Identification of H3N2 NA and PB1-F2 genetic variants and their association with disease symptoms in the 2014-15 influenza season

Deena Blumenkrantz,[1] Thomas Mehoke,[5] Justin Hardick,[2] Kathryn Shaw-Saliba,[1,3] Harrison Powell,[1] , Hsuan Liu,[1] Elizabeth Macias,[7] Jared Evans,[5] Mitra Lewis,[2] Rebecca Medina,[2] Lauren M. Sauer,[2] Andrea Dugas,[2] Anna DuVal,[3] Andrew P Lane,[4] Charlotte Gaydos[2,3], Richard Rothman[2,3], Peter Thielen[5], Andrew Pekosz[1*]

[1]W. Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, [2]Department of Emergency Medicine, [3]Department of Medicine, Division of Infectious Diseases, and [4]Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland; [5]Research and Exploratory Development Department, Johns Hopkins Applied Physics Laboratory, Laurel, Maryland; and [7]Epidemiology Laboratory Service, United States Air Force School of Aerospace Medicine, Wright-Patterson Air Force Base, Ohio

*Corresponding Author contact information: Andrew Pekosz, Ph.D., apekosz1@jhu.edu, 410-502-9306

Conclusions

In 2014-15, H3N2 antigenic drift variants of HA clade 3C.2a were the dominant influenza virus strain identified in the Johns Hopkins Center of Excellence in Influenza Research and Surveillance (JH CEIRS) network.

H3N2 antigenic drift variants (HA clades) and additional mutations in other gene segments were determined by direct sequencing of human nasal swabs or washes.

Virus isolation on hNECs was more efficient than isolation on MDCK cells.

An NA mutation introducing a putative glycosylation site co-segregated with a PB1-F2 mutation (H75P).

Viruses with the NAgly+ H75P genotype were associated with increased reporting of some clinical symptoms.

Abstract

The 2014-15 influenza season saw the emergence of an antigenic drift variant of H3N2 which formed the 3C.2a HA clade. Nasopharyngeal swabs from patients with confirmed influenza A virus infection were used for whole viral genome sequencing to make HA clade calls and identify mutations in the NA (a new glycosylation site, NAg+) and PB1-F2 (H75P mutation). The mutations in NA and PB1-F2 that were present in a subset of clade 3C.2a viruses (NAg+F2P) which came to dominate in the subsequent influenza season. The NAg+F2P genotype viruses were associated with increased wheezing and shortness of breath in infected patients. The use of primary human nasal epithelial cell cultures allowed for more efficient isolation of H3N2 viruses and identified replication differences between H3N2 genotypes. The use of sequencing of clinical samples combined with *in vitro* analysis of virus replication provided a unique system for identifying and characterizing novel variants of influenza A virus.

Keywords : Antigenic drift, H3N2, influenza, symptoms, human, NA, PB1-F2

Introduction

Human influenza epidemics are largely influenced by the emergence and circulation of seasonal viral variants containing mutations that evade pre-existing population immunity (Bedford et al., 2015). The emergence and circulation of influenza virus variants presents multiple challenges including optimizing surveillance and vaccine effectiveness, as well as guiding other key public health decision-making in response to emerging influenza epidemics and pandemics (Bedford et al., 2015; Flannery et al., 2016; Huzly et al., 2016). The seasonal H3N2 virus variants that became dominant during the 2014-2015 influenza season were associated with an increase in outpatient influenza-like illness hospital visits, hospitalizations, and deaths attributable to pneumonia and influenza (Appiah et al., 2015; Epperson et al., 2014). Increased disease severity was also detected in other countries across the northern hemisphere (Broberg et al., 2015; Donadeu et al., 2015). The increased number of cases correlated with a marked diminution of serum hemagglutination inhibition (HI) and neutralizing (NT) antibodies against seasonal virus isolates (Flannery et al., 2016; Skowronski et al., 2016) which was attributed to discreet changes in HA antigenic sites (Chambers et al., 2015; Stucker et al., 2015), primarily through a novel glycosylation site in the H3 hemagglutinin (HA) protein .

The dominant circulating strains of the 2014-15 season largely fell into HA clades 3C.2a and 3C.3a. Furthermore, detailed sequence analysis and protein modeling indicated that mutations in the HA and neuraminidase (NA) genes could be linked to escape from neutralizing antibodies (Broberg et al., 2015; Chambers et al., 2015; Flannery et al., 2016; McCauley et al., 2015; Mostafa et al., 2016; Skowronski et al., 2016; Stucker et al., 2015).

