AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere®

RESEARCH ARTICLE
Clinical Science and Epidemiology

Next-Generation Sequencing Analysis of the Within-Host Genetic Diversity of Influenza A(H1N1)pdm09 Viruses in the Upper and Lower Respiratory Tracts of Patients with Severe Influenza

Ikuyo Takayama,[a] Binh Gia Nguyen,[b] Co Xuan Dao,[b] Thach The Pham,[b] Tuan Quoc Dang,[b] Phuong Thai Truong,[b] Thanh Van Do,[b] Thuy Thi Phuong Pham,[c] Seiichiro Fujisaki,[a] Takato Odagiri,[a] Hideki Hasegawa,[a] Noriko Nakajima[d]

[a]Influenza Virus Research Center, National Institute of Infectious Diseases, Tokyo, Japan
[b]Bach Mai Hospital, Hanoi, Vietnam
[c]NCGM-Bach Mai Hospital Medical Collaboration Center, Hanoi, Vietnam
[d]Department of Pathology, National Institute of Infectious Diseases, Tokyo, Japan

Ikuyo Takayama and Binh Gia Nguyen contributed equally to this article. Author order was decided based on the specific type of contribution within the collaborating research group.

ABSTRACT The influenza A(H1N1)pdm09 virus emerged in April 2009 with an unusual incidence of severe disease and mortality, and currently circulates as a seasonal influenza virus. Previous studies using consensus viral genome sequencing data have overlooked the viral genomic and phenotypic diversity. Next-generation sequencing (NGS) may instead be used to characterize viral populations in an unbiased manner and to measure within-host genetic diversity. In this study, we used NGS analysis to investigate the within-host genetic diversity of influenza A(H1N1)pdm09 virus in the upper and lower respiratory samples from nine patients who were admitted to the intensive care unit (ICU). A total of 47 amino acid substitution positions were found to differ between the upper and lower respiratory tract samples from all patients. However, the D222G/N substitution in hemagglutinin (HA) protein was the only amino acid substitution common to multiple patients. Furthermore, the substitution was detected only in the six samples from the lower respiratory tract. Therefore, it is important to investigate influenza A(H1N1)pdm09 virus populations using multiple paired samples from the upper and lower respiratory tract to avoid overlooking potentially important substitutions, especially in patients with severe disease.

IMPORTANCE The D222G/N substitution in the hemagglutinin (HA) protein of influenza A(H1N1)pdm09 virus has been reported to be associated with disease severity and mortality in numerous previous studies. In the present study, 75% of lower respiratory samples contained heterogeneous influenza populations that carried different amino acids at position 222 of the HA protein, whereas all upper respiratory samples only contained the wild-type 222D. These results suggest the influenza A(H1N1)pdm09 virus has diversified inside the host owing to differences in tissue specificity. In this study, the within-host genetic diversity of influenza A(H1N1)pdm09 virus was investigated for the first time using next-generation sequencing analysis of the viral whole-genome in samples extracted from the upper and lower respiratory tracts of patients with severe disease.

KEYWORDS within-host genetic diversity, H1N1, influenza, D222G/N

The influenza virus is a single-stranded negative-sense RNA virus belonging to the *Orthomyxoviridae* family. It causes severe acute respiratory tract infections that result in high annual hospitalization rates, especially in children and the elderly (1, 2). It

Citation Takayama I, Nguyen BG, Dao CX, Pham TT, Dang TQ, Truong PT, Do TV, Pham TTP, Fujisaki S, Odagiri T, Hasegawa H, Nakajima N. 2021. Next-generation sequencing analysis of the within-host genetic diversity of influenza A(H1N1)pdm09 viruses in the upper and lower respiratory tracts of patients with severe influenza. mSphere 6:e01043-20. https://doi.org/10.1128/mSphere.01043-20.

Editor Nicole M. Bouvier, Mount Sinai School of Medicine

Copyright © 2021 Takayama et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Ikuyo Takayama, i-taka@nih.go.jp.

Received 14 October 2020
Accepted 16 December 2020
Published 6 January 2021

has also caused worldwide pandemics over the past century. The influenza A virus (IAV) genome consists of eight RNA segments encoding the following proteins: polymerase proteins (PB2, PB1, PA), hemagglutinin (HA) protein, neuraminidase (NA) protein, nucleoprotein (NP), matrix (M1 and M2) proteins, and nonstructural (NS1 and NS2) proteins (3). HA and NA are the major viral surface glycoproteins detected by antibodies. The HA protein is responsible for virus attachment and entry into a host cell, whereas NA protein is essential for virus release after budding and, therefore, for the spread of infection.

In April 2009, the influenza A(H1N1)pdm09 virus [A(H1N1)pdm] emerged in Mexico and the United States and spread rapidly throughout the world (4). Infections with A (H1N1)pdm led to relatively mild disease, with moderate and self-limiting upper respiratory tract illness. However, there was an unusually high incidence of viral pneumonia, severe disease, and death in younger age groups compared to that commonly observed with seasonal influenza (5).

IAVs, like other RNA viruses, exist as heterogeneous populations of closely related genetic variants known as quasispecies (6, 7). Human IAV evolves rapidly on a global scale, with these evolutionary dynamics of IAV being driven by processes that take place within and between infected hosts. Evolution of IAV, primarily through antigenic drift and antigenic shift between infected hosts, has been studied using consensus sequencing data that only represents the dominant nucleotide at each viral genome position (8, 9). Consequently, the use of consensus sequencing data has concealed a large proportion of the viral genotypic and phenotypic diversity occurring within an individual host. Recently, high-throughput sequencing technologies, such as next-generation sequencing (NGS), have allowed unbiased assessments of viral populations and measurements of within-host genetic diversity (10, 11).

The evolution of IAV within hosts can be affected by antigenic selection, antiviral drug treatment, tissue-specific selection, and reassortment (12). Human IAV infects heterogeneous, spatially structured populations of epithelial cells in the human airways. A major variation in regions of the human respiratory tract is the distribution of sialic acid receptors that influenza viruses use to enter host cells. In the human body, sialic acid linked to galactose by an α-2,6 linkage (α2,6-SA) is present in the epithelial cells of the upper respiratory tract, trachea, bronchi, and bronchioles. Alternatively, sialic acid linked to galactose by an α-2,3 linkage (α2,3-SA) is found in epithelial cells of the bronchioles and alveoli (13, 14). Most human IAVs preferentially recognize α2,6-SA. In contrast, avian IAVs preferentially recognize α2,3-SA. This tissue specificity of human and avian IAVs can explain why avian-derived viruses cause severe lower lung infections, such as primary viral pneumonia in humans, and show poor human-to-human transmission (14, 15). Although studies on the within-host evolution of human IAV evaluated by high-throughput sequencing technology have increased in the past years (16), most have focused on the diversity linked to drug treatment (17, 18), associated with chronic infection in immunocompetent hosts (19, 20), and as a result of antigenic selection (21, 22).

In this study, the within-host genetic diversity of A(H1N1)pdm was investigated by NGS analysis of the viruses in the nasopharyngeal swab (NPS) and the tracheal lavage aspirate (TLA) from nine patients who were admitted to the intensive care unit (ICU) and required respiratory management with a ventilator.

