Virology 534 (2019) 96–107

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

Deep sequencing of 2009 influenza A/H1N1 virus isolated from volunteer human challenge study participants and natural infections

Yongli Xiao[a,*], Jae-Keun Park[a], Stephanie Williams[a], Mitchell Ramuta[a], Adriana Cervantes-Medina[b], Tyler Bristol[b], Sarah Smith[b], Lindsay Czajkowski[b], Alison Han[b], John C. Kash[a], Matthew J. Memoli[b], Jeffery K. Taubenberger[a]

[a] Viral Pathogenesis and Evolution Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

[b] Clinical Studies Unit, Laboratory if Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

ARTICLE INFO

Keywords:
Influenza
Deep-sequencing
Human challenge study
Intrahost viral evolution

ABSTRACT

Nasal wash samples from 15 human volunteers challenged with GMP manufactured influenza A/California/04/2009(H1N1) and from 5 naturally infected influenza patients of the 2009 pandemic were deep sequenced using viral targeted hybridization enrichment. Ten single nucleotide polymorphism (SNP) positions were found in the challenge virus. Some of the nonsynonymous changes in the inoculant virus were maintained in some challenge participants, but not in others, indicating that virus is evolving away from the Vero cell adapted inoculant, for example SNPs in the neuraminidase. Many SNP sites in challenge patients and naturally infected patients were found, many not identified previously. The SNPs identified, and phylogenetic analyses, showed that intrahost evolution of the virus are different in challenge participants and naturally infected patients. This study, using hybridization enrichment without PCR, provided an accurate and unbiased assessment of differential intrahost viral evolution from a uniform influenza inoculant in humans and comparison to naturally infected patients.

1. Introduction

Influenza A viruses (IAV) are major health threats, in part due to their rapid, unpredictable evolutionary dynamics, leading to annual epidemics and occasional pandemics, through the mechanisms of mutations in the viral surface glycoproteins (antigenic drift) or viral gene segment reassortment (antigenic shift) (Taubenberger and Kash, 2010). The morbidity and mortality impact of influenza epidemics and pandemics is high; for example, the 1918 pandemic caused at least 50 million deaths globally (Johnson and Mueller, 2002). More recently, the unexpected emergence of the 2009 H1N1 pandemic served as a reminder of how quickly influenza viruses can dramatically change (Garten et al., 2009), and even during the recent seasonal influenza epidemics, antigenic drift of influenza viruses has caused vaccine mismatches, which significantly increased morbidity and mortality (Centers for Disease and Prevention, 2010; Garten et al., 2018). Despite years of research, there are still limited countermeasures available, due in part to our lack of full understanding of how these IAV mutate and adapt over time.

The study of how IAV change and adapt over time on a population level through consensus Sanger sequencing has led to the discovery of antigenic shift, drift, intrasubtypic reassortment, and all related mechanisms. It also identified the important changes in IAV genome that allow it to evade population immunity and cause pandemics or seasonal epidemics (Morens et al., 2010). An aspect of viral evolution that has been explored to a far lesser extent is the amount of viral genetic diversity during infection in a single host and analyzing intrahost viral evolution. Because of viral mutations emerging rapidly during a single host infection, characterizing the changes that occur in genetic diversity during infection at the single host level is very important to model development of antiviral resistance (Ghedin et al., 2012; Memoli et al., 2010). This may be key to understand how viruses evolve to evade the immune system driving the antigenic changes observed at the population level, and also to adapt from animal to human hosts.

Aspects of human influenza infection and pathogenesis such as intrahost viral evolution can be studied to some extent *in vitro* or in animal models, but human healthy volunteer challenge trials offer many advantages (Sherman et al., 2019). Studies performed with well-

* Corresponding author. Viral Pathogenesis and Evolution Section, Laboratory of Infectious Diseases, NIH/NIAID, 33 North Drive MSC 3203, Bethesda, MD, 20892-3203, USA.
E-mail address: yongli.xiao@nih.gov (Y. Xiao).

https://doi.org/10.1016/j.virol.2019.06.004
Received 22 April 2019; Received in revised form 5 June 2019; Accepted 6 June 2019
Available online 13 June 2019
0042-6822/ Published by Elsevier Inc.

characterized wild-type influenza viruses in healthy volunteer hosts offer a unique and controlled setting to model multiple aspects of self-limited human influenza infection including viral adaptation, transmission, host response, immunoprotection, biomarkers, and risk factors for severe disease.

Although a number of influenza challenge trials have been performed since the 1930s (Francis, 1940; Smorodintseff et al., 1937) including one reporting exposure of more than 200 human subjects to four active influenza viruses in 1946 (Henle and Henle, 1946), few had been performed in last two decades. Recently, a well-characterized and controlled healthy volunteer challenge model was described, using a wild-type A(H1N1)pdm09 influenza virus generated by reverse genetics at the Clinical Center of the National Institute of Health, Bethesda, Maryland (Memoli et al., 2015). This challenge model has been utilized to evaluate pathogenesis and correlates of protection of A(H1N1) pdm09 infection in humans (Memoli et al., 2016; Park et al., 2018).

For years, researchers have been applying next generation sequencing technology to sequence influenza samples from patients, for example, discovering mixed infection with 2009 pandemic IAV and the emergence of oseltamivir resistance (Ghedin et al., 2011), high throughput sequencing of influenza B viruses (Zhou et al., 2014), revealing antigenic variants at low frequencies in IAV-infected patients (Dinis et al., 2016), high-throughput identification of IAV H3N2 antigenic drift variants (Mishin et al., 2017), and assessing evolution process of influenza virus within and between hosts (McCrone et al., 2018). In all these studies, samples were sequenced from naturally infected patients and used viral specific PCR amplification strategies. In the current study, we utilized nasal wash samples from challenge participants to evaluate genetic diversity and look for evidence of intrahost viral evolution occurring in multiple challenge participants by tracking the genome changes of the A(H1N1)pdm09 challenge virus during these experimental infections. Using next generation sequencing technology, viral RNA was isolated from nasal wash samples and cDNA libraries were sequenced directly from both challenge participants and naturally infected A(H1N1)pdm09 patients without any prior *in vitro* viral culture and influenza specific PCR, but by using whole transcriptome amplification (WTA) and targeted influenza enrichment in the next generation sequencing pipeline (Xiao et al., 2018a, 2018b). Using this novel and unbiased approach, we were able to characterize viral evolution and diversity in the inoculant virus and individuals from challenge studies, and individuals from natural infections from fall wave of 2009 pademic H1N1. This study identified SNP differences between isolates from influenza challenge participants and those from naturally infected patients. Some identified SNPs were not reported previously and some may be functionally important. The unique nature of the controlled influenza challenge setting allows this study to provide a detailed view of the intrahost influenza viral evolution from a uniformed inoculant virus in humans.

2. Materials and methods

2.1. Clinical studies and sample collection

The clinical samples and data used for this study were collected under 3 separate NIAID Institutional Review Board approved protocols, after all participants signed informed consent for use of their samples. All of the clinical trials described were conducted in accordance with the Declaration of Helsinki and good clinical practice guidelines. The samples from volunteer influenza challenge participants were collected from 2 separate previously published influenza challenge trials (Clinical Trials Identifier NCT01646138 and NCT01971255) (Memoli et al., 2015, 2016). In both of these studies healthy volunteers were inoculated intranasally by using a nasal atomizer from Teleflex (Morrisville, NC) with an A/California/04/2009(H1N1)pdm challenge virus that was produced from cloned genes by viral rescue and good manufacturing practices (GMP) manufactured by Charles River Laboratories (Wilmington, MA), as described (Memoli et al., 2015). The participants were sampled daily via nasal wash both prior to inoculation and for a minimum of 7 days post inoculation while they were quarantined at the NIH Clinical Center.

The samples used from naturally infected patients were collected in a outpatient natural history study of influenza (NCT00533182). In this study, patients with diagnosed influenza infection were enrolled and nasal samples were taken every other day for as long as they remained positive. For this study, samples from 5 naturally infected patients were used, all of whom were healthy, non-smokers who were infected with influenza during the 2009 pandemic. None of the patients had severe illness and non required hospitalization. Samples were collected from the 5 patients between July 3, 2009 and November 6, 2009, with the October and November cases corresponding to the main fall wave of the pandemic in 2009 (Jhung et al., 2011).

2.2. Sample selection for deep sequencing

All nasal washes from the seventeen participants were tested using an FDA approved molecular test for influenza (Memoli et al., 2015). Positive tests were then chosen for deep sequencing by identifying the sample with the highest viral load by influenza A virus matrix 1 gene TaqMan real time PCR (Runstadler et al., 2007). In two of the participants, influenza viruses were not detected by RT-PCR in any of the positive nasal washes. In those cases, the sample collected at the latest time point post-challenge was chosen for deep sequencing.

2.3. Enrichment probe design

Enrichment probes were designed by Agilent Technologies (Santa Clara, CA) using the challenge strain influenza A/California/04/2009(H1N1) virus (CA04) sequence as reference. The overlapping probes were 120 nt in length with 5 nt spacing density. Probes were designed and validated using methods previously described (Xiao et al., 2018a).

2.4. RNA isolation, cDNA synthesis and real-time PCR

The PureLink® RNA Mini Kit (Ambion from Thermo Fisher Scientific, Waltham, MA) was used to isolate total RNA from 200 µl of nasal wash. The final volume of the isolated total RNA was approximately 50 µl per sample. Synthesis of cDNA was carried out using 5 µl of isolated RNA. The reverse transcription (RT) primer (5′-AgC RAA AgC Agg-3′) was used to produce 1st strand cDNA using the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, Waltham, MA). Four µl of the 1st strand cDNA generated was used to perform the real-time PCR reaction (Taqman™ assay) for the influenza A virus matrix 1 gene as previously described (Runstadler et al., 2007)

2.5. Library construction and sequencing

Isolated total RNA from nasal wash was amplified using the Ovation RNA-Seq system V2 from NuGEN (NuGEN, San Carlos, CA). For each sample, 5 µl of total RNA was used as input. The amplified total cDNAs were analyzed by an Agilent 2100 Bioanalyzer using the Agilent High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA) and sheared to 150bp on the Covaris S2 machine (Covaris, Woburn, MA). The Agilent 2100 Bioanalyzer was then used to analyze amplified sheared cDNA using the Agilent High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA) again. Approximately 300 ng of amplified cDNA was used to generate the Illumina sequencing library using the Agilent SureSelect^{XT} Target Enrichment Kit (Agilent, Santa Clara, CA) for Illumina Multiplex Sequencing. Enriched Illumina sequencing libraries were then analyzed with the Agilent 2100 Bioanalyzer using the Agilent High Sensitivity DNA Kit. Libraries were then clustered on an Illumina cBot machine and sequenced on an Illumina GAIIx

sequencer (Illumina, San Diego, CA). In this project, more than 1.1 billion reads and a total of more than 46 Gb of sequence were generated. All sequences generated were deposited as a series into the GenBank SRA database (Accession No: PRJNA528931).

