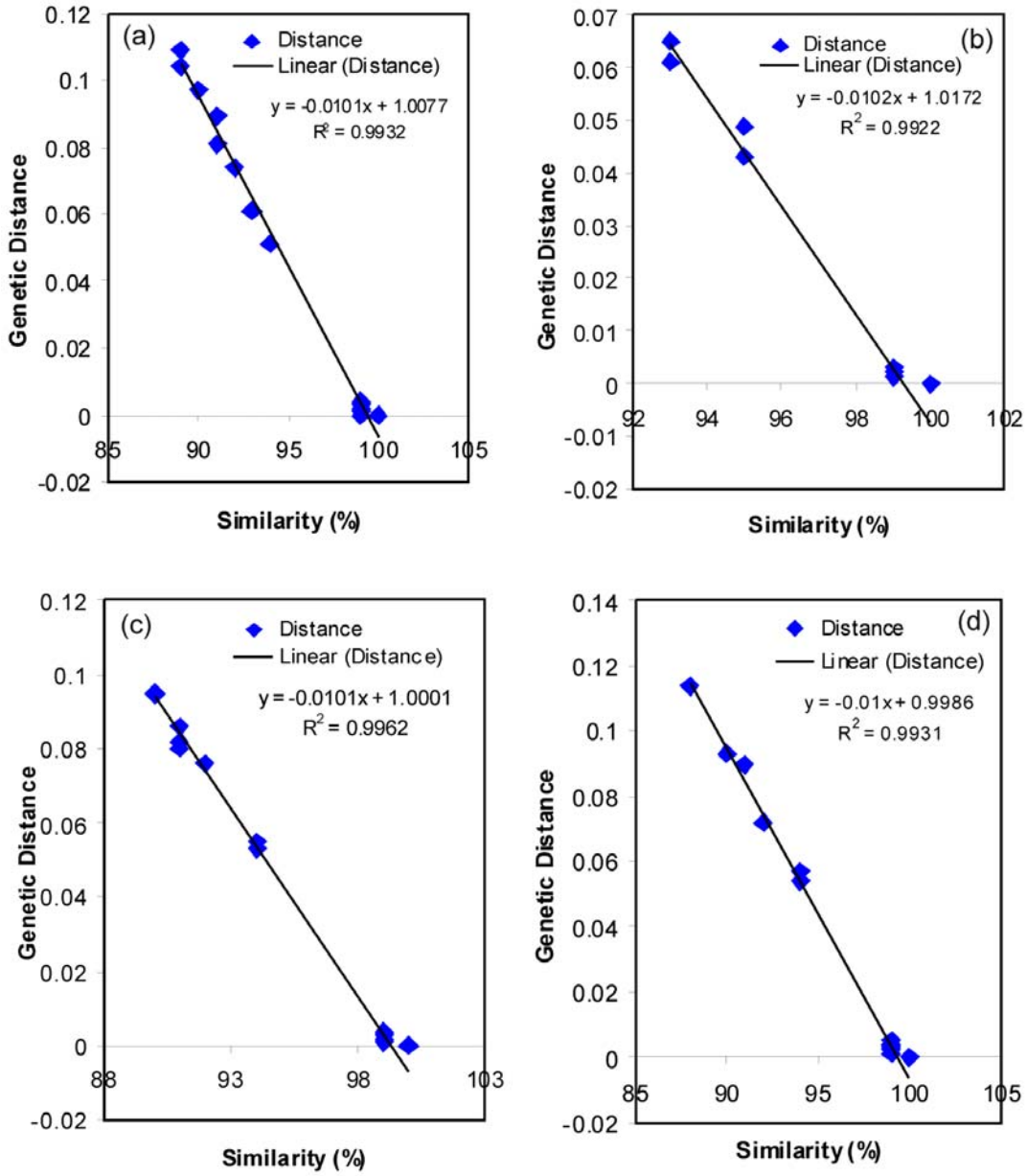

Fig. 4a–d. Similarity and GD were used for perfect negative correlation test. (a) PB2; (b) PB1; (c) PA; (d) HA.

viruses (group B), whereas Garten et al. (2009) and Qu et al. (2011) report show 94–97% and 98–99% sequence identity, respectively. Aforementioned two findings indicate that the NIVs are homogenous lineages and had attempted very little evolution among human after having a long evolutionary history in swine. This considerable divergence from previously sequenced viruses is also shown by the long-branch lengths to the NIVs in the phylogenetic tree of each gene segment. The lack of similarity between NIVs and its relatives indicates that its gene segments have been circulating in undetected animal reservoirs in relatively long period of time. The unsampled diversity of NIVs was calculated using available isolated year of strains from group A and B, because it will give an appropriate results like similarity and diversity analyses.