It is critical for research and public health laboratories to have the methods in place for rapid reliable detection, recovery, and characterization of influenza seasonal drift variants. These capabilities inform decision making regarding vaccine composition, and the public health systems ability to respond according to severity of an ongoing epidemic or pandemic (Flannery et al., 2016; Pebody et al., 2015). In this report, we illustrate the utility of whole influenza virus genome sequencing and characterization of virus isolates on primary human nasal epithelial cell (hNEC) cultures to identify novel genetic variants of influenza viruses that altered virus replication and symptom severity during the 2014-15 influenza season. Genetic changes in the

NA and PB1-F2 proteins were identified and correlated with a higher incidence of wheezing and shortness of breath reported by infected individuals. Finally, efficient virus isolation and characterization was enabled through the use of differentiated human nasal epithelial cell (hNEC) cultures. While changes in the HA protein may have led to increased numbers of influenza cases in 2014-15, our data suggests that changes in other viral genes may have contributed to increased disease severity.

Materials and Methods

Sample collection and ethics

The human subjects protocol was reviewed and approved by the Johns Hopkins School of Medicine Institutional Review Board (IRB00052743) This protocol allowed for collection of residual nasopharyngeal swabs (NPS) from influenza positive patients at three clinical sites (Johns Hopkins Hospital in Baltimore, Howard County General Hospital in Columbia, and Surburban Hospital in Bethesda) and collection of related, de-identified patient data from the electronic health records: including the for variables listed in Table 1, including symptoms (CDC, 2016a) and co-morbidities that elevate the risk of increased morbidity and mortality among individuals infected with influenza virus (CDC, 2016b). The majority of specimens were from pediatric (under 18 years of age) and elderly (over 64) individuals, which represent the groups at increased risk for severe seasonal influenza disease (Luk et al., 2001; Morens et al., 2010; Skowronski et al., 2010). Symptoms were not reported for children under the age of two. Specimens from patients who had severe disease (as characterized by one of the following: supplemental oxygen use, Intensive Care Unit stay, or death) were selected as were samples of patients with known vaccination status. Specimens from age and sex matched individuals, who did not have characteristics of severe disease, were also included.

An IRB-exempt protocol from the same IRB allowed for access to residual nasopharyngeal wash specimens from de-identified patients of the Wright-Patterson Air Force Base (WPAFB). The WPAFB is the central repository for disease specimens from service men and women from around the country. A set of 16 samples that were collected before September from the WPAFB and grew to titers above 3 Log TCID50 in hNECs were selected for sequence analysis. Patient-level demographic and clinical data were not available for these specimens.

NPS and nasopharyngeal wash specimens were collected from each patient and placed in 3 ml of viral transport medium (MicroTest M4RT; Remel, Lenexa, KS, USA). All the specimens were aliquoted and stored at -70°C.

Sample processing for genomic analysis

NPS samples were centrifuged for five minutes at 16,000xg to remove debris. Clarified supernatants were purified per manufacturer's recommendations with the Ambion MagMax Viral RNA purification reagent, with the only protocol modification being omission of carrier RNA to the viral lysis enhancer reagent. RNA was resuspended in 20uL of supplied RNAse free elution buffer. Ambion Turbo DNA-free reagent (Life Technologies) was used to deplete genomic DNA. Manufacturer's instructions for stringent DNAse treatments were followed: 2U of enzyme was added at the beginning of treatment and after 30 minutes of incubation at 37°C. Total duration of DNA depletion was one hour, after which DNAse inactivation reagent was added at 0.2 volumes.

Reverse transcription of viral RNA was carried out with Superscript III (Invitrogen) per manufacturer's instructions for gene specific priming of low input samples (primers in Supplementary Table X). Samples were prepared using either a traditional multi-segment PCR approach (Zhou et al., 2009) or a preamplification-free approach, which we term amplification-limited viral RNA sequencing (vRNAseq). Briefly, primers and dNTPs were added to RNA and annealed by heating to 65°C for five minutes, then allowed to return to 4°C for five minutes. 400U of enzyme was then used in a total reaction volume of 40uL, and reverse transcription was carried out at 50°C. Upon completion, second strand synthesis was carried out at 16°C for 2.5 hours using mRNA Second Strand Synthesis reagent, which is comprised of *E. coli* DNA Pol I, E. coli DNA ligase, RNAse H (NEB), in a total volume of 80µL. Double stranded viral cDNA was purified using one volume of Ampure XP per manufacturer's instructions (Beckman Coulter). Once beads were completely dry, samples were eluted in 7.5µL of nuclease free water for five minutes prior to collection. One microliter was used for quantification with the High Sensitivity Qubit reagent according to manufacturer instructions (Invitrogen).

Nextera XT DNA library preparation was carried out per manufacturer's instructions (Illumina). For samples that resulted in less than one nanogram or unquantifiable amounts of DNA by picogram, library preparation was carried out with the maximum amount of DNA available in 5uL. After tagmentation and addition of sequencing adapters by 12 cycles of PCR, samples were purified with one volume of Ampure XP and eluted in 10µL of nuclease free water. One microliter of the resulting library was then used for quantification with High

Sensitivity Qubit reagents (Invitrogen), and multiplexed libraries were then pooled equi-mass, assuming an even size distribution between samples.