2.6. Data analysis

Reads were mapped to the Bowtie2 (version 2.2.5, http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) indexed A/California/04/2009(H1N1) genome using Tophat2 (release 2.0.13, http://ccb.jhu.edu/software/tophat/index.shtml) downloaded from the Center for Computational Biology, Johns Hopkins University (http://ccb.jhu.edu/) (Trapnell et al., 2009). SAMtools mpileup (version 2.1.0) was used to make SNP calls with minimum base Phred quality score as 25 (Li et al., 2009b). A reported SNP call was one that satisfied the following criteria at the SNP position: 1) more than 100 reads at that position (Barbezange et al., 2018; McGinnis et al., 2016; Xiao et al., 2018b), 2) reads present from both directions, 3) eliminated the variant calls exactly at the end of the read, 4) reads with bases that are different to reference were more than 10% of the aligned reads.

Consensus nucleotide sequence from each segment was constructed using the SNP base that was more than 50% at the site. The generated consensus sequences were concatenated in the same order. After that, multiple sequence alignment was performed using MUSCLE (Edgar, 2004) version 3.8.31 and phylogenetic trees with CA04 as root using approximately maximum likelihood method were constructed with FastTree version 2.1.8 with Jukes-Cantor + CAT model (Price et al., 2010) and displayed in FigTree version 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) including Shimodaira-Hasegawa test values. The precomputed SNP data of all human H1N1 IAVs was downloaded on 04/11/2017 as text files from the Influenza Research Database (IRD) that is updated bimonthly (https://www.fludb.org/brc/snpAnalysis.spg?method=ModifyInputPage&decorator=influenza&ticketNumber=SA_1392488016) and SNPs with score greater than 1 and greater than 10 were used to compare to the SNPs identified in this study respectively. H1N1 viral genomes from 2009 to present were downloaded on 06/03/2019 from Influenza Research Database (IRD) using following criteria: Virus type: A; Subtype: H1N1; Complete Genome Only; Host: human; Region: North America; Date range: March 2009 to May 2019 and 30 genomes were randomly picked from each year (2010–2019) to construct the Supplemental Figure in the same way.

2.7. Protein structure

Crystal structure of H1 hemagglutinin from A/California/04/2009(H1N1) in complex with a neutralizing antibody 3E1(5GJS) (Wang et al., 2016b), neuraminidase from A/California/04/2009 (H1N1) complexed with laninamivir (3TI3) (Vavricka et al., 2011), and matrix protein 1 from influenza A virus (A/California/04/2009 (H1N1)) (3MD2) were downloaded from RCSB Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do) (Berman et al., 2000) and amino acid changes were made and viewed in Swiss-PdbViewer (Guex and Peitsch, 1997).

2.8. Neuraminidase protein expression and enzyme-linked lectin assay

A wild type and two mutant NA constructs (E119K, D151N) were designed based on a previously described method (Xu et al., 2008). Different NA constructs were expressed, purified, concentrated, and quantified as previously described (Park et al., 2018). Different NAs were serially diluted and the NA activities were measured using an enzyme-linked lectin assay (ELLA) as previously described (Wang et al., 2016a).

Table 1
Summary of the volunteers in this study.

Patient	Patient type	Challenge date	Age	Gender	Race	HAI titer
PD11	Challenge	07/31/2012	22	Female	Black	Low (≤1:10)
PD16	Challenge	09/11/2012	43	Male	Black	Low (≤1:10)
PD25	Challenge	10/09/2012	34	Male	White	Low (≤1:10)
PD43	Challenge	03/19/2013	33	Male	Black	Low (≤1:10)
PD45	Challenge	04/02/2013	24	Female	White	Low (≤1:10)
PD46	Challenge	04/02/2013	37	Male	White	Low (≤1:10)
PD47	Challenge	04/02/2013	34	Male	White	Low (≤1:10)
HAI2	Challenge	10/16/2014	27	Female	Black	High (≥1:40)
HAI16	Challenge	12/10/2013	25	Female	Black	Low (≤1:10)
HAI17	Challenge	12/10/2013	30	Female	White	Low (≤1:10)
HAI27	Challenge	03/11/2014	28	Male	White	Low (≤1:10)
HAI35	Challenge	04/22/2014	43	Male	Black	High (≥1:40)
HAI44	Challenge	06/17/2014	34	Female	White	High (≥1:40)
HAI52	Challenge	06/17/2014	23	Female	Asian	High (≥1:40)
HAI57	Challenge	09/09/2014	28	Male	White	Low (≤1:10)
HAI60	Challenge	09/09/2014	25	Male	White	Low (≤1:10)
HAI65	Challenge	09/09/2014	28	Male	White	Low (≤1:10)
201	Natural infection	not applicable	31	Male	White	unavailable
204	Natural infection	not applicable	54	Female	White	unavailable
206	Natural infection	not applicable	15	Male	White	Low (≤1:10)
208	Natural infection	not applicable	41	Male	White	Low (≤1:10)
212	Natural infection	not applicable	25	Female	White	unavailable

3. Results

3.1. Study participants

Demographic and serologic data of the 22 study participants are summarized in Table 1. All 17 influenza challenge participants were healthy adults between the ages of 22–43 with a mean age of 30, and were challenged with the A/California/04/2009 (H1N1) influenza virus (CA04) between Jul 2012 and Oct 2014 (Table 1). All participants (10 male and 7 female) developed mild-to-moderate influenza infections that included both influenza symptoms as well as a positive clinical test for influenza virus in nasal wash after challenge. In addition, five additional participants (3 male and 2 female) from natural history study were naturally infected with a 2009 A(H1N1)pdm virus between 07/03/2009 and 11/06/2009. Of these individuals, 2 naturally infected individuals had low HAI titers (≤1:10) and the other 3 individuals were untested for HAI titers at the time of diagnosis. Of the 17 challenge participants, 13 had pre-challenge HAI titers of ≤1:10, while 4 of them had pre-inoculation HAI titers of ≥1:40.

3.2. Inoculant virus sequencing

Prior Sanger sequencing of this inoculant virus identified 3 non-synonymous changes as compared to the original wild-type isolate from which it was derived (PA: G58S, HA: A388V, and NA: E119K) (Memoli et al., 2015). Illumina sequencing of this inoculant virus generated 3,546,289 reads with 3,050,998 reads mapped to the 8 segments of reference influenza virus genome (CA04) with an average base coverage of 10,763 and its average coverage for each segment are also shown in Supplemental Table 1. From these deep sequencing data, a total of 10 nucleotide changes (7 non-synonymous changes) at different percentages (ranging from 12.3% to 99.5%), were identified in the inoculant virus based on the filtering criteria (Table 2).

3.3. Sequencing of challenge participants samples

Nasal washes from 17 challenge participants at the time of their peak viral shedding as determined by quantitative real-time RT-PCR

Table 2
Nucleotide changes revealed in inoculant virus.

Segment	AA_pos	NT_pos	Ref	Coverage	AA changes	Type	A	T	G	C	N	Avg_Qvalue
CA04_HA	71	245	A	14208	AAA(K)- > AAT(N)	nonsyn	ref = A	T = 7001(0.493)	G = 10(0.001)	C = 14(0.001)	N = 0(0.000)	36.246
CA04_HA	328	1016	A	8553	AAA(K)- > AAG(K)	syn	ref = A	T = 19(0.002)	G = 8509(0.995)	C = 0(0.000)	N = 0(0.000)	35.732
CA04_HA	388	1195	C	7799	GCC(A)- > GTC(V)	nonsyn	A = 9(0.001)	T = 4358(0.559)	G = 16(0.002)	ref = C	N = 0(0.000)	35.795
CA04_HA	451	1384	C	12477	ACT(T)- > AAT(N)	nonsyn	A = 2637(0.211)	T = 7(0.001)	G = 5(0.000)	ref = C	N = 0(0.000)	35.682
CA04_MX	56	192	C	11516	ACT(T)- > ATT(I)	nonsyn	A = 19(0.002)	T = 1647(0.143)	G = 3(0.000)	ref = C	N = 0(0.000)	36.044
CA04_MX	0	1007	A	1884	after_stop	3'utr	ref = A	T = 2(0.001)	G = 231(0.123)	C = 0(0.000)	N = 0(0.000)	35.866
CA04_NA	119	375	G	6230	GAA(E)- > AAA(K)	nonsyn	A = 3055(0.490)	T = 7(0.001)	ref = G	C = 8(0.001)	N = 0(0.000)	35.855
CA04_NA	151	471	G	9681	GAC(D)- > AAC(N)	nonsyn	A = 1192(0.123)	T = 26(0.003)	ref = G	C = 3(0.000)	N = 0(0.000)	36.07
CA04_PA	58	196	G	11889	GGT(G)- > AGT(S)	nonsyn	A = 6732(0.566)	T = 11(0.001)	ref = G	C = 4(0.000)	N = 0(0.000)	35.647
CA04_PB2	10	55	C	490	CTA(L)- > TTA(L)	syn	A = 1(0.002)	T = 120(0.245)	G = 1(0.002)	ref = C	N = 0(0.000)	36.241

were obtained for deep sequencing. The peak shedding day after the inoculation for each challenge participant, the results of influenza A virus matrix 1 gene TaqMan real time PCR (Runstadler et al., 2007) and Illumina deep sequencing are shown in Table 3. RNA from 15 of 17 challenge participant nasal wash samples were Matrix positive by real time RT-PCR and from them, deep sequencing generated an average of 33,103,415 (ranging from 747,289 to 107,430,181) viral reads for each sample. Although the remaining two challenge participants had 1–2 days of positive clinical viral testing (Memoli et al., 2016), TaqMan was unable to detect viral RNA, and deep sequencing revealed only 25 and 81 influenza viral reads, respectively. Therefore, both samples were excluded from subsequent analyses. The mapped read numbers and average coverages for each segment for all challenge samples are shown in Supplemental Table 1. After analysis, a total of 382 SNP positions were identified in these 15 challenge participant samples and the number of synonymous and non-synonymous SNPs that passed the defined filtering criteria (see methods section) from all participant samples are shown in Table 4. Their positions and base call percentages in each segment from each challenge sample are shown in the Supplemental Table 2.