In cluster analysis (figures not shown), clusters were joined based on the average distance between all members in the two/more groups (Hammer et al. 2001). The present study shows that viruses from both outbreak and nearest branch have homologous genes in their genomes, which might be reflecting the phylogenetic relationship of strains as well as generation of genes by mutations (Table 2). Both phylogenetic and cluster analyses confirm that the reassortment takes place between viruses originated from different species and it could generate NIV with pandemic potential.

The unsampled history was observed for every segment by regressing GD and similarity against sampling year of each strains of group A and B using XLSTAT-2010 (figures not shown). It was found that the evolutionary rate proceeding the NIV pandemic is entirely typical for swine influenza. Therefore, to quantify the period of unsampled diversity, and to estimate the pe-

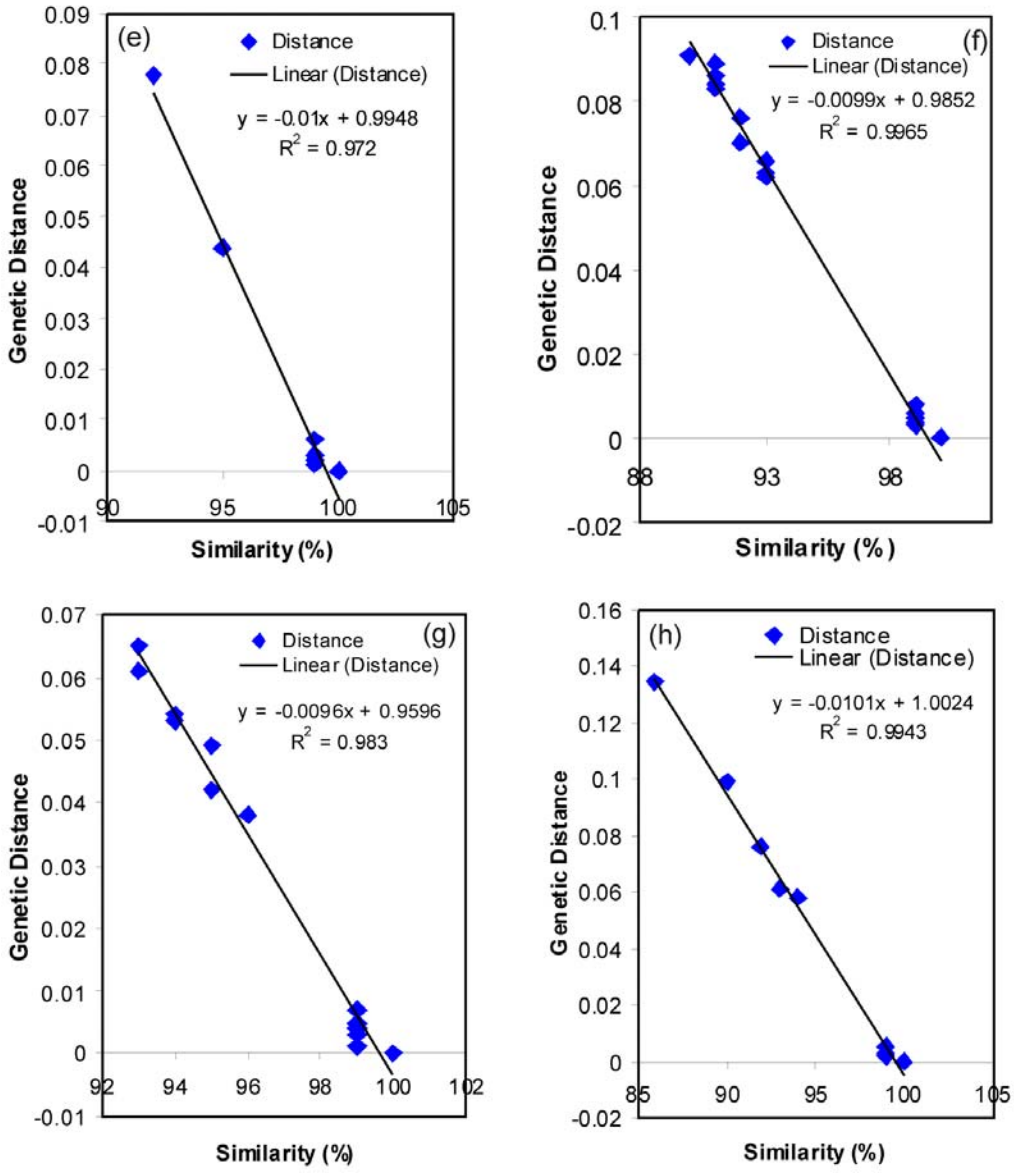

Fig. 4e–h. Similarity and GD were used for perfect negative correlation test. (e) NP; (f) NA; (g) MP; (h) NS.