RESULTS

Study population characteristics. ICU-admitted patients with A(H1N1)pdm infection at Bach Mai hospital in Hanoi, Vietnam, from February 2016 through May 2017 were enrolled in this study. The characteristics of included patients are shown in Table 1. The patients included six men and three women (nine in total), aged 23 to 70 years old (median age, 45 years). All patients presented with pneumonia as confirmed by chest X-ray and exhibited respiratory distress and hypoxemia, thereby requiring respiratory management with a ventilator in the ICU. A(H1N1)pdm infection was diagnosed by

TABLE 1 Patient information

Season	Patient	Sex	Age	Underlying conditions[a]	P/F ratio on ICU[a] admission	Days from onset: ICU admission	Days from onset: Sampling	Total hospital time (days)	Outcome	Virus strain name
2015/16	Pt1	Male	62	Liver disease	90	0	2	5	Death	A/Vietnam/BM1/2016
	Pt2	Male	45	–	281	4	5	29	Alive	A/Vietnam/BM2/2016
	Pt3	Female	70	Diabetes, hypertension	36	8	11	6	Death	A/Vietnam/BM3/2016
	Pt4	Male	64	Liver disease	302	9	11	11	Death	A/Vietnam/BM4/2016
2016/17	Pt5	Male	50	Hypertension, cardiovascular disease, hematological disease	70	7	7	2	Death[b]	A/Vietnam/BM5/2017
	Pt6	Female	23	Pregnancy	234	14	14	11	Death	A/Vietnam/BM6/2017
	Pt7	Female	24	SLE, CKD	64	0	2	6	Alive	A/Vietnam/BM7/2017
	Pt8	Male	40	–	62	6	6	16	Alive	A/Vietnam/BM8/2017
	Pt9	Male	28	–	45	5	5	17	Alive	A/Vietnam/BM9/2017

[a]SLE, systemic lupus erythematosus; CKD, chronic kidney disease; ICU, intensive care unit; –, unknown or no condition.
[b]Severe condition, discharged by family.

real-time reverse transcription (RT)-PCR assay. Three patients had no underlying conditions. Six patients had PaO_2/FIO_2 (P/F) ratios of ≤100, suggesting severe hypoxia and respiratory dysfunction upon admission to the ICU. Five patients died after a median of six hospital days (range: 2 to 11 days).

Viral load of A(H1N1)pdm and other respiratory pathogens. The copy number of the A(H1N1)pdm M gene in each respiratory sample was calculated based on the results of a quantitative real-time RT-PCR assay targeting the M gene (Table 2). Viral loads in three patients (Pt4, Pt6, and Pt9) were below the calculation limit. For the other six patients, the median viral load in NPS was 1.96×10^5 copies/ml (range: 2.08×10^4 to 3.82×10^8), and the median viral load in TLA was 3.00×10^7 copies/ml (range: 3.18×10^6 to 1.39×10^8). These results suggest these patients were severely infected with A(H1N1)pdm in the upper respiratory tract as well as in the lower respiratory tract.

In addition to A(H1N1)pdm, other respiratory pathogens were identified in eight patients (Table 2). Coinfections with human rhinovirus (RV) ($n = 2$), human coronavirus (HCoV) ($n = 2$), human parainfluenza virus (HPIV) ($n = 1$), human bocavirus (HBoV)

TABLE 2 Quantification cycle (C_q) values derived from real-time RT-PCR-mediated detection of respiratory pathogens and influenza A(H1N1) pdm09 viral load in respiratory samples

Patient	Specimen[b]	A(H1N1)pdm viral load (copies/ml)	Virus[b]: H1N1	Virus[b]: RV	Virus[b]: HCoV strain OC43	Virus[b]: HCoV strain 229E	Virus[b]: HPIV4	Virus[b]: HBoV	Bacteria/fungus: *K. pneumoniae*	Bacteria/fungus: *S. aureus*	Bacteria/fungus: *S. pneumoniae*	Bacteria/fungus: *P. jirovecii*
Pt1	NPS[a]	2.04×10^5	30.60	37.99								
	TLA	2.90×10^7	22.89		35.59							
Pt2	NPS[a]	1.15×10^5	32.63									
	TLA	3.18×10^6	27.19									
Pt3	NPS[a]	2.95×10^6	26.84						37.76	36.20		
	TLA	3.10×10^7	23.38						31.96			
Pt4	NPS[a]	$<1 \times 10^4$	35.65						36.98			
	TLA	$<5 \times 10^3$	36.25						36.43	36.04		
Pt5	NPS[a]	3.82×10^8	19.81					37.45	34.99			37.84
	TLA	1.39×10^8	19.71						23.17			28.86
Pt6	NPS	$<5 \times 10^3$	36.41	37.99			36.17		37.97			
	TLA	8.80×10^4	30.22									
Pt7	NPS	2.08×10^4	32.16			31.65			36.77			
	TLA	7.05×10^7	24.34									
Pt8	NPS	1.87×10^5	28.91						37.49	35.72		
	TLA	3.90×10^6	24.93						36.74			
Pt9	NPS	$<5 \times 10^3$	35.82									
	TLA	2.22×10^6	25.63								30.66	

[a]RNA/DNA extractions of five specimens were performed using 100 μl of the specimen.
[b]NPS, nasopharyngeal swab; TLA, tracheal lavage aspirate; RV, human rhinovirus; HCoV, human coronavirus; HPIV, human parainfluenza virus, HBoV, human bocavirus.

($n = 1$), *Klebsiella pneumoniae* ($n = 6$), *Staphylococcus aureus* ($n = 3$), *Streptococcus pneumoniae* ($n = 1$), and *Pneumocystis jirovecii* ($n = 1$) were observed. Among these, HCoV was the only virus detected in both NPS and TLA. Out of six patients, Pt5 had the highest concentration of *K. pneumoniae* detected in TLA (TLA-Pt5). The concentrations of A (H1N1)pdm were equal to or higher than those of other pathogens in each patient, suggesting that A(H1N1)pdm was the causative pathogen of the disease.

Summary of NGS analysis and genetic features of A(H1N1)pdm among the patients. A total of 18 respiratory samples from 9 patients were analyzed by NGS. The percent coverage of the open reading frame (ORF) of each A(H1N1)pdm gene by the final consensus sequences from all the samples is shown in Table S1 in the supplemental material. Viral genes that were not fully covered are indicated by shades of gray in Table S1. All genes of NPS-Pt9 were not fully covered. Except for NPS-Pt9, the ORFs of M and NS genes were fully covered in all the samples. Of the 18 samples, 6 samples (NPS-Pt1, TLA-Pt1, NPS-Pt5, TLA-Pt7, NPS-Pt8, and TLA-Pt8) had 100% coverage for all genes. For the longer polymerase genes, fewer samples (7 to 13 samples) exhibited 100% coverage.