3.4. Sequencing of samples from naturally infected patients

Five naturally influenza infected patients, enrolled in an outpatient study, were sampled during the main wave of the 2009 pandemic (CDC, 2010). None of these patients had severe disease or required hospitalization. Their overall symptoms and clinical courses were similar to those from our challenge participants. Their nasal washes collected at time of diagnosis were processed identically to the samples from the challenge participants. Real time RT-PCR and deep sequencing results from these samples are shown in Table 3 and their average coverage for each segment are shown in Supplemental Table 1. A total of 136 SNP positions were identified and the detected SNP numbers for each sample are shown in Table 4. Their SNP positions and base call percentages in each segment from each sample from naturally infected patients are also shown in the Supplemental Table 2.

3.5. Comparison of SNPs detected in this study and SNPs detected from previously sequenced H1N1 human isolates.

In order to compare the SNPs identified in this study with ones detected from previous studies, the precomputed SNP data of human H1N1 IAVs were downloaded from the Influenza Research Database (IRD) (Squires et al., 2012). When compared to the SNPs from IRD with a SNP score > 1, among all the identified SNP locations in the current study (including from the inoculant, samples from challenge participants, and naturally infected patient samples), there were 271 previously identified at IRD; at a SNP score > 10 level, only 171 SNP position were reported by IRD (Supplemental Table 2). The SNP scores were generated from the modified formula as described in Crooks et al. (Crooks et al., 2004). Briefly, the score is the normalized entropy of the observed allele distribution at each position. The least polymorphic site would have a score of 0 (no polymorphism) and the most polymorphic site would have a score of 200 (highest polymorphism). In addition, at IRD SNP score > 1 level, among the 271 same SNP positions, 219 SNP calls were the same as observed in the current study. Therefore, in the current study, there were 241 SNP positions, including 216 non-synonymous changes from inoculant, challenge participant, and naturally infected patient samples that have not been previously reported at the IRD.

3.6. Comparison of SNPs between challenge and naturally infected patients

The total recovered SNP numbers from each of the naturally infected patients were very similar, from 40 to 49 (with a mean of 45.8 and standard deviation of 3.6). They all share a single SNP that is the

Table 3
Real-time PCR and sequencing results of nasal washes from patients.

Patient	HAI titer	Sample date	TaqMan Ct	Total Reads	Aligned Reads
PD11	Low (≤1:10)	08/06/2012	29.8	82551959	65438485
PD16	Low (≤1:10)	09/14/2012	30.8	26214901	20069140
PD25	Low (≤1:10)	10/15/2012	24	116221507	97162811
PD43	Low (≤1:10)	03/23/2013	30.2	71919481	27743916
PD45	Low (≤1:10)	04/09/2013	19.3	4803656	4263168
PD46	Low (≤1:10)	04/04/2013	30.5	55465375	6901215
PD47	Low (≤1:10)	04/05/2013	26.8	132502898	107430181
HAI2	High (≥1:40)	10/19/2014	31.4	15113819	747289
HAI16	Low (≤1:10)	12/16/2013	30.5	55060301	38214029
HAI17	Low (≤1:10)	12/13/2013	33.7	53107248	34783046
HAI27	Low (≤1:10)	3/15/2014	33.1	42525428	7658412
HAI35	High (≥1:40)	4/27/2014	n/a	15015145	81
HAI44	High (≥1:40)	6/19/2014	36.3	46321232	17868423
HAI52	High (≥1:40)	6/20/2014	n/a	8577145	25
HAI57	Low (≤1:10)	9/13/2014	28.1	49913984	28289874
HAI60	Low (≤1:10)	9/12/2014	31.9	43465379	24423625
HAI65	Low (≤1:10)	9/11/2014	27.8	27705734	15557612
201	Unavailable (natural infection)	07/03/2009	36.7	61026944	51797551
204	Unavailable (natural infection)	10/12/2009	36.9	56618351	45613734
206	Low (≤1:10) (natural infection)	10/14/2009	36.7	46409303	40848733
208	Low (≤1:10) (natural infection)	10/18/2009	37.6	53532122	47207352
212	Unavailable (natural infection)	11/06/2009	35.9	46617339	12277737

Table 4
SNPs numbers from each patient sample.

Patient	HAI titer	Sample day post-challenge	Called SNP #	Nonsyn SNP #	SNP # same as challenge virus
PD11	Low (≤1:10)	6	24	11	2
PD16	Low (≤1:10)	3	17	12	5
PD25	Low (≤1:10)	6	16	5	5
PD43	Low (≤1:10)	4	51	33	4
PD45	Low (≤1:10)	7	7	4	3
PD46	Low (≤1:10)	2	117	85	5
PD47	Low (≤1:10)	3	17	8	6
HAI2	High (≥1:40)	3	28	19	2
HAI16	Low (≤1:10)	6	29	19	3
HAI17	Low (≤1:10)	3	41	29	4
HAI27	Low (≤1:10)	4	35	24	5
HAI35	High (≥1:40)	5	n/a		
HAI44	High (≥1:40)	2	18	13	3
HAI52	High (≥1:40)	3	n/a		
HAI57	Low (≤1:10)	4	29	16	6
HAI60	Low (≤1:10)	3	34	28	3
HAI65	Low (≤1:10)	2	30	22	5
201	Unavailable (natural infection)	not applicable	47	24	1
204	Unavailable (natural infection)	not applicable	48	22	1
206	Low(≤1:10) (natural infection)	not applicable	49	19	1
208	Low(≤1:10) (natural infection)	not applicable	45	20	1
212	Unavailable (natural infection)	not applicable	40	19	1

same as the inoculant, which is an A to G change at matrix segment at nucleotide position 1007 (synonymous change of M2 stop codon TAA- > TAG) (Table 4 and Supplemental Table 2). For the challenge participants, the range of the recovered SNPs was much larger, from 7 (3 SNPs shared with inoculant) to 117 (5 SNPs shared with inoculant) (Table 4) with a mean of 32.9 and standard deviation of 24.9. Considering the whole genome of the CA04 virus, there were a total of 499 positions meeting the criteria as SNP calls from the inoculant, the challenge participant samples, and from the naturally infected patient samples. Of these positions, 73 SNPs were shared in two or more samples. Among those, 17 SNP positions were shared between samples from challenge participants and naturally infected patients, 28 SNP positions were shared between samples only from challenge participants, and 28 positions were shared only between naturally infected patients (Fig. 1, Supplemental Table 3). Examining the SNPs shared only between the naturally infected patients, as compared to the A/California/04/2009(H1N1) (CA04) reference sequence, there were 6 nonsynonymous SNPs in the hemagglutinin gene (Supplemental Table 3), of which D114N, S220T, and E391K distinguish early pandemic sequences like CA04 (isolated on 04/01/2009) from main wave sequences in the fall of 2009, and 3 which distinguish the CA04 reference hemagglutinin from the predominant sequence of 2009 strains at these codons: (P100S, T214A, and I338V). Two nonsynonymous SNPs in the neuraminidase gene (V106I and N248D), one in nucleoprotein (V100I), one in the PA protein (P224S) one in PB1 (R563K), and one in the NS1 gene (V123I) all also distinguish early CA04-like strains from main fall wave sequences.

To further examine the relationship between samples from our challenge participants and naturally infected patients, we constructed approximately maximum-likelihood phylogenetic tree on concatenated nucleotide consensus sequences (with SNP calls at bases over 50%) using CA04 sequence as the root (Fig. 2). The sequences from the naturally infected samples form a well-defined cluster, while those from the challenge samples are clustered together with sub-clusters. The viruses sequenced from the naturally infected patients share many genetic features with other contemporary fall wave viral genomes when

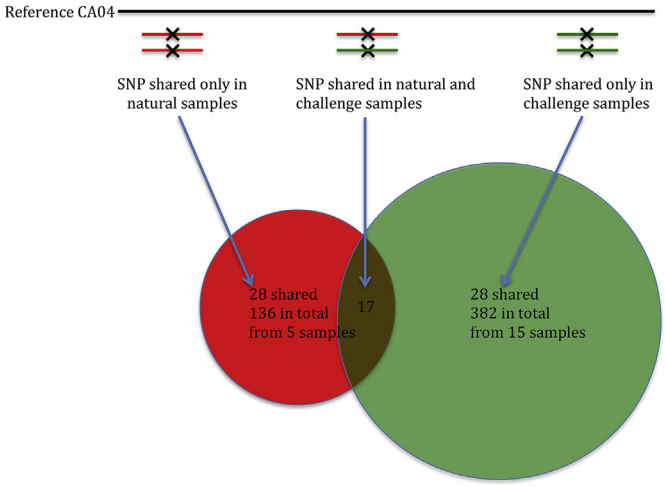

Fig. 1. Venn diagram of shared SNPs identified in this study as compared to the refence genome sequence of CA04. Green: challenge samples. Red: natural samples. Cross: SNPs shared between challenge and natural samples.

we added strains from 2009 to present downloaded from IRD to construct the phylogenetic tree (Supplemental Figure), reflective of the distinct genetic features that characterize fall wave viruses from earlier CA04-like isolates (Jin Gao et al., 2019; Morlighem et al., 2011).