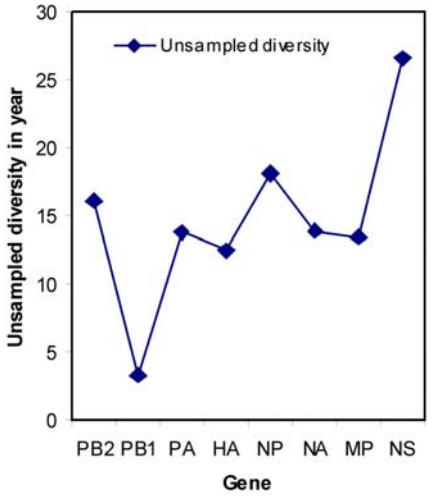

Fig. 5. Estimation of unsampled diversity (in year) and age of NIV by calculating year mean of groups A as well as B and the degree of excess of year could be considered for each gene.

Table 1. The genetic distance values of groups A and B.[a]

Gene	R^2	P value
PB2	0.9932	<0.0001
PB1	0.9922	<0.0001
PA	0.9962	<0.0001
HA	0.9931	<0.0001
NP	0.9720	<0.0001
NA	0.9965	<0.0001
MP	0.9830	<0.0001
NS	0.9943	<0.0001

[a] The values were individually regressed against similarity values to find out the relationship of both variables, which was obtained by calculating R^2 and P values using ANOVA that was performed at 95% confidence interval. Value of P less than 0.05 indicate that model terms are significant.

riod of origin for the NIV, we found the year mean of groups A as well as B; and the degree of excess of year could be considered for each gene. We revealed that the

Table 2. The mutations responsible for adaptation of NIV in human.[a]