Phylogenetic analysis was conducted for the A(H1N1)pdm HA genes that had 100% coverage (i.e., HA genes from NPS-Pt4, TLA-Pt4, NPS-Pt6, and NPS-Pt9 were excluded from the analysis) (Fig. 1). In the 2015/16 influenza season, viruses within the subclade 6B.1 and 6B.2 circulated globally, including in Vietnam (23). As shown in Fig. 1, viruses from Pt1 (A/Vietnam/BM1/2015) belong to clade 6B.1 and viruses from Pt2 and Pt3 (A/Vietnam/BM2/2016 and A/Vietnam/BM3/2016, respectively) belong to clade 6B.2; these results are consistent with the global prevalence of influenza virus in the 2015/16 influenza season. All viruses from Pt5 to Pt9 collected in the 2016/17 influenza season were classified under clade 6B.1, which was globally predominant in that season. Clade 6B.2 is characterized by the amino acid substitutions V152T, V173I, E491G, and D501E (24, 25); viruses from Pt2 and Pt3 possessed these substitutions (Fig. 1). All the amino acid differences between the consensus sequences (including the partially covered consensus sequences) and the consensus sequence of NPS-Pt1 were investigated. Except for the above substitutions characterized in the HA protein of clade 6B.2 and the amino acid substitutions at position 222 of the HA protein, no previously reported amino acid markers were found in any of the proteins of A(H1N1)pdm among the patients (data not shown). Analysis of the amino acid substitution at position 222 of the HA protein is described below as a within-host substitution.

To estimate the genetic diversity of A(H1N1)pdm among the patients, the ratio of nonsynonymous to synonymous substitution rates (dN/dS ratio) averaged over the entire gene of each sample was calculated against the consensus sequence of NPS-Pt1 (Table S1). The dN/dS ratio could not be calculated when either dN or dS was 0. In addition, only consensus sequences with 100% coverage were used for this analysis. For the HA gene, the dN/dS ratios for both NPS and TLA could be determined in five patients (Pt2, Pt3, Pt5, Pt7, and Pt8). Higher ratios were observed in samples classified into clade 6B.1 (Pt5, Pt7, and Pt8) than in samples classified into clade 6B.2 (Pt2 and Pt3) in NPSs and TLAs, respectively. However, the differences were not statistically significant (both $P = 0.2$; Mann–Whitney U test).

Within-host genetic diversity of A(H1N1)pdm. To investigate the tissue-specific within-host diversity of A(H1N1)pdm, the dN/dS ratios of paired NPS and TLA samples were compared within 33 pairs in 8 viral genes for which the ratio could be determined. The difference between the ratio of NPS and that of TLA was below 0.002 in almost all pairs of samples. In four pairs (NP gene from Pt6 and Pt7, M gene from Pt2, and NS gene from Pt6), the ratio of NPS was 0.005 or higher than the ratio of TLA.

To investigate the detailed population of A(H1N1)pdm within each patient, the differences in the viral protein sequence between NPS and TLA from each patient were analyzed (Table 3). However, amino acid positions that did not have sufficient mapped reads were excluded from this analysis. A total of 47 amino acid substitution positions were found to be different between the paired NPS and TLA samples. Aside from the substitutions at amino acid position 222 of HA protein, which were present in multiple patients, the remaining 46 substitutions were dispersed among the nine patients with

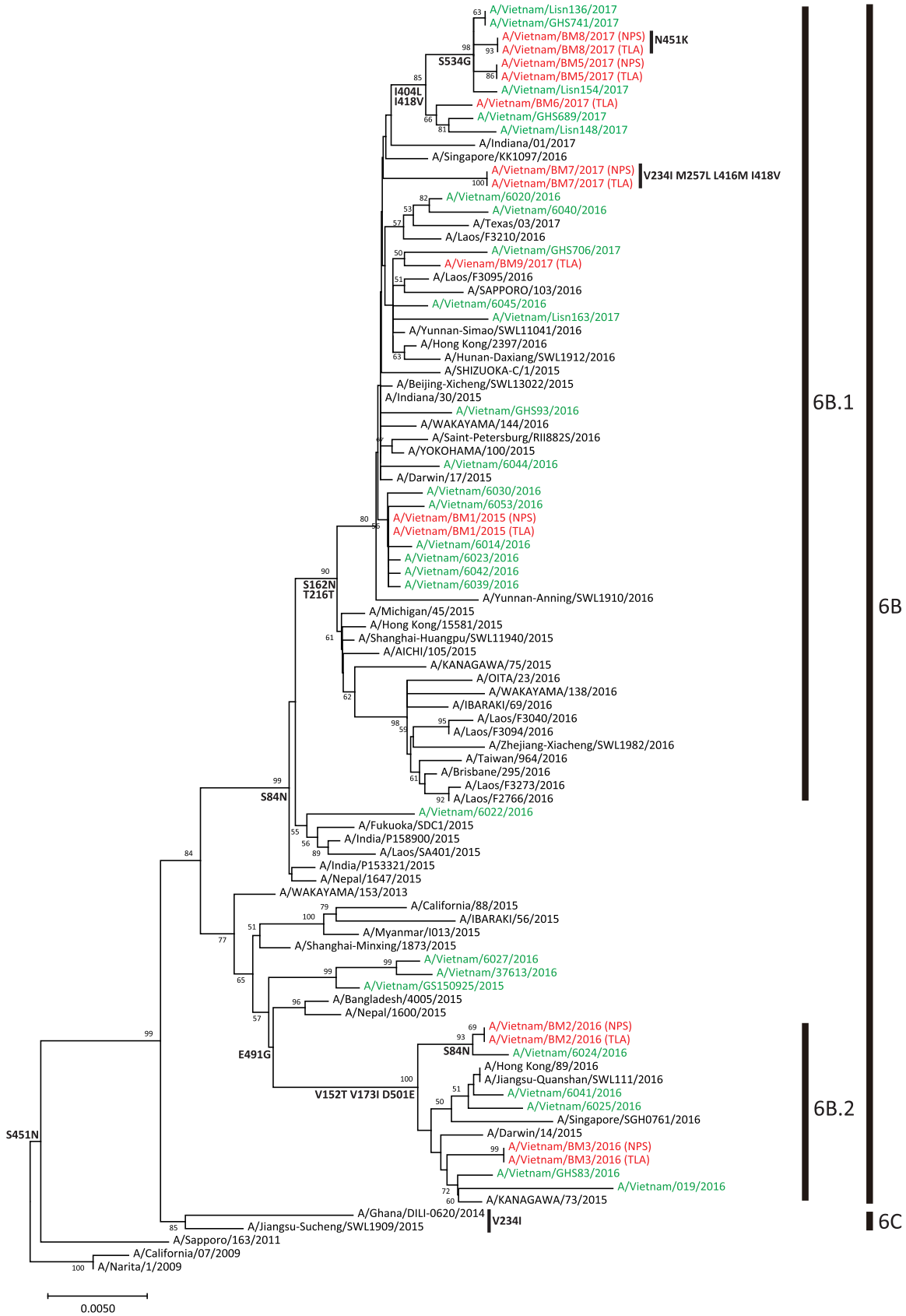

FIG 1 Phylogenetic relationships of the HA gene in influenza A(H1N1)pdm09 viruses. Reference sequences of viruses circulating in Asia in 2015 to 2017 were randomly selected from the GISAID database, and the Vietnamese isolates in 2015 to 2017 are shown in

(Continued on next page)

each substitution found in only a single patient. The number of amino acid substitutions between NPS and TLA in each patient ranged from 0 to 16; Pt1 did not have any substitutions, whereas Pt7 had 16 substitutions. The correlation between the number of amino acid substitutions between NPS and TLA and the number of days from onset to sampling was low ($R^2 = 0.05$) (data not shown), which is similar to the results of previous studies (21, 22). In addition, the correlation between the number of amino acid substitutions between NPS and TLA and the number of mapped reads was also low between samples ($R^2 = 0.17$), as observed in a previous study (19). Notably, there were seven substitution positions for which the frequency of variant amino acids exceeded 50% (Table 3). Among these substitutions, HA-A134T (26) and HA-D222N (27) have been characterized in the HA receptor-binding site. The functions of the other variants have not been reported to date.