3.7. Intrahost evolution of the viral inoculant in challenge participants

As shown in Table 2, there were 10 pre-existing SNP changes (7 nonsynonymous SNPs) in the inoculant virus, as compared to the reference CA04 sequence from which the GMP-manufactured challenge stock was produced (Memoli et al., 2015), with 4 of them not reported in IRD (Supplemental Table 4). Fig. 3 and Supplemental Table 4 show the base composition of these 10 SNP sites in the inoculant and in all the challenge patient samples analyzed. One of the 3 synonymous SNPs, at HA nt 1016 site, is 99.5% G with coverage of 8553 reads. All the challenge participant samples with coverage at this site, maintained the same base composition (> 99.9% G) as the inoculant. Another synonymous site is at PB2 nt 55 (C- > T). In the inoculant, it was 75.1% C and 24.5% T with coverage of 490 reads. At this site, 11 samples had coverage with 4 samples showing a C/T mixture as in the inoculant and 7 samples only showing a C at this site. Another synonymous SNP site in the inoculant is in the M2 stop codon in the matrix segment, at site nt 1007 (A87.6%, G12.3%). At this site, 12 challenge samples maintained a similar composition of A (~80–90%) and C (~10–20%) and 3 samples have 100% A. Therefore, at synonymous SNP sites in inoculant, in all the challenge participant samples with read coverage, approximately 74% maintained the same synonymous SNPs (at least 10% of the base calls) while 26% of them preserved only reference base at these positions. .

Among the 7 nonsynonymous SNP sites detected in the inoculant (Table 2), 2 are at NA segment at nt 375 (E119K, 49%) and nt 471 (D151N, 12.3%). At both these sites, all 11 challenge participant samples with adequate read coverage lost these two SNPs observed in the inoculant and only the reference sequences (E119 and D151) were kept, matching the reference sequence, A/California/04/2009 H1N1 (Fig. 3). These observations supported selection pressure at these sites in viruses replicating *in vivo*, suggesting their functional importance. Structural modeling revealed that these two nonsynonymous changes were near the NA active site (Fig. 4 A and B). To evaluate the functional significance of these two NA SNPs, recombinant NA proteins were produced in insect cells and evaluated for enzymatic activity *in vitro* using an enzyme-linked lectin assay (ELLA) with the large glycoprotein

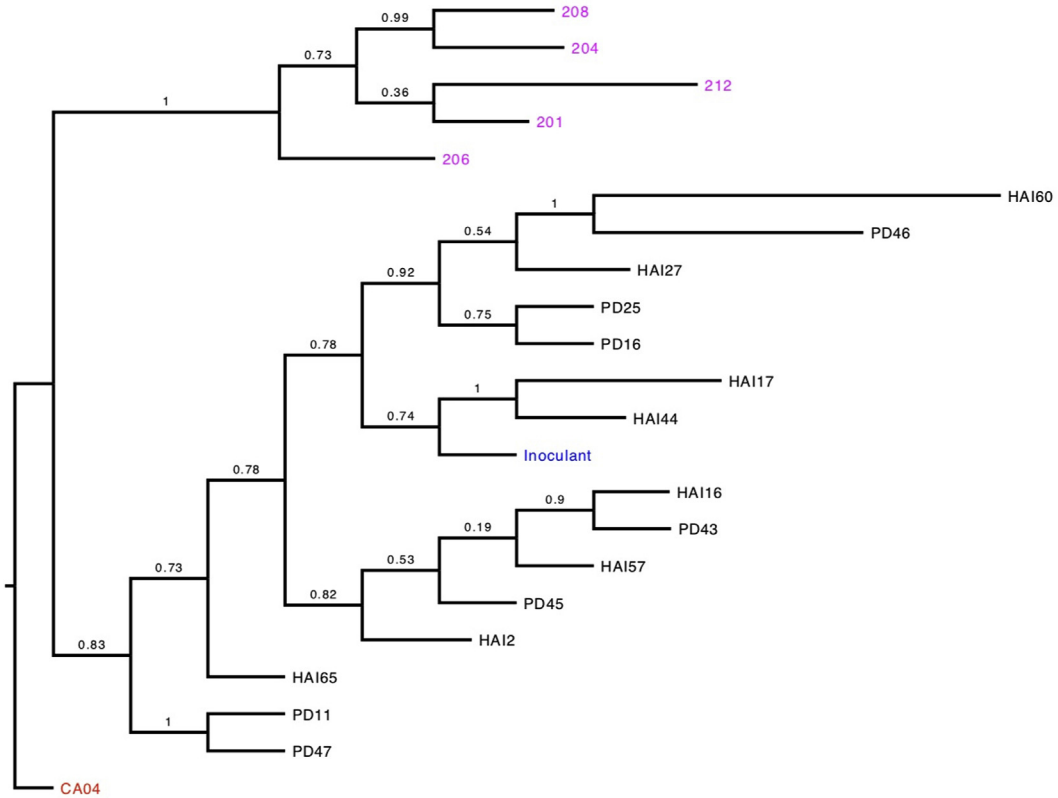

Fig. 2. Phylogenetic tree of concatenated consensus sequences (SNP frequency > 50%) of the samples. Approximately-maximum-likelihood (rooted with CA04: A/California/04/2009(H1N1)) phylogenetic trees with Shimodaira-Hasegawa (SH) values. Red: Cal04 (as root). Blue: inoculant. Black: challenge samples. Magenta: naturally infected samples.

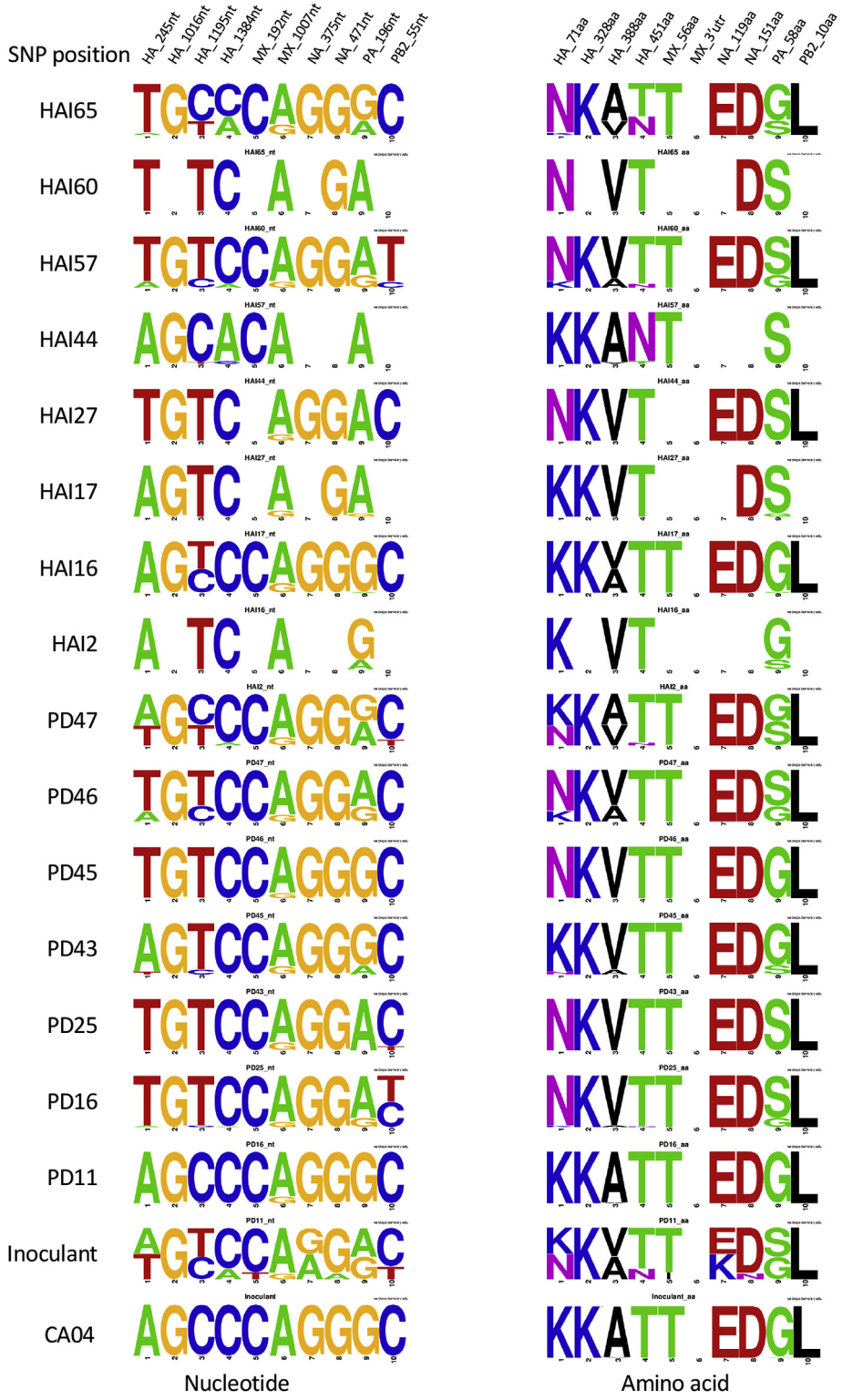

Fig. 3. Nucleotide (nt) base and amino acid (aa) compositions of 10 pre-existing SNP sites in all challenge samples. 1. HA nt site 245 or aa site 71; 2. HA nt site 1016 or aa site 328; 3. HA nt site 1195 or aa site 388; 4. HA nt site 1384 or aa site 451; 5. MX nt site 192 or aa site 56; 6. MX nt site 1007 or M1 3′ UTR/M2 stop site; 7. NA nt site 375 or aa site 119; 8. NA nt site 471 or aa site 151; 9. PA nt site 196 or aa site 58; 10. PB2 nt site 55 or aa site 10. Empty space: no coverage at this site in this sample.

substrate fetuin. As shown in Fig. 4C, the NA protein encoded by the reference CA04 had higher enzymatic activity in the assay than that from the NA proteins with either of the mutations (E119K or D151N).