Sequence data analysis

Raw reads from the Illumina MiSeq sequencer were first filtered using Trim Galore! [http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/] using an adapter sequence of CTGTCTCTTATACACATCT and a quality score cutoff of 30. QC-trimmed reads were then individually classified using Kraken [https://github.com/DerrickWood/kraken] against a custom influenza-specific kraken database containing all complete sequences in EpiFlu with a modified taxonomy, which for flu types A, B, and C, identifies the likely segment, subtype, host, and year for each read. The most likely subtype for each sample was identified from the Kraken results, and 85 H3N2 samples were pulled out for further analysis. QC-trimmed reads for all 85 H3N2-identified samples were aligned to the A/Victoria/361/2011 reference sequence, and consensus FASTA sequences were generated for each sample. From these consensus FASTA sequences, 63 samples had at least 85% coverage of the HA segment and were used for subsequent phylogenetic analysis.

Reference sequences for the 2014-2015 influenza season were obtained from the GISAID EpiFlu database on October 22, 2016, by downloading all human H3N2 isolate sequences with a collection date between 2014-10-01 and 2015-05-31. In total, there were 5,642 unique EpiFlu ID numbers across 5,338 unique strain names. A subset of EpiFlu sequences was selected in two steps. First strains that contained exactly one of each of the 8 genome segments were selected, then for strain names that contained more than one EpiFlu ID number, only the strain with the lowest ID number was chosen. Sequences for the A/Victoria/361/2011 and A/Texas/50/2012 vaccine strains were obtained from NCBI BioProject PRJNA205638. The ultimate subset contained full genomes for 1,477 H3N2 strains. The same method was applied to select isolates for analysis from the 2013-2014 and 2015-2016 seasons, which led to 99 and 1140 isolates, respectively. The number of isolates from the 2013-14 season was an order of magnitude lower than the numbers for the other seasons because sequences for genes other than HA were often not available.

Phylogenetic analysis

Data from each segment was combined into a segment-specific FASTA file for the 1,478 EpiFlu reference strains, the 63 sequenced samples with over 85% coverage in each segment, and the A/Texas/50/2012 vaccine strain, for a total of 1,542 isolates. Each segment was then aligned using MAFFT [http://mafft.cbrc.jp/alignment/software/], all deletions ('-') were replaced with 'N's, and all sequences were padded with 'N's to the length of the vaccine strain sequence. Any insertions with respect to the vaccine sequence were removed (any positions in the vaccine sequence that now contain 'N's), and all sequences were trimmed to the length of the vaccine sequence. Phylogenetic trees were created for each segment using the nucleotide option of FastTree [http://www.microbesonline.org/fasttree/], and trees were then re-rooted to the vaccine strain using the nw_reroot tool from the Newick utilities package [https://github.com/xflouris/newick-tools]. Trees were viewed and colored using FigTree [http://tree.bio.ed.ac.uk/software/figtree/] and HA clades were identified by adding known clade references to these trees and identifying branches for the 3C.2a, 3C.3, 3C.3a, and 3C.3b clades. From these branch identifications, a set of rules was developed (see below) so that a clade label could be added to each strain label. All clade information was added to the name of each strain name before tree generation, and then nodes were colored by searching for a specific clade text and coloring all nodes in that selection.

Clade and genotype identification

Sequence data from each segment of the 1,542 uniquely named isolates were uploaded into a database using R. HA clades were identified using rules that selected sequences by amino acids (AAs) that associated with branches (https://hackmd.io/CwVgHAzAhgnAxgRgLQhiAZk4ATBMkwDsAbIUlFOsGMAEzHEAMEYQA===?view). The defining residues used were: 159Y for 3C.2a; 138S or 159S for 3C.3a; 157S for 3C.3b; 142R for 3C.2; 3L and 128A for 3C.3; and 3I and 128T for 3C.2. H3 numbering was used for all positions. Using these rules, all 1,542 strains were labeled according to the branch they fell on with the following exceptions: A/Brazil/9517/2015 is labeled 3C.3a but falls on the 3C.3b

branch; A/Manitoba/RV3623/2014 and A/Nevada/30/2014 are labeled 3C.3b but fall on the 3C.3 branch; 01-13-P-0057 is labeled uncertain but falls on the 3C.2a branch; and 01-90-P-0102 is labeled 3C.3 but falls on the 3C.3b branch.