Polymorphisms at amino acid position 222 of the HA protein. As we observed the within-host population dynamics at position 222 of the HA protein in multiple patients, the ratio of the wild-type and variant amino acids at this position are shown for each sample (Fig. 2). Because sufficient sequence reads could not be mapped for this amino acid position in samples TLA-Pt4, NPS-Pt6, and NPS-Pt9, no data were available for these samples in this region. Although all seven NPS samples only revealed wild-type 222D in the entire viral population, six out of eight (75%) TLA samples had viral populations that carried different amino acids at position 222 of the HA protein. In detail, only Pt9 had a mixture of 222 D/G, whereas the other samples had 222 D/N/G. However, the frequency of 222G in Pt3 and Pt8 and 222N in Pt7 and Pt8 was below 5%, which is the threshold of variant calling in this study.

Furthermore, we estimated the 47 amino acid differences between NPS and TLA from each viral gene were under positive selection pressure using two selective pressure models: single-likelihood ancestor counting (SLAC) and fixed-effects likelihood (FEL) models (28). Consensus sequences with 100% coverage were used for these analyses. No specific amino acid positions of any viral proteins were detected to be evolving under positive selection using the SLAC model ($P < 0.1$). Conversely, amino acids at position 451 of the NP protein and at position 339 of the NA protein were detected to be evolving under positive selection using the FEL model ($P < 0.1$). At each position, the amino acid differed between Pt1 and the other patients. The results suggested that the substitutions at amino acid position 222 of the HA protein, as well as the other 46 substitutions, were not under positive selection.

DISCUSSION

IAVs replicate as quasispecies within the host and are known for their high evolutionary rate. Most within-host diversity is lost upon transmission, but a few infectious viral populations initially escape the mucus of the new host (22, 29). Understanding within-host evolution is important for determining the evolutionary and epidemiological factors that shape the global evolution of influenza virus. In this study, we quantified and evaluated the within-host genetic diversity of A(H1N1)pdm using NGS analysis of the viral whole-genome sequences within NPS and TLA samples from nine patients with severe influenza infection who were admitted to the ICU, requiring respiratory management with a ventilator. The consensus sequences of all viral genes with 100% coverage could be determined in 6 of the 18 samples (NPS-Pt1, TLA-Pt1, NPS-Pt5, TLA-Pt7, NPS-Pt8, and TLA-Pt8) (Table S1). Alternatively, no or partial viral RNA sequences were determined in the remaining samples, possibly owing to low viral load. As the correlation between the number of amino acid substitutions between NPS and TLA samples and the number of mapped reads was low between samples ($R^2 = 0.17$), the

FIG 1 Legend (Continued)

green. The consensus sequences of the A(H1N1)pdm HA genes generated in this study with 100% coverage are shown in red. Bootstrap values were calculated from 1,000 replicates and values higher than 50% are shown next to the branches. The scale bar indicates the nucleotide substitution per site. Fixed amino acid substitutions are mapped at the major nodes of the tree or next to the virus strain name.

TABLE 3 Amino acid substitutions of influenza A(H1N1)pdm09 viruses that were different between the nasopharyngeal swab (NPS) and the tracheal lavage aspirate (TLA) samples of each patient

Viral protein[c]	Amino acid position	Patient no.	Consensus	Substitution	Frequency of substitution (%)[c]	
					NPS	TLA
HA	134	4	A	T	–[a]	100
	222[b]					
	406	4	N	S	27.1	–
	471	6	C	R	6.8	–
	492	4	E	G	–	13.5
	500	6	I	V	–	47.6
NA	86	4	A	T	–	100
	145	4	S	P	6.9	–
	184	6	C	R	9.7	–
	255	4	I	T	–	6.0
	344	4	N	D	7.5	–
	431	2	P	L	13.9	–
	436	7	I	V	6.0	–
	452	7	T	A	7.5	–
PB2	82	6	N	I	6.8	–
	227	8	V	A	48.9	–
	314	7	A	T	15.1	–
	401	7	A	D	34.8	–
	547	7	V	I	–	22.3
	724	2	V	G	–	22.8
PB1	7	7	L	P	17.9	–
	42	3	T	K	9.0	–
	62	3	G	D	5.1	–
	190	7	R	S	48.3	–
	388	7	K	R	8.2	–
	615	2	L	P	5.7	–
PA	8	7	C	Y	5.3	–
	99	7	G	R	12.2	–
	220	7	P	S	9.2	–
	697	9	D	G	6.8	–
NP	24	6	E	V	15.0	–
	66	6	M	I	6.3	–
	146	6	A	T	100	46.6
	223	7	C	Y	5.3	–
	329	6	V	L	55.4	–
	360	4	T	A	5.7	–
	384	4	R	G	–	7.3
M1	35	4	K	E	–	9.6
	110	7	H	Y	100	–
	236	2	N	K	100	–
M2	60	7	K	E	5.9	–
NS1	9	6	F	S	18.3	–
	67	4	W	R	–	13.1
	80	7	T	A	–	31.0
	98	6	M	I	77.4	–
NS2	9	6	F	S	18.3	–
	101	6	Q	R	16.7	–

[a]Variants were not detected at a 5% detection threshold with NGS analysis.
[b]The variants at codon 222 are shown in Fig. 2.
[c]HA, hemagglutinin; NA, neuraminidase; NP, nucleoprotein; M, matrix; NS, nonstructural.