On the HA segment, there were 3 nonsynonymous SNP sites in the inoculant, K71N, A388V, and T451N (Table 2) compared to CA04 reference genome. These sites are shown modeled on the structure of the reference CA04 HA (Fig. 5). The HA K71N SNP in the inoculant was present at a frequency of 49.3%. Following challenge, 6 participant samples had a K71 dominance (≥93% K), reflective of the wild type CA04 sequence, 6 participants had N71 dominance (≥95% N), and the

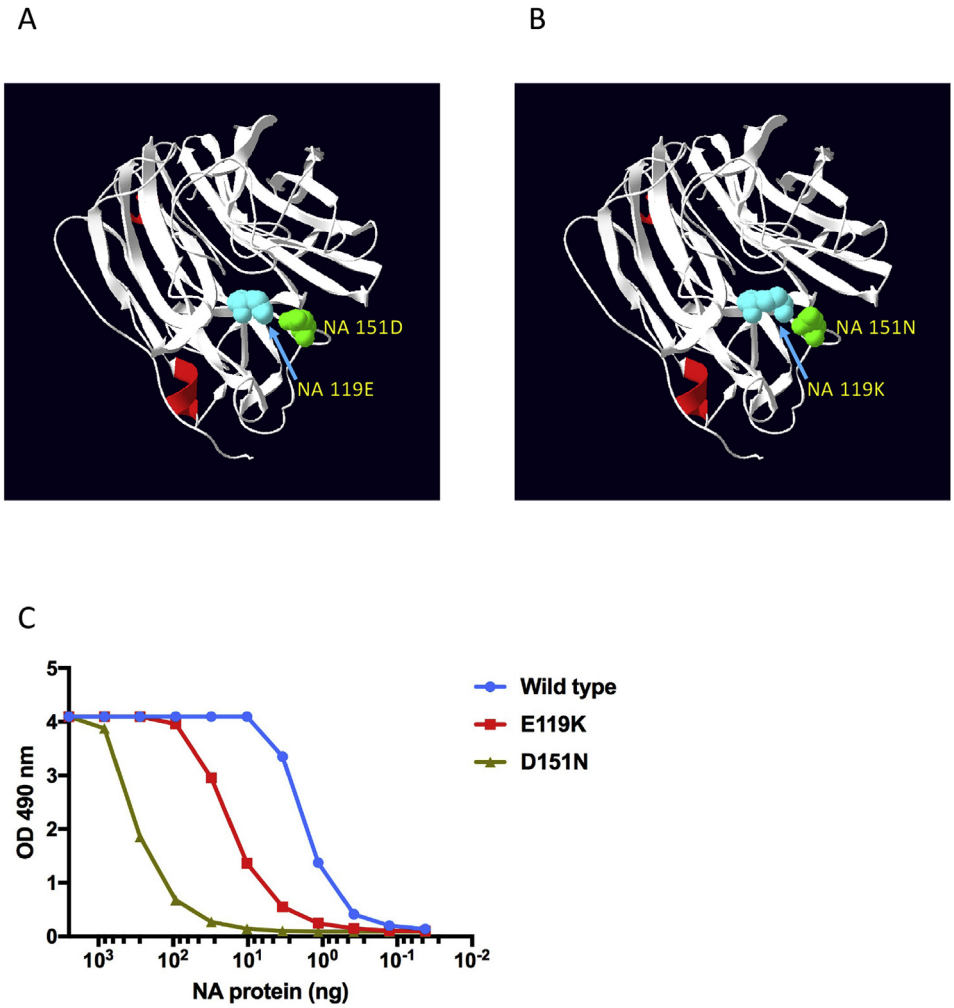

Fig. 4. The structure of 2 NA nonsynonymous SNPs mutations on neuraminidase of A/California/04/2009. A) Reference of A/California/04/2009 with these 2 NA sites shown. B) 2 NA nonsynonymous mutations shown on H1 NA monomer (A/California/04/2009, see Methods). C) NA activity of these 2 mutations.

other 3 samples had a mixture of K and N codons. At HA A388V SNP, 2 samples had an A388 dominance (≥96% A), like the wild type CA04 sequence, 7 samples had a V388 dominance (≥95% V), and 6 samples had a mixture of A/V. At the HA T451N site, only one challenge sample (HAI65) still maintained the SNP with a similar percentage (63.6%T and 36.3%N), one sample (HAI44) showed N451 dominance (94%), and the remaining samples all had the dominant T451 (≥91%), like the reference CA04 sequence.

There were two other nonsynonymous SNPs in the inoculant, in Matrix 1 (T56I) and PA G58S. For the Matrix 1 nonsynonymous SNP the inoculant showed T61I at a frequency of 14.3% (Table 2). In all 11 of the challenge samples with reads over this codon only the wild type T56 was observed at this site, suggestion its functional importance of this site *in vivo*. The amino acid changes of this position were shown on the structure of the reference CA04 HA (Fig. 6). On the PA segment, there was a nonsynonymous SNP G58S (43.3% G and 56.3% S). For this site, all of the 15 challenge samples had adequate coverage. Among them, 3 samples had G58 dominance (≥97% G) like the reference CA04 sequence, 5 samples had an S58 dominance (≥92% S), and 7 samples had a mixture of G/S.

4. Discussion

In this study, deep sequence analysis was performed on viral cDNA libraries generated from nasal wash specimens from human H1N1 influenza challenge study participants (Memoli et al., 2015, 2016) and compared to the sequence of the GMP-manufactured inoculant virus, the reference CA04 sequence, and a small number of naturally infected patients from 2009.

Following rescue, five passages in Vero cells was required to create a seed stock and 1 further passage for GMP manufacture, therefore the inoculant challenge virus stock had undergone a total of 6 passages in Vero cells. Deep-sequencing of inoculant in this study revealed a total of 10 base changes (Table 2). Previous studies have reported that passage in cell lines or fertilized chicken eggs may be associated with nucleotide compositional changes (Bush et al., 2000; Lee et al., 2013; Li et al., 2009a; Roedig et al., 2011; Wu et al., 2017), associated with the error-prone RNA-dependent RNA polymerase of influenza virus (Chen and Holmes, 2006; Jenkins et al., 2002). In a previous study utilizing the same procedures in the control experiments, the highest false positive SNP rate observed was 2.3% (Xiao et al., 2018b). The Illumina sequencing platform has been reported to have an error rate as high as 5% towards the end of the read (Kao et al., 2011). Therefore, in this study we conservatively only reported nucleotide polymorphisms occurring with a frequency greater than 10% from the reference genomes with more than 100 reads covering that position, and as observed in reads from both directions. We also eliminated the variant calls exactly at the first or last mapping positions (Kircher et al., 2009; Nakamura et al., 2011). Whether the nonsynonymous SNPs observed in the inoculant were associated with selection in Vero cells *in vitro* is uncertain, but it was of interest to evaluate intrahost viral evolution of the inoculant virus stock.

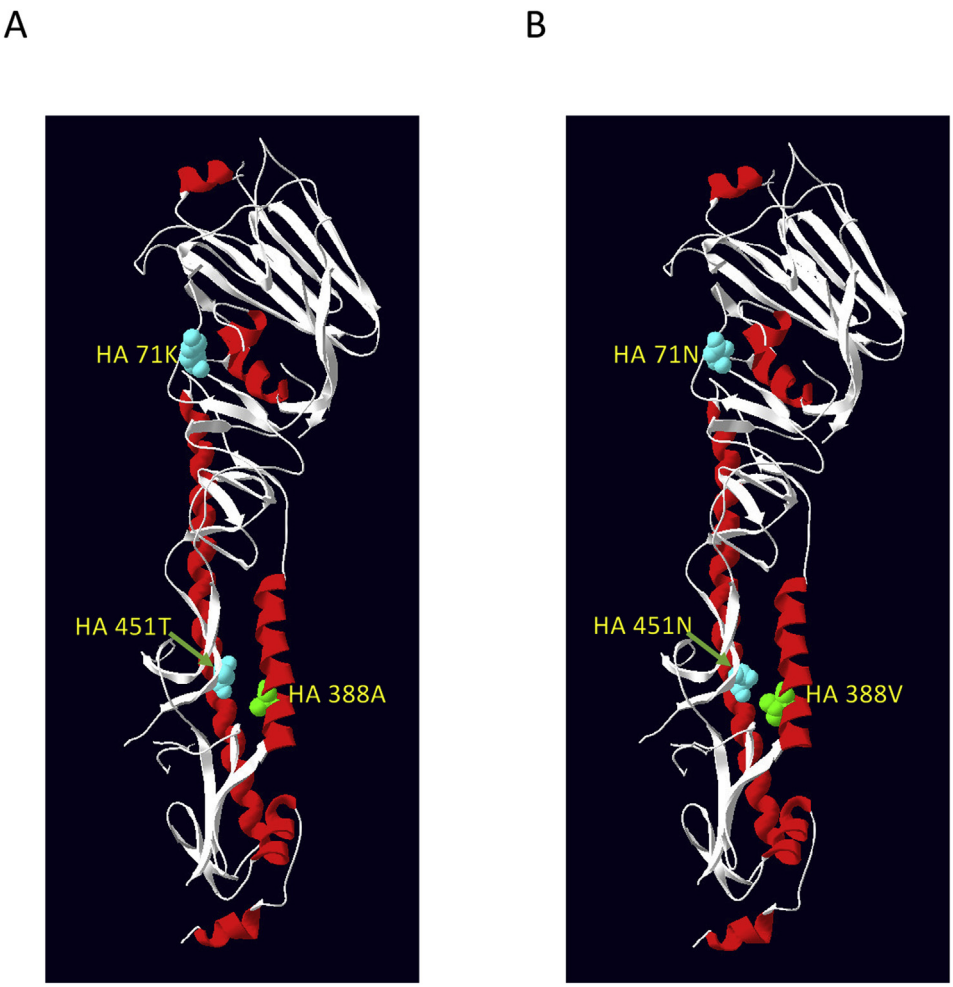

Fig. 5. The structure of 3 HA nonsynonymous SNPs mutations on H1 hemagglutinin of A/California/04/2009. A) Reference of A/California/04/2009. B) 3 HA nonsynonymous mutations shown on an H1 HA monomer (A/California/04/2009, see Methods).