Genomic segment	Query sequence	Total base pairs	Variable sites[b]	Unique mutated sites[c]	Distinctive nucleotide mutated sites between both groups A and B
PB2	A/duck/Alberta/35/1976 (H1N1)Canada	2280	467 (20.5%)	40 (8.6%)	A21G, A161G, T390C, C438T, C489T, A531G, A550G, A612G, G/T645A, G660A, T663C, T666C, G708A, A714G, A874G, A/G909T, G/T945A, C972T, A996G, T/C1035A, A1152G, T1161C, T1212C, C1357T, C1491T, C1513A, C1522A, A1563G, G1638A, C1659A, C/G1676T, T1770C, G1788A, A/T1824G, T1839C, A1990C, G1992A, G2050T, A2055G, G2211A
PB1	A/California/07/2004 (H3N2)/USA	2274	260 (11.4%)	50 (19.2%)	G34A, C129A, T147C, G159A, G/A186T, A243G, G267A, T/G303A, G357A, T447C, C499T, T525C, A549C, A603G, G634C, G672A, A705G, A741G, G750T, G756A, A810G, C892A, A1056G, C1090A, C1092T, T1140C, G1157A, T1207C, C1212T, C1304T, C1329A, C1344T, T1365C, T1392C, T/G1527A, T1533C, A1549G, T1566C, A1629G, A1689G, G1854T, C1899T, C1992T, C2013T, C2016T, C2082T, G2085A, T2106C, A2182G, T2205C
PA	A/blue-winged-teal/LA/ B228/1986(H1N1)/USA	2151	395 (18.4%)	40 (10.1%)	C123T, C/A132T, T201C, A204G, C214T, C254T, C282A, C285T, C291A, G321T, G366A, C390T, C420T, C519T, G556A, G611A, C630T, G638A, A711G, T741C, G767A, A789G, T801C, C824T, T829C, C830A, T847C, C868T, T942C, C/T1006A, C1035A, A1067G, A1085G, C/T1164A, T1167C, C1254T, C1311T, T1404C, T1419C, C1648T
HA	A/Swine/Arizona/148/ 1977(H1N1)/USA	1701	337 (19.8%)	44 (13.1%)	C21T, T90A, T177C, A302G, G315A, G/T363A, A390G, A433T, G447A, G488A, G504A, G/A/C/507T, C552T, A553G, G564A, C603T, C642A, A687G, C698T, C722A, T759G, T/C813A, T832G, T876C, G882A, A942G, A1005G, A1071G, A1080G, A1084G, A1125G, G1172A, C1176T, A1218G, T1296C, T1302C, T1359C, T1437C, T1464C, G1468A, C1485T, T1506C, C1534T, G1545A
NP	A/Swine/Arizona/148/ 1977(H1N1)/USA	1497	154 (10.3%)	37 (24.0%)	A54G, C153T, A174G, A195G, C204T, T264C, T276A, T378C, C/T393A, C420T, T426C, C/G441A, T/A471C, A/G531T, A546G, C615T, A654T, T675C, G693A, G726A, G738A, T765C, A798G, G819A, T822C, T888C, T930C, T937G, C/A948G, G1017A, T1089C, G1117A, G1197A, A1227G, A1245G, T1284A, G1338A
NA	A/Swine/France/WVL3/ 1984(H1N1)/France	1410	312 (22.1%)	45 (14.4%)	T12C, G15A, A/T57G, G62A, A/G118C, T/A219C, T224C, A230G, G/A235T, A240G, A246C, T309C, A354G, G357A, T/C366A, C415T, G474A, G566A, G618A, A621G, T627C, C699T, G702A, G747A, A770G, G783A, A786G, G787A, G804A, T805A, G856A, T885C, T921C, G933A, A939G, T1053C, C1094T, G1106A, A1131G, A1146G, G1165A, C1190A, A1242G, G/T1254A, T1347C
MP	A/swine/France/WVL3/ 1984(H1N1)/France	759	111 (14.6%)	13 (11.7%)	G88A, C154T, T273C, C291T, C357T, A402G, C510T, T558A, C/A618T, G639A, A654T, A726G, G729A
NS	A/swine/1931(H1N1)/ USA	693	133 (19.2%)	17 (12.8%)	T15C, G16A, A60G, C127T, C159T, A/T192C, T204C, A270T, G271T, A355T, C/T384A, A385G, T429C, G/A511T, C592A, C616T, A619G

[a]The mutations have been determined between NIV lineage and their closely related group B. The query sequence for each gene segments has been selected based on the earliest year of isolation. The sequence alignments were carried out with MEGA4.1 using ClustalW program and then variable sites were noted.
[b]The percentage of variable nucleotide sites among all sites are indicated in parentheses.
[c]The percentage of unique mutated nucleotide sites among variable sites are indicated in parentheses.

common ancestor of NIV existed between 3.2 and 26.6 years ago (Fig. 5), depending on the genomic segment, therefore the ancestors of the epidemic have been circulating undetected for about decades. Table 1 shows R^2 and P values, which were obtained by regressing GD values against similarity. The relationship of both variables was represented by P values, which was less than 0.05 and indicates that model terms are significant. The present study shows also that all gene segments had statistically significant values. These results confirm a hypothesis that has been suggested by Smith et al. (2009b) and Qu et al. (2011). Regressing the estimate of the gaps in genetic surveillance indicates a long period of unsampled ancestry before the NIV outbreak, suggesting that the reassortment of swine lineages may have occurred years before emergence in human. The long unsampled history of each gene suggests that the multiple reassortments of Eurasian and North American swine, avian and human lineages might not have occurred recently (Smith et al. 2009b; Qu et al. 2011). This genomic structure might thus have been circulating in pigs for several years before emerging in humans,

and we urge caution in making inferences about human adaptation on the basis of the individual genes of ancestry.

Responsible mutations for NIV adaptation in human
The mutations responsible for adaptation of NIV in human are shown in Table 2. The representative strain from group B, which was isolated earlier than other strains, is used as a query to compare it with other strains present in groups A-B. We already found that the NIVs have less divergence in nucleotide level than their background lineages (Fig. 3b,c). Relative to their neighbouring lineages, the majority of mutation sites are distinctive in the NIV lineage, implying an exclusive fixation. The occurrence of distinctive nucleotide mutated sites between both groups A-B is an important indicator of adaptive selection at a molecular level. For PB2, 467 out of 2,280 sites were variable (20.5%). Among the 467 variable sites, the unique mutated sites were observed in 40 sites (8.6%). The mutation profiles for all the other genes were similar to that for PB2 (Table 2), suggesting that some of these mutants will turn out to be useful for the adaptation of organisms to their environment, but also providing a potential evidence of selective pressure in early evolution of the NIV. However, the wide study should be made on functionally important receptor-binding sites that may help researchers to develop more effective vaccines and drugs for 'swine flu'.