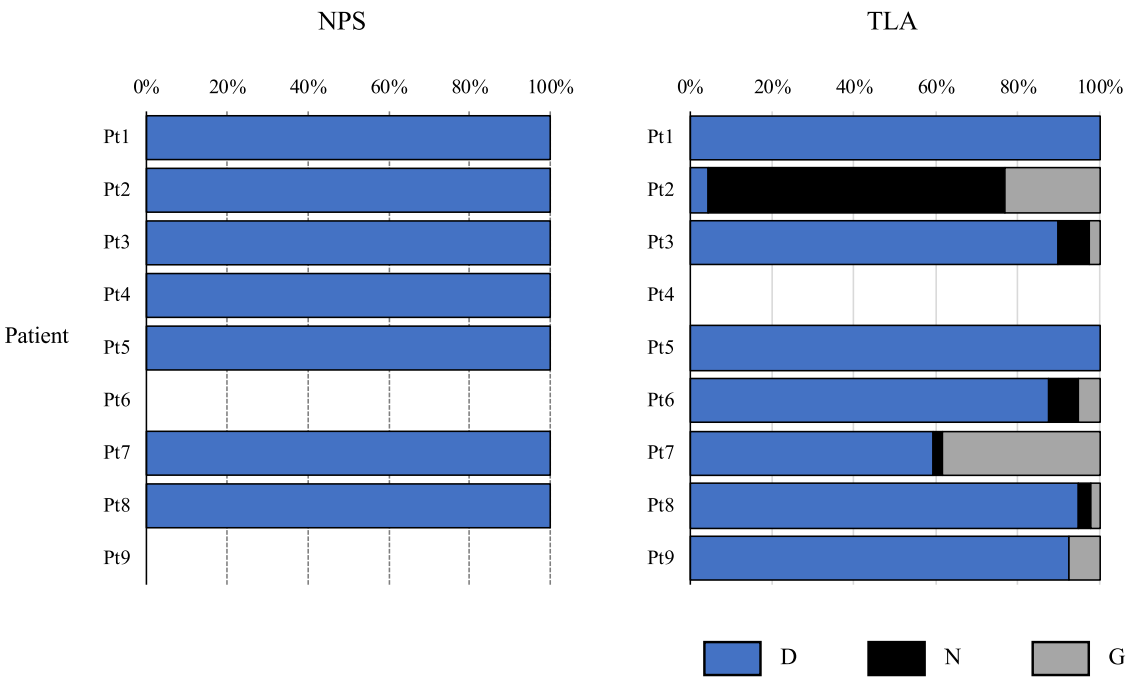

FIG 2 Polymorphisms at amino acid position 222 of the HA protein of influenza A(H1N1)pdm09 virus from nine patients.

NGS analysis results did not appear to be affected by the bias of viral concentration in each sample. A(H1N1)pdm variants among the patients were classified into clade 6B.1 or 6B.2 based on the HA gene genetic clade, with the results being consistent with the global prevalence of A(H1N1)pdm in the 2015/16 and 2016/17 influenza seasons (Fig. 1). In the analysis of the dN/dS ratio averaged over the entire viral gene to estimate genetic diversity, higher ratios in the HA gene were observed in samples classified into clade 6B.1 (Pt5, Pt7, and Pt8) than in samples classified into clade 6B.2 (Pt2 and Pt3) in NPS and TLA samples, respectively, although the ratio differences were not statistically significant. A previous study has also reported that the dN/dS ratios of the HA gene have changed over influenza seasons (30).

Next, to investigate the tissue-specific within-host diversity of A(H1N1)pdm, we compared the dN/dS ratios of NPS and TLA pairs for 33 pairs of all viral genes for which the ratio could be determined. In four pairs (NP gene from Pt6 and Pt7, M gene from Pt2, and NS gene from Pt6), the ratio of NPS was 0.005 or higher than the ratio of TLA. In a previous study, the dN/dS ratio of the NA gene in left lung clones was reported to be lower than that of the NPS sample in one patient (31). From this result, the authors suggested that negative selection preferentially contributed to the within-host genetic diversity of A(H1N1)pdm in the lower respiratory tract. In the present study, preference for negative selection in the lower respiratory tracts was observed in NP, M, and NS genes in a few samples. However, whether negative selection contributed to the diversity of the NP, M, and NS genes of A(H1N1)pdm in the lower respiratory tracts, rather than in the upper respiratory tracts, was not clarified in this study and should be verified using more samples.

The differences in viral protein sequences between paired NPS and TLA samples in each patient were also investigated (Table 3). A total of 47 amino acid substitution positions were identified. Among these, multiple patients had substitutions at amino acid position 222 of the HA protein, whereas the remaining 46 positions were each only substituted in a single patient. It thus seems these 46 substitutions appeared randomly, whereas the HA-D222G/N substitution may have occurred owing to tissue-specific virus diversity. The HA-D222G/N substitution has been reported in several

previous studies. Some studies do not support the association of D222G/N substitution with severe disease, because this substitution is detected even in mild outpatient cases (32). In contrast, numerous other studies have associated HA-D222G with disease severity as this substitution has been described in Europe, the United States, and other countries with a significant frequency in fatal and severe cases (33–37). In recent studies, the HA-D222G/N substitution has been detected in 30% (4/13) of ICU-admitted patients in Italy (38) and the HA-D222G substitution has been detected in 44% (4/9) of severe cases in Norway (33). In both studies, these substitutions were detected only from lower respiratory tract samples. In the present study, 75% (6/8) of TLA samples showed mixed substitutions at amino acid position 222 of the HA protein, whereas all seven NPS samples only showed the wild-type 222D in the entire viral population (Fig. 2). However, we could not find a clear association between the HA-D222G/N substitution and increased disease severity because it is difficult to collect TLA from patients with mild infection; thus, we could not compare prior results with those from patients with mild disease in this study.

The HA-D222G/N substitution has been mapped to the receptor-binding site of HA protein (27). In previous reports, the HA-D222G substitution resulted in a higher binding intensity to α2,3-SA and a lower binding intensity to α2,6-SA in isolated viruses (39, 40). The HA-D222G/N substitution exists as a minor population, whereas HA-D222G appears in $<$1.8% of A(H1N1)pdm worldwide (41). Based on these considerations, the results of the present study suggest that A(H1N1)pdm with a small population of HA-222G/N infected the patients, and the viral population with HA-222G/N showed enhanced viral invasion to deeper areas of the respiratory tract by amplifying viral tropism consequent to viral receptor distribution in the respiratory tract.

In the present study, other than the HA-D222G/N substitution, no amino acid substitution previously associated with severity or other viral phenotypic characteristics was identified in the samples. Furthermore, the selective pressures acting on each amino acid difference between paired NPS and TLA samples were assessed using two methods. The results showed the HA-D222G/N, as well as the other 46 substitutions, were not under positive selection. Conversely, previous studies have identified that the HA-D222G/N and other substitutions in the antigenic sites were under positive selection through alignment of HA genes in A(H1N1)pdm from clinical specimens, including lower respiratory samples (42). In the present study, any potential genetic selection of A(H1N1)pdm caused by the viral tissue specificity and stimulation by a mechanical ventilator was not determined. Moreover, owing to the small number of patients analyzed in this study, substitution under positive selection may not be detected.

In the present study, within-host genetic diversity of A(H1N1)pdm was investigated for the first time using NGS analysis on the viral whole genome in the upper and lower respiratory tracts of patients with severe disease. Owing to the small number of samples analyzed and the viral load of some samples being too low for NGS analysis, more than half of the samples failed to yield a fully covered consensus sequence of A(H1N1)pdm. Thus, further investigation with more samples is needed. However, this study demonstrated the presence of genetic diversity between the upper and lower respiratory tracts of patients with severe disease, especially at amino acid position 222 of the HA protein. Therefore, it is important to investigate A(H1N1)pdm populations using multiple specimens, such as upper and lower respiratory tract paired samples, in order to avoid overlooking potentially important substitutions, especially in patients with severe disease.