Among the 10 SNPs in the inoculant (Table 2, Fig. 3) as compared to the reference CA04 genome, after challenging participants, it was found that 7 pre-existing SNPs were preserved in one or more patient samples, whereas the other 3, all nonsynonymous SNPs (NA nt position 375 and 471, and M1 nt position 192) were not detected in any of the patient samples, even when we lowered the SNP call cutoff in challenge samples to 5% without any read number, read direction, or read end restrictions, which indicates likely selection pressure on these sites in all the challenge participants, although recent studies have shown that positive selection effect has limited role in evolution of influenza viruses (Han et al., 2019; McCrone et al., 2018). From the crystal structure of neuraminidase of 1918 H1N1 virus, E119 (NA nt position 375) is one of the conserved residues required to structurally stabilize NA catalytic site and D151 (NA nt position 471) belongs to the so-called "the 150 loop" that is conserved among group 1 NAs (Xu et al., 2008) and is the target of neuraminidase inhibitors (Vavricka et al., 2011)

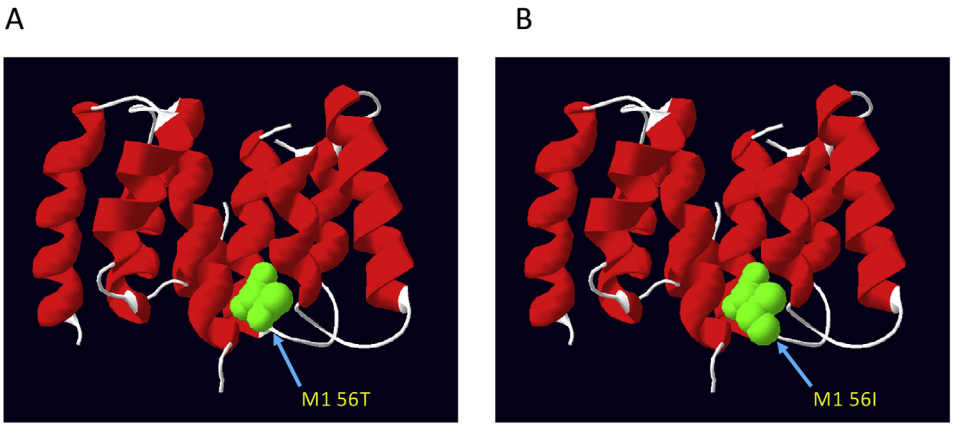

Fig. 6. The structure of the nonsynonymous SNP mutations on matrix protein 1 of A/California/04/2009. A) Reference of A/California/04/2009. B) The nonsynonymous mutation shown on M1 of A/California/04/2009.

(Fig. 4), which clearly indicates their functional importance *in vivo* (Fig. 4A and B). This was confirmed by evaluating NA enzymatic activity *in vitro*, comparing wild type CA04 NA with NAs encoding the nonsynonymous changes (NA E199K or NA D151N). Wild type NA had higher enzymatic activity using a fetuin substrate *in vitro* than either of the NA proteins with the nonsynonymous mutations (Fig. 4C). Together, these data support that there was strong intrahost selection for wild type NA in the challenge participants. Whether there was *in vitro* selection pressure in Vero cells for these nonsynonymous changes to reduce NA activity will require further study.

The third nonsynonymous SNP present in the inoculant but not detected in samples from challenge participants was M1 T56I at a frequency of 14.3%. While this SNP has been noted in Matrix gene sequences of human H1N1 viruses, it has been rarely observed. An analysis of 6845 complete M1 sequences (containing 298 unique M1 protein sequences) from US isolates from 2009 to 2019 revealed only 4 sequences with T56I. It is close to one cavity formed by L3 (49–53), which is structurally important for transportin 1 binding to M1 that promotes the removal of M1 and induces disassembly of vRNP bundles (Miyake et al., 2019).

Among the 7 pre-existing SNPs in inoculant that were preserved in one or more challenge participants nasal wash samples (Fig. 3), 2 were synonymous changes, 1 was in the 3' noncoding region of the Matrix segment (sense orientation), while the remaining 4 were nonsynonymous changes (3 in HA and 1 in PA); percentages at these sites from each sample are shown in Supplemental Table 4.

For the 3 nonsynonymous SNP sites in HA, K71N is located in the globular head domain and A388V and T451N are in the HA stalk (Fig. 5). In the inoculant, the nonsynonymous SNP K71N was present at a frequency of 49.3%. HA amino acid 71 was an asparagine in most human seasonal H1N1 viruses from the 1980s, as part of an N-linked glycosylation site that became fixed after 1986 (Medina et al., 2013). The 2009 H1N1pdm HA encoded a lysine at this site, as had earlier classical swine H1N1 HAs and also the 1918 pandemic HA (Reid et al., 1999). Removal of this HA glycosylation site from a 1991 seasonal H1 virus led to enhanced pathogenicity in mice (Medina et al., 2013). K71N is rare in post-2009 H1N1pdm viruses, and even when present does not form a putative N-linked glycosylation site. This residue is located between the Cb and Ca2 epitopes on H1 HAs (Medina et al., 2013), and nonsynonymous changes at this site may thus alter HA antigenicity.

Another HA nonsynonymous SNP in the inoculant was A388V, in the HA stalk, previously identified after passaging A/California/04/2009/E3 virus 16 times in MDCK cells in the presence of the 6F12 mAb that targets HA stalk region (Anderson et al., 2017; Tan et al., 2012). Pre-challenge and post-challenge HA stalk antibody titers were recently examined in this set of challenge participants (Park et al., 2018). In a parallel study, we have shown that pre-challenge HA stalk antibody titers may have influenced intrahost selection of the A388V SNP (unpublished data).

The third nonsynonymous inoculant SNP in the HA was T451N, which located in the long helix region of HA stalk (Fig. 5). The significance of this site will also require further investigation; however, this site has been mapped experimentally as a T cell epitope (Schanen et al., 2011).

The one nonsynonymous SNP site in PA was G58S (43.3%G, 56.6% S) in the inoculant virus, which has also been mapped as a T cell epitope previously (Tan et al., 2010).

Overall, the observations that some SNPs in inoculant virus been lost, maintained or changed may reflect selection during intrahost replication for changes enhancing viral fitness or selection based on differences in immune pressures between the participants.

An advantage of influenza challenge studies is the ability to characterize both the viral inoculant and the pre-challenge anti-influenza antibody serum titers from the participants prior to the onset of infection (Memoli et al., 2016; Park et al., 2018). From Table 4, one can observe that although the same virus stock was inoculated into each of the challenge participants, the detected SNP numbers emerging during intrahost replication differed widely among the samples, but was not correlated with day of sample collection post-challenge (Table 4). In addition, the difference in the number of called SNPs between different challenge participants was not due to the sequence read numbers produced from them, as was also shown in a previous study (Xiao et al., 2018b). For example, from PD46, there were 6901215 reads mapped to the reference genome with average segment coverage of 97.5%. It has 117 SNPs called while from sample PD45, with total of 4263168 reads mapped with average segment coverage of 99.8%, only 7 SNPs were called. Therefore, the intrahost evolution of the same influenza virus stock in different participants could be quite different. Previous studies have demonstrated that human host factors are important during influenza infections. For example, more than 200 host factors are required for influenza virus replication as demonstrated by genome-wide RNA interference screening (Karlas et al., 2010; Konig et al., 2010). Other studies showed that interferon-inducible transmembrane 3 (IFITM3) is essential for defending the host against influenza A virus *in vivo* (Everitt et al., 2012; Wang et al., 2014). Even influenza vaccination at different times of day has been reported to induce different antibody responses (Long et al., 2016). A recent study of whole genome exon capture and sequencing of 18 H7N9 patients identified 21 genes that were highly associated with H7N9 influenza infection, some of which may be associated with H7N9 influenza susceptibility (Chen et al., 2016). Additionally, different ages and pre-existing immunities in different challenge participants may put different pressures on the challenge virus (Nunez et al., 2017). Therefore, different hosts, varying immunity to influenza from past exposures and/or vaccinations, host factors, and stochastic changes (McCrone et al., 2018) to viral quasispecies due to bottlenecks during the initiation of infection could all play a role in the variable intrahost viral evolution observed here.

In this study, the viral sequences from naturally infected patient samples are more closely related genetically, and the samples from challenge participants share greater genetic relatedness (Fig. 2). The naturally infected patients acquired infections mostly in the fall peak of the pandemic (October–November 2009), when the infection strain may have already further adapted or evolved in human circulation (Domingo et al., 2012), such that fall wave viruses share many nonsynonymous changes not observed in early pandemic viruses, e.g., HA S220T (Morlighem et al., 2011) and E391K (Maurer-Stroh et al., 2010), NA V106I and N248D (Jin Gao et al., 2019), and NP V100I (Xu et al., 2011), which is also associated with 2009 pandemic fall wave isolates (Pan et al., 2010), and has been identified as a mutation associated with human adaptation (Miotto et al., 2010; Reid et al., 2004). Other changes previously reported (Xu et al., 2011), including HA P100S, A214T, and V338, NS I123V, and PA S224P were shared among the naturally infected samples (Supplemental Table 3). When genome relationship was compared adding additional H1N1 viral genomes from 2009 to present, the naturally infected patient genomes were typical of those contemporaneously circulating and they clustered together (Supplemental Figure).

It is also of note that, when SNPs identified here were compared to the precomputed SNPs from human H1N1 sequences at the Influenza Research Database (http://www.fludb.org/brc/home.spg?decorator=influenza) (Squires et al., 2012), we found that 74.1% of the SNP calls from the naturally infected samples had been reported in IRD, whereas only 45.6% of the SNP calls from the challenge participant samples had been previously reported, which suggests that either stochastic bottlenecks or Vero cell-adaptive pressures differ from very similar virus circulation in human populations. In addition, IRD SNP analysis represents population-based consensus sequence SNPs and population-based evolution while the our data represents intrahost individual viral evolution.

To our knowledge, this is the first report that analyzed influenza evolution directly from nasal wash samples from different participants

challenged with the same influenza viral inoculant without culturing the virus *in vitro* or influenza specific PCR amplification. We observed that the reverse-genetics derived CA04 stock influenza virus quickly evolved quasispecies at several different genome positions, even with limited passages in Vero cells during GMP production. The intrahost evolution of the inoculant virus in different participants varied, which might be due to different host factors, immune history, and infection bottlenecks. In addition, the viral sequences from challenge samples were phylogenetically distinct from these from naturally infected samples, which might indicate some adaptation of the inoculant to culture in Vero cells, while still being very similar to the reference CA04 genome.

Conflicts of interest

The authors claim no conflict of interest.

Acknowledgements:

This work was supported by the Intramural Research Program of the National Institutes of Health and National Institute of Allergy and Infectious Diseases.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2019.06.004.

References

Anderson, C.S., Ortega, S., Chaves, F.A., Clark, A.M., Yang, H., Topham, D.J., DeDiego, M.L., 2017. Natural and directed antigenic drift of the H1 influenza virus hemagglutinin stalk domain. Sci. Rep. 7, 14614.

Barbezange, C., Jones, L., Blanc, H., Isakov, O., Celniker, G., Enouf, V., Shomron, N., Vignuzzi, M., van der Werf, S., 2018. Seasonal genetic drift of human influenza a virus quasispecies revealed by deep sequencing. Front. Microbiol. 9, 2596.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein Data Bank. Nucleic Acids Res. 28, 235–242.