Consensus sequences of the twelve proteins PB2, PB1, PA, HA, NP, NA, M1, NS1 (primary open reading frames (ORFs)) and PB1-F2, PAX, M2, and NEP (alternative ORFs) were generated for each isolate that had over 85% sequence coverage and all of the reference strains. Sequences of the eight primary ORF proteins were generated by translation from the first ATG in each gene to the first stop codon. Sequences of the four alternative ORF proteins were generated with the following set of rules: translation of PB1-F2 was programmed to start at the 4th ATG of segment 2 and stop at the first in frame stop codon ; translation of PAX started from the first start codon in segment 3, included the first 573 nucleotides, then shifted +1 to nt 575 and carried on to the next stop codon in that frame; M2 was translated from the first start codon of the segment 7 through 51 nts, joined to nt 715 and stopped at the first stop codon in frame 2; NEP was translated from the first start codon in segment 8 through the 30th nt, joined with base 503 and stopped at the in-frame stop codon.

Genetic variation among clade 3C.2a isolates was analyzed further. Amino acid alignments were generated using MAFFT and the frequency of each AA residue at every location was determined using Excel (Microsoft). Instances where the consensus AA fell below 75% frequency were noted. Mutations associated with putative NA glycosylation signatures were identified by the residues at positions 245-247. Isolates were labeled as NAg+ if those residues were NAT or NAS and NAg- for all other sequences. Residues at AA position 75 within PB1-F2 were also identified for each isolate.

Cell culture and serum free media

Madin-Darby canine kidney cells (MDCK) were cultured as previously described (Grantham et al., 2010). MDCK-SIAT cells were kindly donated by Scott Hensley and cultured exactly the same as MDCK cells. Cultures of human primary, differentiated nasal epithelial cells (hNEC) were obtained from volunteers and differentiated at an air-liquid interface (ALI), as previously described (Fischer et al., 2015; Ramanathan et al., 2009).

Virus isolation

Virus isolation with MDCK cells was performed by incubating 100µl of NPS on twice-PBS-washed cells in 48-well plates for one hour at 32°C with 5% CO_2 with agitation every 15 minutes. After inoculation, NPS media was aspirated and cells were incubated again at 32°C with 5% CO_2 in infection media (Dulbecco's Modified minimal media; 0.3% bovine serum albumin) with 5 µg/ml N-acetyl trypsin (NAT). Media was collected and replaced daily.

Virus isolation with hNEC cultures was performed by incubating 100µl of NPS on the twice-PBS-washed apical surface of hNEC cells in 12-well transwell plates (Corning) for two hours at 32°C with 5% CO_2. Inoculum was removed and cells were returned to incubate at 32°C with 5% CO_2 for up to 7 days. The hNEC cultures were maintained with media on the basolateral side, which was changed every 2-3 days. Apical washes, acquired by adding 300ul infection media to the apical surface of hNECs and incubating for 10 minutes at 32°C with 5% CO_2, were collected daily. All samples were stored at -70C.

Infectious virus quantification

The infectious virus titers were determined by calculating the 50% tissue culture infectious dose (TCID50) using the Reed and Muench method (McCown and Pekosz, 2005; Reed and Muench, 1938)..

Virus generation and growth analysis

Infected-hNEC culture supernatants (4 dpi of hNECs) were inoculated into MDCK cells with a multiplicity of infection (MOI) of 0.01 or 0.001 TCID50 units per cell and incubated in infection media with 5µg/ml NAT for 5 days at 32°C to make virus stocks. A/Columbia/P0041/2014 (Co14) has the NAg-F2H genotype while A/Bethesda/P0055/2015 (Be15) has the NAg+F2P genotype. Infectious virus was quantified by $TCID_{50}$. HA, NA, and PB1 genes of amplified virus stocks were sequenced by Sanger sequencing. The Sanger sequences of virus stocks were compared to the WGS consensus sequences and the HA of both Co14 and Be15 were found to have the non-synonymous nucleotide mutation C478A, which encoded the

mutation T160K (H3 Numbering). This mutation resulted in the loss of a predicted glycosylation site at position 158 in the HA protein.

For MDCK and MDCK-SIAT cell growth curves, cells in 24-well plates were washed twice with PBS containing Ca2+ (0.9mM) and Mg2+ (0.5mM) (Sigma) and infected with MOI 0.001 virus, diluted in 100 ul of infection media. After 1 hour inoculation, the supernatant was removed, cells washed 3 times with PBS with calcium and magnesium and incubated in infection media with 2.5 ug/ml NAT at 32°C with 5% CO_2. Samples were collected at indicated times by removing and replacing the media. Supernatants were stored at -70°C and titrated by $TCID_{50}$.

For hNEC growth curves, the apical surface of 24-well hNEC cultures was washed twice with PBS and infected with MOI 0.01 virus in 100ul in infection media. hNECs were incubated with the inoculum for 2 hours at 32°C with 5% CO_2. Inoculum was removed and the apical surface was washed twice with PBS before cells were returned to the incubator. Apical wash samples were collected, at indicated times, by incubating 100 ul of infection media on the apical surface for 10 minutes at 32°C as previously described (Fischer et al., 2015). Supernatants were stored at -70°C and infectious virus quantified by $TCID_{50}$.