MATERIALS AND METHODS

Patients and clinical specimens. This study targeted ICU-admitted patients with A(H1N1)pdm infection at Bach Mai hospital in Hanoi, Vietnam. Samples from patients receiving mechanical ventilation were analyzed in this study. From February 2016 through May 2017, both nasopharyngeal swab (NPS) and tracheal lavage aspirate (TLA) samples were collected at the same time point from nine patients within 72 h from ICU admission. NPS samples were collected in 1 ml of universal transport medium (UTM; Copan, Brescia, Italy), and all respiratory samples were frozen at −80°C until use. Clinical information, including age, sex, underlying commodity, clinical course, and P/F ratio was collected.

This study was conducted in accordance with the Declaration of Helsinki and was approved by the institutional medical ethical committees of Bach Mai Hospital and the National Institute of Infectious Diseases, Japan. All specimens were collected after obtaining written informed consent.

Real-time PCR assay. Total nucleic acids were extracted using the MagMAX CORE nucleic acid purification kit (Thermo Fisher Scientific, Waltham, MA) using 100 or 200 μl of respiratory samples according to the manufacturer's instructions, with an elution volume of 100 μl. Extracts were tested for 33 pathogens using a LightCycler 480 II real-time PCR system (Roche, Basel, Switzerland) and the FTD respiratory pathogens 33 kit (Fast Track Diagnostics, Esch-sur-Alzette, Luxembourg). This kit consists of multiplex real-time PCR assays for the detection of influenza A, B, and C viruses; influenza A subtype H1 pdm09 virus; respiratory syncytial virus; human parainfluenza virus types 1, 2, 3, and 4; human rhinoviruses; human metapneumovirus; human coronavirus OC43, 229E, NL63, and HKU1; human bocavirus; human adenovirus; human enterovirus; human parechovirus; *Mycoplasma pneumoniae*; *Staphylococcus aureus*; *Chlamydia pneumoniae*; *Haemophilus influenzae*; *Haemophilus influenzae* type b; *Streptococcus pneumoniae*; *Legionella* species; *Klebsiella pneumoniae*; *Salmonella* species; *Moraxella catarrhalis*; *Bordetella pertussis*; and *Pneumocystis jirovecii*. The number of IAV M gene copies per ml of respiratory sample was determined by quantitative real-time RT-PCR assay, as described previously (43).

Next-generation sequencing. Viral RNA was extracted from respiratory samples using the QIAamp viral RNA minikit (Qiagen, Hilden, Germany) or the MagMAX CORE nucleic acid purification kit according to the manufacturer's instructions. All eight viral RNA segments of IAV were simultaneously amplified using the SuperScript III one-step RT-PCR system with platinum *Taq* DNA polymerase (Thermo Fisher Scientific) with two slightly modified universal primers (MBTuni-12v2 [5′-ACG CGT GAT CAG CRA AAG CAG G-3′] and MBTuni-13 [5′-ACG CGT GAT CAG TAG AAA CAA GG-3′]) targeting the highly conserved sequence of viral RNA termini, as reported previously (44). The multisegment RT-PCR (M-RTPCR) products were processed for sequencing libraries using the NEBNext Ultra II DNA library prep kit for Illumina (New England BioLabs, Ipswich, MA). Next-generation sequencing (NGS) was performed using the MiSeq reagent kit v3 (150 cycles) (Illumina, San Diego, CA). Following automated cluster generation in MiSeq, the sequences were processed and genetic sequence reads were obtained as FASTQ files.

NGS data analysis. Tentative consensus sequences of each sample were generated using FLUGAS software (version 11, World Fusion, Tokyo, Japan). Trimmed reads of samples for each patient were mapped to the tentative consensus sequences for NPS from each patient or to the tentative consensus sequences for TLA if the consensus sequence for NPS was not obtained, using the CLC Genomics Workbench (version 11, Qiagen). The final consensus sequences were extracted from the mapping results for each segment with every base covered by at least 100 reads. For nucleotide sites with $<$100 reads, a consensus base was not reported, being instead reported as an "N." Variant calling was performed at a frequency of $\geq$5%, covered by a minimum of 100 reads. Throughout the manuscript, amino acid positions of the HA protein are based on H1 numbering without the signal peptide. All raw sequencing reads mapped to the tentative consensus sequences were deposited in the DDBJ Sequence Read Archive (DRA). The final consensus sequences with $\geq$90% coverage of the ORF of each gene were deposited in the Global Initiative on Sharing All Influenza Data (GISAID) database. For each sample and ORF, the dN/dS ratio averaged over the entire gene was calculated compared to the consensus sequence of NPS-Pt1 using the Synonymous and Nonsynonymous Analysis Program (SNAP) provided by the HIV database web site (https://www.hiv.lanl.gov) based on the method described by Nei and Gojobori (45). Tests for positive selection were conducted using SLAC and FEL methods on the Datamonkey 2.0 server (28, 46). For SLAC and FEL methods, a *P* value less than 0.1 indicated significance.

Phylogenetic analysis. Phylogenetic analysis for the HA gene was carried out using Molecular Evolutionary Genetics Analysis software (MEGA, version 7.0) (47). The consensus sequences of A(H1N1) pdm HA genes with 100% coverage generated in this study were analyzed. Evolutionary history was inferred using the neighbor-joining method (48). Evolutionary distances were computed using Kimura's two-parameter method (49). Bootstrap values of HA genes were calculated from 1,000 replicates (50). The representative sequences used in this study for phylogenetic comparison were obtained from the GISAID database.

Statistical analysis. The statistical significance of the dN/dS ratio was calculated using the nonparametric Mann–Whitney U test using R software (version 4.0.2) (51). A *P* value of less than 0.05 indicated statistical significance.

Data availability. All raw sequencing reads mapped to the tentative consensus sequences were deposited in the DRA database (DRA009446). The final consensus sequences with $\geq$ 90% coverage of the ORF of each gene were deposited to the GISAID database under accession numbers EPI_ISL_400648, 400649, 400683 to 400686, 400691, 400694 to 400696, 400698 to 400700, 400705, 400708, 400709, and 400711.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TABLE S1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Jin Takasaki, National Center for Global Health and Medicine, and Tsutomu Kageyama, National Institute of Infectious Diseases, for providing technical assistance in the collection of clinical specimens.

This research was supported by AMED under grant numbers JP18fk0108003 and JP19fk0108082.