Bush, R.M., Smith, C.B., Cox, N.J., Fitch, W.M., 2000. Effects of passage history and sampling bias on phylogenetic reconstruction of human influenza A evolution. Proc. Natl. Acad. Sci. U. S. A. 97, 6974–6980.

CDC, 2010. Update: Influenza Activity — United States, 2009–10 Season.

Centers for Disease, C., Prevention, 2010. Estimates of deaths associated with seasonal influenza — United States, 1976-2007. MMWR Morb. Mortal. Wkly. Rep. 59, 1057–1062.

Chen, C., Wang, M., Zhu, Z., Qu, J., Xi, X., Tang, X., Lao, X., Seeley, E., Li, T., Fan, X., Du, C., Wang, Q., Yang, L., Hu, Y., Bai, C., Zhang, Z., Lu, S., Song, Y., Zhou, W., 2016. Multiple gene mutations identified in patients infected with influenza A (H7N9) virus. Sci. Rep. 6, 25614.

Chen, R., Holmes, E.C., 2006. Avian influenza virus exhibits rapid evolutionary dynamics. Mol. Biol. Evol. 23, 2336–2341.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.

Dinis, J.M., Florek, N.W., Fatola, O.O., Moncla, L.H., Mutschler, J.P., Charlier, O.K., Meece, J.K., Belongia, E.A., Friedrich, T.C., 2016. Deep sequencing reveals potential antigenic variants at low frequencies in influenza a virus-infected humans. J. Virol. 90, 3355–3365.

Domingo, E., Sheldon, J., Perales, C., 2012. Viral quasispecies evolution. Microbiol. Mol. Biol. Rev. 76, 159–216.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Everitt, A.R., Clare, S., Pertel, T., John, S.P., Wash, R.S., Smith, S.E., Chin, C.R., Feeley, E.M., Sims, J.S., Adams, D.J., Wise, H.M., Kane, L., Goulding, D., Digard, P., Anttila, V., Baillie, J.K., Walsh, T.S., Hume, D.A., Palotie, A., Xue, Y., Colonna, V., Tyler-Smith, C., Dunning, J., Gordon, S.B., Gen, I.I., Investigators, M., Smyth, R.L., Openshaw, P.J., Dougan, G., Brass, A.L., Kellam, P., 2012. IFITM3 restricts the morbidity and mortality associated with influenza. Nature 484, 519–523.

Francis, T., 1940. Intranasal inoculation of human individuals with the virus of epidemic influenza. Exp. Biol. Med. 43, 337–339.

Garten, R., Blanton, L., Elal, A.I.A., Alabi, N., Barnes, J., Biggerstaff, M., Brammer, L., Budd, A.P., Burns, E., Cummings, C.N., Davis, T., Garg, S., Gubareva, L., Jang, Y., Kniss, K., Kramer, N., Lindstrom, S., Mustaquim, D., O'Halloran, A., Sessions, W., Taylor, C., Xu, X., Dugan, V.G., Fry, A.M., Wentworth, D.E., Katz, J., Jernigan, D., 2018. Update: influenza activity in the United States during the 2017-18 season and composition of the 2018-19 influenza vaccine. MMWR Morb. Mortal. Wkly. Rep. 67, 634–642.

Garten, R.J., Davis, C.T., Russell, C.A., Shu, B., Lindstrom, S., Balish, A., Sessions, W.M., Xu, X., Skepner, E., Deyde, V., Okomo-Adhiambo, M., Gubareva, L., Barnes, J., Smith, C.B., Emery, S.L., Hillman, M.J., Rivailler, P., Smagala, J., de Graaf, M., Burke, D.F., Fouchier, R.A., Pappas, C., Alpuche-Aranda, C.M., Lopez-Gatell, H., Olivera, H., Lopez, I., Myers, C.A., Faix, D., Blair, P.J., Yu, C., Keene, K.M., Dotson Jr., P.D., Boxrud, D., Sambol, A.R., Abid, S.H., St George, K., Bannerman, T., Moore, A.L., Stringer, D.J., Blevins, P., Demmler-Harrison, G.J., Ginsberg, M., Kriner, P., Waterman, S., Smole, S., Guevara, H.F., Belongia, E.A., Clark, P.A., Beatrice, S.T., Donis, R., Katz, J., Finelli, L., Bridges, C.B., Shaw, M., Jernigan, D.B., Uyeki, T.M., Smith, D.J., Klimov, A.I., Cox, N.J., 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 325, 197–201.

Ghedin, E., Holmes, E.C., DePasse, J.V., Pinilla, L.T., Fitch, A., Hamelin, M.E., Papenburg, J., Boivin, G., 2012. Presence of oseltamivir-resistant pandemic A/H1N1 minor variants before drug therapy with subsequent selection and transmission. J. Infect. Dis. 206, 1504–1511.

Ghedin, E., Laplante, J., DePasse, J., Wentworth, D.E., Santos, R.P., Lepow, M.L., Porter, J., Stellrecht, K., Lin, X., Operario, D., Griesemer, S., Fitch, A., Halpin, R.A., Stockwell, T.B., Spiro, D.J., Holmes, E.C., 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.

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.

Han, A.X., Maurer-Stroh, S., Russell, C.A., 2019. Individual immune selection pressure has limited impact on seasonal influenza virus evolution. Nat Ecol Evol 3, 302.

Henle, W., Henle, G., et al., 1946. Experimental exposure of human subjects to viruses of influenza. J. Immunol. 52, 145–165.

Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. 54, 156–165.

Jhung, M.A., Swerdlow, D., Olsen, S.J., Jernigan, D., Biggerstaff, M., Kamimoto, L., Kniss, K., Reed, C., Fry, A., Brammer, L., Gindler, J., Gregg, W.J., Bresee, J., Finelli, L., 2011. Epidemiology of 2009 pandemic influenza A (H1N1) in the United States. Clin. Infect. Dis. 52 (Suppl. 1), S13–S26.

Jin Gao, L.C., Burke, David, Wan, Hongquan, Wilson, Patrick, Memoli, Matthew, Xu, Xiyan, Harvey, Ruth, Wrammert, Jens, Ahmed, Rafi, Taubenberger, Jeffery, Smith, Derek, Fouchier, Ron, Eichelberger, Maryna, 2019. Antigenic Drift of the Influenza A(H1N1)pdm09 Neuraminidase Results in Reduced Effectiveness of A/California/7/2009 (H1N1pdm09)-specific Antibodies. MBio In press.

Johnson, N.P., Mueller, J., 2002. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. Bull. Hist. Med. 76, 105–115.

Kao, W.C., Chan, A.H., Song, Y.S., 2011. ECHO: a reference-free short-read error correction algorithm. Genome Res. 21, 1181–1192.

Karlas, A., Machuy, N., Shin, Y., Pleissner, K.P., Artarini, A., Heuer, D., Becker, D., Khalil, H., Ogilvie, L.A., Hess, S., Maurer, A.P., Muller, E., Wolff, T., Rudel, T., Meyer, T.F., 2010. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. Nature 463, 818–822.

Kircher, M., Stenzel, U., Kelso, J., 2009. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. Genome Biol. 10, R83.

Konig, R., Stertz, S., Zhou, Y., Inoue, A., Hoffmann, H.H., Bhattacharyya, S., Alamares, J.G., Tscherne, D.M., Ortigoza, M.B., Liang, Y., Gao, Q., Andrews, S.E., Bandyopadhyay, S., De Jesus, P., Tu, B.P., Pache, L., Shih, C., Orth, A., Bonamy, G., Miraglia, L., Ideker, T., Garcia-Sastre, A., Young, J.A., Palese, P., Shaw, M.L., Chanda, S.K., 2010. Human host factors required for influenza virus replication. Nature 463, 813–817.

Lee, H.K., Tang, J.W., Kong, D.H., Loh, T.P., Chiang, D.K., Lam, T.T., Koay, E.S., 2013. Comparison of mutation patterns in full-genome A/H3N2 influenza sequences obtained directly from clinical samples and the same samples after a single MDCK passage. PLoS One 8 e79252.

Li, D., Saito, R., Suzuki, Y., Sato, I., Zaraket, H., Dapat, C., Caperig-Dapat, I.M., Suzuki, H., 2009a. In vivo and in vitro alterations in influenza A/H3N2 virus M2 and hemagglutinin genes: effect of passage in MDCK-SIAT1 cells and conventional MDCK cells. J. Clin. Microbiol. 47, 466–468.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S., 2009b. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Long, J.E., Drayson, M.T., Taylor, A.E., Toellner, K.M., Lord, J.M., Phillips, A.C., 2016. Morning vaccination enhances antibody response over afternoon vaccination: a cluster-randomised trial. Vaccine 34, 2679–2685.

Maurer-Stroh, S., Lee, R.T., Eisenhaber, F., Cui, L., Phuah, S.P., Lin, R.T., 2010. A new common mutation in the hemagglutinin of the 2009 (H1N1) influenza A virus. PLoS Curr 2, RRN1162.

McCrone, J.T., Woods, R.J., Martin, E.T., Malosh, R.E., Monto, A.S., Lauring, A.S., 2018. Stochastic processes constrain the within and between host evolution of influenza virus. Elife 7.

McGinnis, J., Laplante, J., Shudt, M., George, K.S., 2016. Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. J. Clin. Virol. 79, 44–50.

Medina, R.A., Stertz, S., Manicassamy, B., Zimmermann, P., Sun, X., Albrecht, R.A., Uusi-Kerttula, H., Zagordi, O., Belshe, R.B., Frey, S.E., Tumpey, T.M., Garcia-Sastre, A., 2013. Glycosylations in the globular head of the hemagglutinin protein modulate the virulence and antigenic properties of the H1N1 influenza viruses. Sci. Transl. Med. 5 187ra170.

Memoli, M.J., Czajkowski, L., Reed, S., Athota, R., Bristol, T., Proudfoot, K., Fargis, S., Stein, M., Dunfee, R.L., Shaw, P.A., Davey, R.T., Taubenberger, J.K., 2015. Validation of the wild-type influenza A human challenge model H1N1pdMIST: an A(H1N1) pdm09 dose-finding investigational new drug study. Clin. Infect. Dis. 60, 693–702.