Each virus was inoculated into 4 wells of each cell type and growth curves were performed on two separate occasions, for a total of 8 wells per virus, per cell type, per time point. The average viral titers from the 8 equally treated wells were graphed for virus comparisons.

Plaque assay

MDCK cells, at approximately 100% confluence were washed twice with PBS with calcium and magnesium. Virus was serially diluted 10-fold in infection media and 250 uL of each dilution was added to cells. The plates were incubated for one hour at 32°C in 5% CO_2 with agitation, then inoculum was removed and cells were overlaid with 1X MEM with the addition of .3% BSA, 10mM Pen/Strep, 5mM Glutamine, 1% agarose (Invitrogen), 1 mM Hepes (Gibco), and 5ug/ml NAT and incubated at 32°C in 5%CO_2 for 72 hours, after which cells were fixed with 4% paraformaldehyde (Fisher) and stained with Napthol Blue Black (Sigma). One image per well

was collected using an Olympus OM-D E-M5 Mark II digital camera. The scale for all images was determined by photographing a ruler. Images were processed in ImageJ v1.49 Plaque images were cropped to remove area outside of the well, then they were turned into 8-bit images. Next, plaques were circled manually using the selection tool. Finally, plaques were analyzed using the ImageJ analyze particle size command with a size minimum of 200 pixels and a circularity of 0.10-1.00. In total 40-60 plaques (between 10-20 plaques per well) were analyzed per virus.

Statistical analysis.

Differences between the average viral titers of isolates grown on MDCK cells or hNECs were determined by paired, two-tailed, t-test. Virus replication differences in low MOI growth curves were determined with repeated measures two-way ANOVA followed by a Bonferonni post-test. Differences in clinical and demographic data were determined using a Chi-squared test.

Results

Clinical characteristics of patients

Ninety four influenza positive NPS specimens with demographic and clinical data were collected from three hospitals in the Johns Hopkins Health System (JHMI)(Table 1). There were 54% females and 46% males. Vaccination status was known for 36% of patients, of which 73% were vaccinated. 61% of patients were treated as outpatients and 10% required inpatient care. 30% of patients required intensive care and most of those (97%) required supplemental oxygen. 5% of the patients died. 57% of patients had one or multiple co-morbidities. See Table 1 for a more detailed breakdown of symptoms and co-morbidities. The JHMI samples were supplemented with 16 nasal wash specimens from the Wright-Patterson Air Force Base (WPAFB) repository. Clinical data for the WPAFB patients was not accessible.

Sequence characterization of virus

Whole genome sequencing (WGS) was performed using an amplification strategy designed to specifically detect influenza genomic sequences. A combination of Illumina sequencing and Kraken analyses were used to generate and categorize consensus sequences for the eight influenza virus gene segments in each specimen. Ninety three specimens were processed using WGS (77 from JHMI and 16 from WPAFB)

The influenza virus type A was determined for 86% (66/77) of the JHMI samples and 88% (14/16) of the WPAFB samples (Table 2). The influenza H3N2 subtype was determined for 84% (65/77), and the HA clade was determined for 77% (59/77), of JHMI isolates. The influenza subtype and clade were determined for 88% (14/16) of the WPAFB samples.

HA clade distribution

WPAFB samples were collected earlier in the 2014-15 influenza season and across a broader geographic area than the JHMI samples. The majority of WPAFB (57%) and JHMI (93%) isolates belonged to the H3N2 clade 3C.2a (Figure 1 A,B). The remaining isolates were characterized as either 3C.3a or 3C.3b, with one WPAFB isolate characterized as a 3C.3. As a

percentage of the samples analyzed, there was a higher level of clade diversity among the WPAFB viruses compared to the JHMI isolates (Figure 1 A,B).

For comparison, the HA clade distributions of viral isolates with full genomes published on the EpiFlu database were determined for the winter seasons preceding, during, and after the season in which our samples were collected. While the 2013-14 influenza season had a diverse distribution of clades, the 2014-15 influenza season, as previously reported (Flannery et al., 2016), was dominated by clade 3C.2a viruses (Figure 1C). JHMI and WPAFB sequences were included in the analysis for the 2014-15 season for a total of 1551 isolates, of which 1179 (76%) belonged to clade 3C.2a (Figure 1D. 3C.2a clade viruses continued to dominate during the 2015-16 season (Figure 1 E). Taken together, the HA clade distribution of isolates from the JHMI and WPAFB collections were representative of that seen in the database of all samples sequenced during the winter season in 2014-15 even though the WPAFB samples were collected earlier in the season than the JHMI samples.