REFERENCES

1. Sugaya N, Mitamura K, Nirasawa M, Takahashi K. 2000. The impact of winter epidemics of influenza and respiratory syncytial virus on paediatric admissions to an Urban General Hospital. J Med Virol 60:102–106. https://doi.org/10.1002/(SICI)1096-9071(200001)60:1<102::AID-JMV17>3.0.CO;2-D.
2. Zhou H, Thompson WW, Viboud CG, Ringholz CM, Cheng PY, Steiner C, Abedi GR, Anderson LJ, Brammer L, Shay DK. 2012. Hospitalizations associated with influenza and respiratory syncytial virus in the United States, 1993–2008. Clin Infect Dis 54:1427–1436. https://doi.org/10.1093/cid/cis211.
3. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. Microbiol Rev 56:152–179. https://doi.org/10.1128/MR.56.1.152-179.1992.
4. Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team. 2009. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med 360:2605–2615. https://doi.org/10.1056/NEJMoa0903810.
5. Gao R, Bhatnagar J, Blau DM, Greer P, Rollin DC, Denison AM, Deleon-Carnes M, Shieh WJ, Sambhara S, Tumpey TM, Patel M, Liu L, Paddock C, Drew C, Shu Y, Katz JM, Zaki SR. 2013. Cytokine and chemokine profiles in lung tissues from fatal cases of 2009 pandemic influenza A (H1N1): role of the host immune response in pathogenesis. Am J Pathol 183:1258–1268. https://doi.org/10.1016/j.ajpath.2013.06.023.
6. Lauring AS, Andino R. 2010. Quasispecies theory and the behavior of RNA viruses. PLoS Pathog 6:e1001005. https://doi.org/10.1371/journal.ppat.1001005.
7. Nowak M. 1992. What is a quasispecies? Trends Ecol Evol 7:118–121. https://doi.org/10.1016/0169-5347(92)90145-2.
8. Ferguson NM, Galvani AP, Bush RM. 2003. Ecological and immunological determinants of influenza evolution. Nature 422:428–433. https://doi.org/10.1038/nature01509.
9. Smith DJ, Lapedes AS, De Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus A, Fouchier R. 2004. Mapping the antigenic and genetic evolution of influenza virus. Science 305:371–377. https://doi.org/10.1126/science.1097211.
10. Beerenwinkel N, Günthard HF, Roth V, Metzner KJ. 2012. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front Microbiol 3:329. https://doi.org/10.3389/fmicb.2012.00329.
11. Posada-Cespedes S, Seifert D, Beerenwinkel N. 2017. Recent advances in inferring viral diversity from high-throughput sequencing data. Virus Res 239:17–32. https://doi.org/10.1016/j.virusres.2016.09.016.
12. Xue KS, Moncla LH, Bedford T, Bloom JD. 2018. Within-host evolution of human influenza virus. Trends Microbiol 26:781–793. https://doi.org/10.1016/j.tim.2018.02.007.
13. Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y. 2006. Avian flu: influenza virus receptors in the human airway. Nature 440:435–436. https://doi.org/10.1038/440435a.
14. Van Riel D, Munster VJ, De Wit E, Rimmelzwaan GF, Fouchier RAM, Osterhaus ADME, Kuiken T. 2006. H5N1 virus attachment to lower respiratory tract. Science 312:399. https://doi.org/10.1126/science.1125548.
15. Xu R, de Vries RP, Zhu X, Nycholat CM, McBride R, Yu W, Paulson JC, Wilson IA. 2013. Preferential recognition of avian-like receptors in human influenza A H7N9 viruses. Science 342:1230–1235. https://doi.org/10.1126/science.1243761.
16. Quiñones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA. 2014. Deep sequencing: becoming a critical tool in clinical virology. J Clin Virol 61:9–19. https://doi.org/10.1016/j.jcv.2014.06.013.
17. Ghedin E, Laplante J, DePasse J, Wentworth DE, Santos RP, Lepow ML, Porter J, Stellrecht K, Lin X, Operario D, Griesemer S, Fitch A, Halpin RA, Stockwell TB, Spiro DJ, Holmes EC, St George K. 2011. Deep sequencing reveals mixed infection with 2009 pandemic influenza A (H1N1) virus strains and the emergence of oseltamivir resistance. J Infect Dis 203:168–174. https://doi.org/10.1093/infdis/jiq040.
18. Hu Y, Lu S, Song Z, Wang W, Hao P, Li J, Zhang X, Yen HL, Shi B, Li T, Guan W, Xu L, Liu Y, Wang S, Zhang X, Tian D, Zhu Z, He J, Huang K, Chen H, Zheng L, Li X, Ping J, Kang B, Xi X, Zha L, Li Y, Zhang Z, Peiris M, Yuan Z. 2013. Association between adverse clinical outcome in human disease caused by novel influenza A H7N9 virus and sustained viral shedding and emergence of antiviral resistance. Lancet 381:2273–2279. https://doi.org/10.1016/S0140-6736(13)61125-3.
19. Taubenberger JK, Wu F, Xiao Y, Wang J, Qi L, Kash JC, Xiao Y-L, Ren L, Zhang X. 2018. Deep sequencing of H7N9 influenza A viruses from 16 infected patients from 2013 to 2015 in Shanghai reveals genetic diversity and antigenic drift. mSphere 3:e00462-18. https://doi.org/10.1128/mSphereDirect.00462-18.
20. Pichon M, Simon B, Valette M, Bal A, Picard C, Escuret V, Ottmann M, Gillet Y, Ader F, Lina B, Josset L. 2018. Evolution of influenza genome diversity during infection in immunocompetent patients. bioRxiv https://doi.org/10.1101/435263.
21. Debbink K, McCrone JT, Petrie JG, Truscon R, Johnson E, Mantlo EK, Monto AS, Lauring AS. 2017. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. PLoS Pathog 13:e1006194. https://doi.org/10.1371/journal.ppat.1006194.
22. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. 2018. Stochastic processes constrain the within and between host evolution of influenza virus. Elife 7:e35962. https://doi.org/10.7554/eLife.35962.
23. World Health Organization. 2016. Recommended composition of influenza virus vaccines for use in the 2016–2017 northern hemisphere influenza season. https://www.who.int/influenza/vaccines/virus/recommendations/2016_17_north/en/.
24. Broberg E, Melidou A, Prosenc K, Bragstad K, Hungnes O, WHO European Region and the European Influenza Surveillance Network members of the reporting countries. 2016. Predominance of influenza A(H1N1)pdm09 virus genetic subclade 6B.1 and influenza B/Victoria lineage viruses at the start of the 2015/16 influenza season in Europe. Eurosurveillance 21. https://doi.org/10.2807/1560-7917.ES.2016.21.13.30184.
25. Pebody R, Warburton F, Ellis J, Andrews N, Potts A, Cottrell S, Johnston J, Reynolds A, Gunson R, Thompson C, Galiano M, Robertson C, Mullett D, Gallagher N, Sinnathamby M, Yonova I, Moore C, McMenamin J, de Lusignan S, Zambon M. 2016. Effectiveness of seasonal influenza vaccine in preventing laboratory-confirmed influenza in primary care in the United Kingdom: 2015/16 mid-season results. Eurosurveillance 21:30184. https://doi.org/10.2807/1560-7917.ES.2016.21.13.30179.
26. Khandaker I, Suzuki A, Kamigaki T, Tohma K, Odagiri T, Okada T, Ohno A, Otani K, Sawayama R, Kawamura K, Okamoto M, Oshitani H. 2013. Molecular evolution of the hemagglutinin and neuraminidase genes of pandemic (H1N1) 2009 influenza viruses in Sendai, Japan, during 2009–2011. Virus Genes 47:456–466. https://doi.org/10.1007/s11262-013-0980-5.
27. Igarashi M, Ito K, Yoshida R, Tomabechi D, Kida H, Takada A. 2010. Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. PLoS One 5:e8553. https://doi.org/10.1371/journal.pone.0008553.
28. Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208–1222. https://doi.org/10.1093/molbev/msi105.
29. Poon LLM, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, Sebra R, Halpin RA, Guan Y, Twaddle A, DePasse JV, Stockwell TB, Wentworth DE, Holmes EC, Greenbaum B, Peiris JSM, Cowling BJ, Ghedin E. 2016. Quantifying influenza virus diversity and transmission in humans. Nat Genet 48:195–200. https://doi.org/10.1038/ng.3479.
30. Ma Y, Liu K, Yin Y, Qin J, Zhou Y-H, Yang J, Li S, Poon LLM, Zhang C. 2020. The phylodynamics of seasonal influenza A/H1N1pdm virus in China between 2009 and 2019. Front Microbiol 11:735. https://doi.org/10.3389/fmicb.2020.00735.
31. Hamada N, Imamura Y, Hara K, Kashiwagi T, Imamura Y, Nakazono Y, Chijiwa K, Watanabe H. 2012. Intrahost emergent dynamics of oseltamivir-resistant virus of pandemic influenza A (H1N1) 2009 in a fatally immunocompromised patient. J Infect Chemother 18:865–871. https://doi.org/10.1007/s10156-012-0429-0.
32. Wang B, Dwyer DE, Soedjono M, Shi H, Matlho K, Ratnamohan M, Blyth C, McPhie K, Cunningham AL, Saksena NK. 2011. Evidence of the circulation of pandemic influenza (H1N1) 2009 with D222D/G/N/S hemagglutinin polymorphisms during the first wave of the 2009 influenza pandemic. J Clin Virol 52:304–306. https://doi.org/10.1016/j.jcv.2011.08.023.
33. Kilander A, Rykkvin R, Dudman SG, Hungnes O. 2010. Observed