Based on our observations, we conclude that the NIV originated in pigs by multiple genetic reassortment of Eurasian and North American swine, avian and human viruses. The each segment PB2, PB1, PA, NP, NA and MP have been derived from related influenza viruses present in group B. Presence of swine viruses in group B belongs to different lineages of hosts, which suggests that multiple reassortment events occurred. Both phylogenetic and cluster analyses confirm that the gene exchange takes place between viruses originated from different species (group B) and it could generate NIV (group A) with unpredictable pandemic potential. An extensive study should thus be made to recognize, which reassortment groups are closely related to NIVs. These ancestral groups might give us clues of when and where these reassortant events have most likely occurred (Garten et al. 2009; Qu et al. 2011). Moreover, 286 new unique mutations are found to be exclusively fixed in whole genome of the NIV lineage, suggesting mutations that may be associated with increased virulence and playing a role of selective sweep in the early evolution. It is also necessary to determine sites in the genes of NIV under greatest or least selection pressure, which will ultimately be important in the effective design of a NIV vaccine and drugs.

Acknowledgements

We thank Prof. K.G. Sivaramakrishnan for his fruitful comments and C. Selvakumar for technical support.

References

Antonovics J., Hood M.E. & Baker C.H. 2006. Molecular virology: was the 1918 flu avian in origin? Nature **440:** E9.

Babakir-Mina M., Dimonte S., Perno C.F. & Ciotti M. 2009. Origin of the 2009 Mexico influenza virus: a comparative phylogentic analysis of the principal external antigens and matrix protein. Arch. Virol. **154:** 1349–1352.

Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J. & Sayers E.W. 2011. GenBank. Nucleic Acids Res. **39:** D32–D37.

Centers for Disease Control and Prevention (CDC). 2009a. Swine influenza A(H1N1) infection in two children – Southern California, March-April 2009. Morbidity and Mortality Weekly Report MMWR **58:** 400–402.

Centers for Disease Control and Prevention (CDC). 2009b. Update swine influenza A(H1N1) infections – California and Texas, April 2009. Morbidity and Mortality Weekly Report MMWR **58:** 1–3.

Dawood F.S., Jain S., Finelli L., Shaw M.W., ... & Zipprich J. 2009. Emergence of a novel swine-origin influenza A(H1N1) virus in humans. N. Engl. J. Med. **360:** 2605–2615.

Ding N., Wu N., Xu Q., Chen K. & Zhang C. 2009. Molecular evolution of novel swine-origin A/H1N1 influenza viruses among and before human. Virus Genes **39:** 293–300.

Fraser C., Donnelly C.A., Cauchemez S., Hanage W.P., ... & Roth C. 2009. Pandemic potential of a strain of influenza A (H1N1): early findings. Science **324:** 1557–1561.

Gambotto A., Barratt-Boyes S.M., Jong M.D., Neumann G. & Kawaoka Y. 2008. Human infection with highly pathogenic H5N1 influenza virus. Lancet **371:** 1464–1475.

Garten R.J., Davis C.T., Russell C.A., Shu B., ... & Cox N.J. 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science **325:** 197–201.

Gibbs M. & Gibbs A. 2006. Molecular virology: was the 1918 pandemic caused by a bird flu? Nature **440:** E8.

Hammer O., Harper D.A.T. & Ryan P.D. 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontologia Electronica **4:** art. 4.

Hatta M., Hatta Y., Kim J.H., Watanabe S., Shinya K., Nguyen T., Lien P.S., Le Q.M. & Kawaoka Y. 2007. Growth of H5N1 influenza A viruses in the upper respiratory tracts of mice. PLoS Pathog. **3:** e133.

Kryazhimskiy S., Bazykin G.A. & Dushoff J. 2008. Natural selection for nucleotide usage at synonymous and nonsynonymous sites in influenza A virus genes. J. Virol. **82:** 4938–4945.

Li Y.G., Chittaganpitch M., Waicharoen S., Kanai Y., Bai G.R., Kameoka M., Takeda N., Ikuta K. & Sawanpanyalert P. 2010. Characterization of H5N1 influenza viruses isolated from humans *in vitro*. Virol. J. **7:** 112.