Identification and frequency of NA glycosylation and PB1-F2 AA 75 mutations

Further analysis of the influenza virus sequences revealed that over one third (38%; 20/53, Figure 1G) of JHMI isolates had a non-synonymous nucleotide mutation (ntG734A; open reading frame (ORF) numbering throughout). This mutation changed a Serine [S] residue to Asparagine [N], thereby creating a new putative N-linked glycosylation signal in NA at amino acid residues 245-247 (NAg+). In addition, 18 of the 20 isolates had another non-synonymous mutation (ntT739A), which changed the Serine [S] residue to Threonine [T] at amino acid 247 of NA. Since the putative glycosylation sequence is N-X-S/T, the second non-synonymous mutation in NA presumably would not affect the glycosylation state at this site. This glycosylation site was previously observed in about a third (27%; 3/11) of isolates (Mostafa et al., 2016), as well as in other studies of 2014-15 influenza viruses. Interestingly, the WPAFB samples did not contain the NAg+ mutations. Since these samples were collected earlier in the 2014-15 season, the data may indicate that the NAg+ mutation emerged later in the season (Figure 1F).

The only other non-synonymous mutation that occurred with greater than 25% frequency among the 3C.2a clade JHMI isolates was an A318C mutation in the PB1 gene, which resulted in

a change from Histidine to Proline at amino acid 75 (H75P) of the PB1-F2 ORF (F2P). The same 20 JHMI 3C.2a clade viruses that were NAg+ contained the F2P mutation while all other 3C.2a isolates lacked both mutations (NAg-F2H) (Figure 1G). None of the 3C.2a clade isolates from the WPAFB had the F2P genotype (Figure 1F). This data indicated that both the NAg+ and F2P genotypes fixed in clade 3C.2a viruses at approximately the same time.

The total clade 3C.2a influenza winter season isolates from 2013-2016 that had WGS were characterized into six genotypes, which were determined by the presence or absence of a glycosylation motif at residues 245-247 in the NA protein and the amino acid residue at position 75 in the PB1-F2 protein (Figure 1 H-J). Besides His [H] and Pro [P], the only other amino acid that was present at position 75 in PB1-F2 was Arginine [R]. In samples from 2013-14, only the NAg- F2H genotype was detected. The NAg-F2H was dominant during the 2014-15 winter season (75%) with the NAg+F2P genotype being the second most common genotype (18%) (Figure 1I). The frequency of NAg+F2P isolates grew to 68% of total clade 3C.2a isolates in 2015-16 while the NAg-F2H genotype shrank from 75% to 12%. The only other genotype to increase in frequency was NAg+F2H, which grew from 4% to 19% (Figure 1 I and J). Taken together, the data indicate that the NAg+F2P genotype rapidly became dominant among clade 3C.2a viruses (compare Figure 1 I to J), while the frequency of 3C.2a viruses remained relatively similar (compare Figure 1 D to E) between the 2014-15 and 2015-16 seasons. Furthermore, viruses with NAg+ accounted for 87% of isolates in the 2015-16 season (Figure 1J).

Evolution of NA glycosylation and PB1-F2_75P

HA, NA, and PB1 phylogenetic trees with the 1487 2014-15 winter season EpiFlu, JHMI, and WPAFB isolate sequences were generated to determine the phylogenetic relationships between genotypes (Figure 2). HA clades 3C.2a, 3C.3, 3C.3a, and 3C.3b were all represented in the HA phylogenetic tree (Figure 2A). The branches of the NA and PB1 phylogenetic trees use the HA clade color scheme, with the exception of viruses encoding the NAg+ or F2P mutations (salmon and burgundy in the NA and PB1 trees, respectively, figure 2B,C), which use the 3C.2a genotype color scheme from Figure 1. In the NA phylogenetic tree, the large set of salmon colored branches at the top indicates that most detected NAg+ isolates arose from a common

ancestor. In the PB1 phylogenetic tree, the large set of burgundy colored branches at the top similarly indicates that all detected F2P isolates arose from a common ancestor. The putative NA glycosylation site was mapped to the X-ray crystal structure of the NA protein from influenza A/Tanzania/205/2010 (H3N2)(Zhu et al., 2012), which shows that the site is on the surface of the NA protein and directly adjacent to the NA enzymatic site (Figure 2D). The location of residue 75 was mapped to the NMR structure of the C-terminus tail of the PB1-F2 protein from influenza A/Puerto Rico/8/1934 (H1N1)(Bruns et al., 2007) and it shows that the mutation site is about two-thirds of the way down the C-terminal domain (Figure 2E). Since the NAg+ and F2P genotypes appear together in most virus strains (Figure 1 I and J), we conclude that both genotypes fixed in the same common ancestor, leading to the dominance of this genotype in the 2015-16 clade 3C.2a viruses.