association between the HA1 mutation D222G in the 2009 pandemic influenza A(H1N1) virus and severe clinical outcome, Norway 2009–2010. Eurosurveillance 15:19498. https://doi.org/10.2807/ese.15.09.19498-en.

34. Resende PC, Motta FC, Oliveira MDLA, Gregianini TS, Fernandes SB, Cury ALF, Rosa MDCD, Souza TML, Siqueira MM. 2014. Polymorphisms at residue 222 of the hemagglutinin of pandemic influenza A(H1N1)pdm09: association of quasi-species to morbidity and mortality in different risk categories. PLoS One 9:e92789. https://doi.org/10.1371/journal.pone.0092789.
35. Marzano A, Marengo A, Ruggiero T, Allice T, Sanna C, Alessandria C, Morgando A, Sciandrello MC, Franzin AM, Rizzetto M, Ghisetti V. 2013. Clinical impact of A/H1/N1/09 influenza in patients with cirrhosis: experience from a nosocomial cluster of infection. J Med Virol 85:1–7. https://doi.org/10.1002/jmv.23454.
36. Devries A, Wotton J, Lees C, Boxrud D, Uyeki T, Lynfield R. 2012. Neuraminidase H275Y and hemagglutinin D222G mutations in a fatal case of 2009 pandemic influenza A (H1N1) virus infection. Influenza Other Respi Viruses 6:85–88. https://doi.org/10.1111/j.1750-2659.2011.00329.x.
37. Chen H, Wen X, To KKW, Wang P, Tse H, Chan JFW, Tsoi H, Fung KSC, Tse CWS, Lee RA, Chan K, Yuen K. 2010. Quasispecies of the D225G substitution in the hemagglutinin of pandemic influenza A(H1N1) 2009 virus from patients with severe disease in Hong Kong, China. J Infect Dis 201:1517–1521. https://doi.org/10.1086/652661.
38. Piralla A, Rovida F, Girello A, Premoli M, Mojoli F, Belliato M, Braschi A, Iotti G, Pariani E, Bubba L, Zanetti AR, Baldanti F. 2017. Frequency of respiratory virus infections and next-generation analysis of influenza A/H1N1pdm09 dynamics in the lower respiratory tract of patients admitted to the ICU. PLoS One 12:e0178926. https://doi.org/10.1371/journal.pone.0178926.
39. Casalegno JS, Ferraris O, Escuret V, Bouscambert M, Bergeron C, Linès L, Excoffier T, Valette M, Frobert E, Pillet S, Pozzetto B, Lina B, Ottmann M. 2014. Functional balance between the hemagglutinin and neuraminidase of influenza A(H1N1)pdm09 HA D222 variants. PLoS One 9:e104009. https://doi.org/10.1371/journal.pone.0104009.
40. Chutinimitkul S, Herfst S, Steel J, Lowen AC, Ye J, van Riel D, Schrauwen EJA, Bestebroer TM, Koel B, Burke DF, Sutherland-Cash KH, Whittleston CS, Russell CA, Wales DJ, Smith DJ, Jonges M, Meijer A, Koopmans M, Rimmelzwaan GF, Kuiken T, Osterhaus ADME, Garcia-Sastre A, Perez DR, Fouchier RAM. 2010. Virulence-associated substitution D222G in the hemagglutinin of 2009 pandemic influenza A(H1N1) virus affects receptor binding. J Virol 84:11802–11813. https://doi.org/10.1128/JVI.01136-10.
41. World Health Organization. 2010. Preliminary review of D222G amino acid substitution in the haemagglutinin of pandemic influenza A(H1N1) 2009 viruses. Wkly Epidemiol Rec 85:21–22.
42. Piralla A, Pariani E, Giardina F, Galli C, Sapia D, Pellegrinelli L, Novazzi F, Anselmi G, Rovida F, Mojoli F, Cereda D, Senatore S, Baldanti F. 2019. Molecular characterization of influenza strains in patients admitted to intensive care units during the 2017–2018 season. Int J Mol Sci 20:2664. https://doi.org/10.3390/ijms20112664.
43. Nakauchi M, Yasui Y, Miyoshi T, Minagawa H, Tanaka T, Tashiro M, Kageyama T. 2011. One-step real-time reverse transcription-PCR assays for detecting and subtyping pandemic influenza A/H1N1 2009, seasonal influenza A/H1N1, and seasonal influenza A/H3N2 viruses. J Virol Methods 171:156–162. https://doi.org/10.1016/j.jviromet.2010.10.018.
44. Zhou B, Donnelly ME, Scholes DT, George KS, Hatta M, Kawaoka Y, Wentworth DE, George S, David E. 2009. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and swine origin human influenza A viruses. J Virol 83:10309–10313. https://doi.org/10.1128/JVI.01109-09.
45. Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426. https://doi.org/10.1093/oxfordjournals.molbev.a040410.
46. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol 35:773–777. https://doi.org/10.1093/molbev/msx335.
47. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.
48. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.
49. Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120. https://doi.org/10.1007/BF01731581.
50. Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.
51. R Core Team. 2020. R: A language and environment for statistical computing. https://www.r-project.org/.