Memoli, M.J., Hrabal, R.J., Hassantoufighi, A., Eichelberger, M.C., Taubenberger, J.K., 2010. Rapid selection of oseltamivir- and peramivir-resistant pandemic H1N1 virus during therapy in 2 immunocompromised hosts. Clin. Infect. Dis. 50, 1252–1255.

Memoli, M.J., Shaw, P.A., Han, A., Czajkowski, L., Reed, S., Athota, R., Bristol, T., Fargis, S., Risos, K., Powers, J.H., Davey Jr., R.T., Taubenberger, J.K., 2016. Evaluation of antihemagglutinin and antineuraminidase antibodies as correlates of protection in an influenza A/H1N1 virus healthy human challenge model. MBio 7 e00417-00416.

Miotto, O., Heiny, A.T., Albrecht, R., Garcia-Sastre, A., Tan, T.W., August, J.T., Brusic, V., 2010. Complete-proteome mapping of human influenza A adaptive mutations: implications for human transmissibility of zoonotic strains. PLoS One 5 e9025.

Mishin, V.P., Baranovich, T., Garten, R., Chesnokov, A., Abd Elal, A.I., Adamczyk, M., LaPlante, J., St George, K., Fry, A.M., Barnes, J., Chester, S.C., Xu, X., Katz, J.M., Wentworth, D.E., Gubareva, L.V., 2017. A pyrosequencing-based approach to high-throughput identification of influenza A(H3N2) virus clades harboring antigenic drift variants. J. Clin. Microbiol. 55, 145–154.

Miyake, Y., Keusch, J.J., Decamps, L., Ho-Xuan, H., Iketani, S., Gut, H., Kutay, U., Helenius, A., Yamauchi, Y., 2019 Apr. Influenza virus uses transportin 1 for vRNP debundling during cell entry. Nat Microbiol 4 (4), 578–586. https://doi.org/10.1038/s41564-018-0332-2. Epub 2019 Jan 28.

Morens, D.M., Taubenberger, J.K., Fauci, A.S., 2010. The 2009 H1N1 pandemic influenza virus: what next? MBio 1.

Morlighem, J.E., Aoki, S., Kishima, M., Hanami, M., Ogawa, C., Jalloh, A., Takahashi, Y., Kawai, Y., Saga, S., Hayashi, E., Ban, T., Izumi, S., Wada, A., Mano, M., Fukunaga, M., Kijima, Y., Shiomi, M., Inoue, K., Hata, T., Koretsune, Y., Kudo, K., Himeno, Y., Hirai, A., Takahashi, K., Sakai-Tagawa, Y., Iwatsuki-Horimoto, K., Kawaoka, Y., Hayashizaki, Y., Ishikawa, T., 2011. Mutation analysis of 2009 pandemic influenza A(H1N1) viruses collected in Japan during the peak phase of the pandemic. PLoS One 6 e18956.

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., Kanaya, S., 2011. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 39, e90.

Nunez, I.A., Carlock, M.A., Allen, J.D., Owino, S.O., Moehling, K.K., Nowalk, P., Susick, M., Diagle, K., Sweeney, K., Mundle, S., Vogel, T.U., Delagrave, S., Ramgopal, M., Zimmerman, R.K., Kleanthous, H., Ross, T.M., 2017. Impact of age and pre-existing influenza immune responses in humans receiving split inactivated influenza vaccine on the induction of the breadth of antibodies to influenza A strains. PLoS One 12 e0185666.

Pan, C., Cheung, B., Tan, S., Li, C., Li, L., Liu, S., Jiang, S., 2010. Genomic signature and mutation trend analysis of pandemic (H1N1) 2009 influenza A virus. PLoS One 5, e9549.

Park, J.K., Han, A., Czajkowski, L., Reed, S., Athota, R., Bristol, T., Rosas, L.A., Cervantes-Medina, A., Taubenberger, J.K., Memoli, M.J., 2018. Evaluation of preexisting anti-hemagglutinin stalk antibody as a correlate of protection in a healthy volunteer challenge with influenza A/H1N1pdm virus. MBio 9.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5 e9490.

Reid, A.H., Fanning, T.G., Hultin, J.V., Taubenberger, J.K., 1999. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. Proc. Natl. Acad. Sci. U. S. A. 96, 1651–1656.

Reid, A.H., Fanning, T.G., Janczewski, T.A., Lourens, R.M., Taubenberger, J.K., 2004. Novel origin of the 1918 pandemic influenza virus nucleoprotein gene. J. Virol. 78, 12462–12470.

Roedig, J.V., Rapp, E., Hoper, D., Genzel, Y., Reichl, U., 2011. Impact of host cell line adaptation on quasispecies composition and glycosylation of influenza A virus hemagglutinin. PLoS One 6 e27989.

Runstadler, J.A., Happ, G.M., Slemons, R.D., Sheng, Z.M., Gundlach, N., Petrula, M., Senne, D., Nolting, J., Evers, D.L., Modrell, A., Huson, H., Hills, S., Rothe, T., Marr, T., Taubenberger, J.K., 2007. Using RRT-PCR analysis and virus isolation to determine the prevalence of avian influenza virus infections in ducks at Minto Flats State Game Refuge, Alaska, during August 2005. Arch. Virol. 152, 1901–1910.

Schanen, B.C., De Groot, A.S., Moise, L., Ardito, M., McClaine, E., Martin, W., Wittman, V., Warren, W.L., Drake 3rd, D.R., 2011. Coupling sensitive in vitro and in silico techniques to assess cross-reactive CD4(+) T cells against the swine-origin H1N1 influenza virus. Vaccine 29, 3299–3309.

Sherman, A.C., Mehta, A., Dickert, N.W., Anderson, E.J., Rouphael, N., 2019. The future of flu: a Review of the human challenge model and systems Biology for advancement of influenza vaccinology. Front Cell Infect Microbiol 9, 107.

Smorodintseff, A.A., Tushinsky, M.D., Drobyshevskaya, A.I., Korovin, A.A., Osetroff, A.I., 1937. Investigation on volunteers infected with the influenza virus. Am. J. Med. Sci. 194, 159–170.

Squires, R.B., Noronha, J., Hunt, V., Garcia-Sastre, A., Macken, C., Baumgarth, N., Suarez, D., Pickett, B.E., Zhang, Y., Larsen, C.N., Ramsey, A., Zhou, L., Zaremba, S., Kumar, S., Deitrich, J., Klem, E., Scheuermann, R.H., 2012. Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. Influenza Other Respir Viruses 6, 404–416.

Tan, G.S., Krammer, F., Eggink, D., Kongchanagul, A., Moran, T.M., Palese, P., 2012. A pan-H1 anti-hemagglutinin monoclonal antibody with potent broad-spectrum efficacy in vivo. J. Virol. 86, 6179–6188.

Tan, P.T., Heiny, A.T., Miotto, O., Salmon, J., Marques, E.T., Lemonnier, F., August, J.T., 2010. Conservation and diversity of influenza A H1N1 HLA-restricted T cell epitope candidates for epitope-based vaccines. PLoS One 5 e8754.

Taubenberger, J.K., Kash, J.C., 2010. Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe 7, 440–451.

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Vavricka, C.J., Li, Q., Wu, Y., Qi, J., Wang, M., Liu, Y., Gao, F., Liu, J., Feng, E., He, J., Wang, J., Liu, H., Jiang, H., Gao, G.F., 2011. Structural and functional analysis of laninamivir and its octanoate prodrug reveals group specific mechanisms for influenza NA inhibition. PLoS Pathog. 7 e1002249.

Wang, D., Yang, L., Zhu, W., Zhang, Y., Zou, S., Bo, H., Gao, R., Dong, J., Huang, W., Guo, J., Li, Z., Zhao, X., Li, X., Xin, L., Zhou, J., Chen, T., Dong, L., Wei, H., Li, X., Liu, L., Tang, J., Lan, Y., Yang, J., Shu, Y., 2016a. Two outbreak sources of influenza a (H7N9) viruses have been established in China. J. Virol. 90, 5561–5573.

Wang, S., Chi, X., Wei, H., Chen, Y., Chen, Z., Huang, S., Chen, J.L., 2014. Influenza A virus-induced degradation of eukaryotic translation initiation factor 4B contributes to viral replication by suppressing IFITM3 protein expression. J. Virol. 88, 8375–8385.

Wang, W., Sun, X., Li, Y., Su, J., Ling, Z., Zhang, T., Wang, F., Zhang, H., Chen, H., Ding, J., Sun, B., 2016b. Human antibody 3E1 targets the HA stem region of H1N1 and H5N6 influenza A viruses. Nat. Commun. 7, 13577.

Wu, N.C., Zost, S.J., Thompson, A.J., Oyen, D., Nycholat, C.M., McBride, R., Paulson, J.C., Hensley, S.E., Wilson, I.A., 2017. A structural explanation for the low effectiveness of the seasonal influenza H3N2 vaccine. PLoS Pathog. 13 e1006682.

Xiao, Y., Nolting, J.M., Sheng, Z.M., Bristol, T., Qi, L., Bowman, A.S., Taubenberger, J.K., 2018a. Design and validation of a universal influenza virus enrichment probe set and its utility in deep sequence analysis of primary cloacal swab surveillance samples of wild birds. Virology 524, 182–191.

Xiao, Y.L., Ren, L., Zhang, X., Qi, L., Kash, J.C., Xiao, Y., Wu, F., Wang, J., Taubenberger, J.K., 2018b. 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.

Xu, L., Bao, L., Zhou, J., Wang, D., Deng, W., Lv, Q., Ma, Y., Li, F., Sun, H., Zhan, L., Zhu, H., Ma, C., Shu, Y., Qin, C., 2011. Genomic polymorphism of the pandemic A (H1N1) influenza viruses correlates with viral replication, virulence, and pathogenicity in vitro and in vivo. PLoS One 6 e20698.

Xu, X., Zhu, X., Dwek, R.A., Stevens, J., Wilson, I.A., 2008. Structural characterization of the 1918 influenza virus H1N1 neuraminidase. J. Virol. 82, 10493–10501.

Zhou, B., Lin, X., Wang, W., Halpin, R.A., Bera, J., Stockwell, T.B., Barr, I.G., Wentworth, D.E., 2014. Universal influenza B virus genomic amplification facilitates sequencing, diagnostics, and reverse genetics. J. Clin. Microbiol. 52, 1330–1337.