To confirm the genetic relatedness of the JHMI and WPAFB viruses, we performed a phylogenetic analysis on concatemerized genome segments from the JHMI and WPAFB samples (Figure 3A). The NAg+F2P genotype viruses all clustered together from a single branch of the tree (salmon color at top of tree), indicating that there was a common ancestor that acquired both the NAg+ and the F2P mutations. To determine if this result was the same in the general population of viruses we performed a similar phylogenetic analysis on all H3N2 viruses with full genome sequences available from the 2014-15 and 2015-16 seasons (Figure 3B). The NAg+F2P viruses clustered together, as was observed for the JHMI and WPAFB samples, providing further support for the presence of a common ancestor from which all the NAg+F2P genotype viruses derived. The locations of the NAg-F2P viruses (cyan branches) within the NAg+F2P viruses (salmon branches) indicate that the NAg-F2P genotype likely arose from mutations that reversed the NAg+ to the NAg- genotype in the NAg+F2P genotype lineage. The group of NAg+F2H viruses (plum) that branched from the trunk of the tree before the NAg+F2P group, and the other NAg+F2H virus locations suggest that the NA glycosylation may have been selected slightly before the H75P mutation and that the 75P mutation was only selected when NAg+ was paired with a specific genetic cassette.

The branching of the three NAg+F2R viruses (yellow) among small group of NAg-F2R viruses (magenta) indicates that a subset of F2R viruses created a background in which NAg+

was selected, but this was unrelated to the emergence of the NAg+F2P genotype. In the PB1 phylogenetic tree all F2P and all F2R viruses shared a single node (Figure 2C), indicating that, evolutionarily, these genes were closely related. In summary, it seems that the NAg+F2P genotype clade 3C.2a viruses descended from a common ancestor, which suggests a selective advantage of that genotype over other viruses in the 3C.2a clade.

Virus isolation on hNEC cultures was more successful than on MDCK cells

Virus isolation was attempted from 104 NPS specimens. Primary differentiated hNEC cultures were inoculated with all 104 NPS specimens while MDCK cells were inoculated with a subset of 96 specimens. Infectious virus titers were determined by $TCID_{50}$ assay for supernatants collected at four days post infection (dpi). Inoculation of hNEC cultures led to isolation of viruses from 66% of the NPS specimens while inoculation of MDCK cells only led to isolation of viruses from 23% of the specimens (Figure 4A), a nearly three-fold higher success rate. The median infectious virus titers were 3.6 $\log_{10}$ $TCID_{50}$/ml from hNEC cultures and 2.4 $\log_{10}$ $TCID_{50}$/ml, indicating the clinical isolates grew to higher infectious virus titers in hNECs compared to MDCK cells (Figure 4A). The higher virus isolation frequency and higher viral titers achieved with hNEC cultures were consistently observed, irrespective of the HA clade or genotype of virus isolated (Figure 4B). This data indicates that hNEC cultures, which represent the cells that influenza viruses encounter in the upper respiratory tract, are more efficient than MDCK cells at isolating virus from clinical specimens.

Virus growth and plaque morphology

Viruses representative of the two 3C.2a genotypes – A/Columbia/P0041/2014 (NAg-F2H) and A/Bethesda/P0055/2015 (NAg+F2P) – were expanded and assessed for differences in replication on three different cell types. Plaque assays were performed on MDCK cells and the two genotypes showed similar plaque morphologies (Figure 4C and D) and areas (Figure 4E). In low MOI growth curves, the two viruses grew to similar titers on MDCK cells (Figure 4F) and MDCK-SIAT cells (Figure 4G), which overexpress the α2,6-linked sialic acid utilized by human influenza viruses as a cellular receptor (Matrosovich et al., 2003). The NAg-F2H virus grew to

higher titers on hNEC cultures compared to NAg+ F2P (Figure 4H) and this difference was statistically significant at 24, 36, and 48 hpi. The replication difference in hNECs was transient, and both viruses reached the same peak titer at the same day post infection. Taken together, the data indicate the two genotypes replicate to similar extents in immortalized cell lines but show different replication kinetics in hNEC cultures, suggesting the NAg+ F2P genotype is slightly inhibited in its replication in primary respiratory epithelial cell cultures.

Viruses from the 3C.2a NAg+F2P genotype associated with breathing difficulties

The clinical and demographic information from 20 patients infected with the NAg+F2P genotype was compared to the 34 patients infected with clade 3C.2a NAg-F2H genotype viruses (Table 1). The only clinical factors that differed between the patient populations was an increased reporting of wheezing and shortness of breath in individuals infected with the NAg+F2P genotype. A multivariate logistic regression analysis was performed to determine if any underlying co-morbidities were responsible for the correlations. We found no association of co-morbidities with increased wheezing or shortness of breath.