Ma W., Vincent A.L., Lager K.M., Janke B.H., Henry S.C., Rowland R.R.R., Hesse R.A. & Richt J.A. 2010. Identification and characterization of a highly riulent triple reassortant H1N1 swine influenza virus in the United States. Virus Genes **40:** 28–36.

Munster V.J., de Wit E., van den Brand J.M.A., Herfst S., ... & Fouchier R.A. 2009. Pathogenesis and transmission of swine-origin 2009 A(H1N1) influenza virus in ferrets. Science **325:** 481–483.

Nei M. & Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Inc., New York.

Olsen C.W., Karasin A. & Erickson G. 2003. Characterization of a swine-like reassortant H1N2 influenza virus isolated from a wild duck in the United States. Virus Res. **93:** 115–121.

Olsen W.C. 2002. The emergence of novel swine influenza viruses in North America. Virus Res. **85:** 199–210.

Palese P. 2004. Influenza: old and new threats. Nat. Med. **10:** 82–87.

Qu Y., Zhang R., Cui P., Song G., Duan Z. & Lei F. 2011. Evolutionary genomics of the pandemic 2009 H1N1 influenza viruses (pH1N1v). Virol. J. **8:** 250.

Reid A.H., Taubenberger J.K. & Fanning T.G. 2004. Evidence of an absence: the genetic origins of the 1918 pandemic influenza virus. Nat. Rev. Microbiol. **2:** 909–914.

Reis M.D., Hay A.J. & Goldstein R.A. 2009. Using non-homogeneous models of nucleotide substitution to identify host shift events: application to the origin of the 1918 'Spanish' influenza pandemic virus. J. Mol. Evol. **69:** 333–345.

Schafer J.R., Kawaoka Y., Bean W.J., Suss J., Senne D. & Webster R.G. 1993. Origin of the pandemic 1957 H2 influenza A virus and the persistence of its possible progenitors in the avian reservoir. Virology **194:** 781–788.

Schnitzler S.U. & Schnitzler P. 2009. An update on swine-origin influenza virus A/H1N1: a review. Virus Genes **39:** 279–292.

Scholtissek C. 1990. Pigs as 'mixing vessels' for the creation of new pandemic influenza A viruses. Med. Principles Pract. **2:** 65–71.

Smith G.J.D., Bahl J., Vijaykrishna D., Zhang J.Z., Poon L.L., Chen H., Webster R.G., Peiris J.S. & Guan Y. 2009a. Dating the emergence of pandemic influenza viruses. Proc. Natl. Acad. Sci. USA **106:** 11709–11712.

Smith G.J., Vijaykrishna D., Bahl J., Lycett S.J., Worobey M., Pybus O.G., Ma S.K., Cheung C.L., Raghwani J., Bhatt S., Peiris J S., Guan Y. & Rambaut A. 2009b. Origin and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature **459:** 1122–1126.

Suzuki Y. 2006. Natural selection on the influenza virus genome. Mol. Biol. Evol. **23:** 1902–1911.

Tamura K., Dudley J., Nei M. & Kumar S. 2007. MEGA 4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:** 1596–1599.

Tamura K. & Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10:** 512–526.

Tamura K., Nei M. & Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. USA **101:** 11030–11035.

Tamuri A.U., Reis D.M., Hay A.J. & Goldstein R.A. 2009. Identifying changes in selective constraints: host shifts in influenza. PLoS Comput. Biol. **5:** e1000564.

Taubenberger J.K. 2006. The origin and virulence of the 1918 "Spanish" influenza virus. Proc. Amer. Phil. Soc. **150:** 86–112.

Taubenberger J.K., Reid A.H., Lourens R.M., Wang R., Jin G. & Fanning T.G. 2005. Characterization of the 1918 influenza virus polymerase genes. Nature **437:** 889–893.

Thompson J.D., Higgins D.G. & Gibson T.J. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Van-Reeth K. & Nicoll A. 2009. A human case of swine influenza virus infection in Europe – implications for human health and research. Eurosurveillance. **14:** art. 19124.

Webby R.J. & Webster R.G. 2001. Emergence of influenza A viruses. Phil. Trans. R. Soc. B. **356:** 1817–1828.

Webster R.G., Bean W.J., Gorman O.T., Chambers T.M. & Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. Microbiol. Rev. **56:** 152–179.

Webster R.G., Peiris M., Chen H. & Guan Y. 2006. H5N1 outbreaks and enzootic influenza. Emerg. Infect. Dis. **12:** 3–8.

Received April 29, 2011
Accepted October 5, 2011