Discussion

The 2014-15 Northern hemisphere influenza season had unusually high numbers of cases and an increase in disease severity. This was attributed to antigenic drift of the H3N2 HA protein mediated by a new glycosylation site at amino acid 160 which partially masked neutralizing antibody epitopes. WGS led to identification of a new, dominant genotype within HA clade 3C.2a that we called NAg+F2P. Viruses of this genotype had a putative glycosylation site on the rim of the NA sialidase pocket and the mutation H75P in the PB1-F2 protein. The 3C.2a H3N2 viruses with the NAg-F2H genotype were eventually outcompeted by NAg+ F2P viruses in the human population, despite the NAg+ F2P viruses showing slightly reduced infectious virus production in hNEC cultures. Infection with the NAg+ F2P viruses was associated with increased reporting of shortness of breath and wheezing.

Since both of the clade 3C.2a genotypes had the HA mutations associated with antigenic drift, factors outside of HA antibody pressure must have led to the dominance of the NAg+F2P genotype within the clade. The appearance of both mutations in a common ancestor and the continued maintenance of both mutations together suggests that both contributed to increased fitness.

The main function of the influenza NA protein is to cleave sialic acid; cleavage from cellular receptors leads to release of new virions and cleavage from decoy receptors allows virions to breach the mucosal barrier. Glycosylation of NA can assist in protein folding, prevent proteolytic cleavage, mask antigenic regions, and regulate the enzyme's catalytic activity (Chen et al., 2012; Wu et al., 2009). Antigenic escape epitopes of human seasonal N2 NA proteins have been mapped using monoclonal antibodies and molecular biology and mostly localize around the enzyme site (Gulati et al., 2002; Lentz et al., 1984). NA drift has also been linked to evasion of human polyclonal antibodies (Sandbulte et al., 2011). Recently, the NAg+ mutation has been demonstrated to contribute to H3N2 NA antigenic drift and also alters virus replication fitness on hNECs and NA enzymatic activity. It follows that the amino acid mutations S245N and S245T, which encode a glycosylation site near the sialidase rim of the 3C.2a NA, could have been selected to evade human immunity and may reduce NA enzymatic activity.

Furthermore, since the introduction of this novel NA genotype in the 2014/15 season nearly all circulating human H3N2 isolates contain this NA glycosylation site; highlighting its importance in evading humoral immunity regardless of associated fitness costs.

PB1-F2 is a virulence factor that can affect virus replication and apoptosis or morbidity in cell culture and animal models, respectively. The effects of PB1-F2 differ in a cell type- and virus strain-dependent manner. The location of the mutation in PB1-F2 has undergone some changes in human seasonal H3N2 viruses. In 1986, PB1-F2 acquired a R75H mutation, which fixed in circulating H3N2 viruses until 2014-15, when the described H75P mutation became dominant. Whether mutations at position 75 of PB1-F2 can alter virus replication or host responses to infection should be investigated further. PB1-F2 is located in the C-terminal portion of the protein and is part of a mitochondrial association domain (cite) further suggesting it may play a role in H3N2 virus replication, virulence, and host cell innate immune signaling.

Associations between different influenza A genotypes and altered disease phenotypes are expected across subtypes, but rarely detected within lineages. Human seasonal H3N2 infection is associated with a higher population-wide disease burden compared to H1N1, especially in the elderly. Within human H1N1 viruses, recent pandemic strains with the HA mutation D225G associated with infection of the lower airways, likely due to the increased binding of a2,3-linked sialic acid, and increased disease. However, these viruses have not outcompeted D225 encoding viruses, suggesting they lack some fitness for spread in the human population.

Isolation of influenza viruses in hNEC cultures allowed expansion of viruses from two thirds of positive specimens, three times more than the quarter that expanded after MDCK cell culture. This efficient isolation allowed for a greater group of virus genotypes to be studied more extensively in a faster timeline. The use of hNEC cultures to tease apart viral mutations that alter replication has been successful previously, suggesting that replication on primary epithelial cells does place unique pressures on influenza virus when compared to transformed or immortalized cell lines.

Our surveillance approach of WGS from nasal swabs, coupled with virus isolation and characterization on hNEC cultures identified a novel H3N2 genotype which resulted in an altered phenotype (reduced replication in hNEC cultures) and increased reporting of shortness of breath and wheezing. This type of deep surveillance for influenza at one site has the benefit of linking patient demographic and clinical data with detailed virological data. With recent advances in sequencing technology and infectious clone methods, we expect that the identification of influenza variants with altered disease potential or antigenic structure will soon be performed within a single season, which may have implications for patient treatment and public health efforts.

Acknowledgements

Thank you to Samantha Lycett (University of Edinburgh, UK) and Colin Russell (Cambridge University, UK) for sharing their phylogenetic tree branch coloring protocols. Thank you to Monica Galiano (Public Health England) for the discussion about correlations between genotype and phenotype. We thank Sabra Klein and Kimberley Davis and their laboratories for discussion and comments on the data. This work was supported by the NIH/NIAID Center of Excellence in Influenza Research and Surveillance contract HHS N272201400007C (Johns Hopkins University) and the Richard Eliasberg Family Foundation.

References

Appiah, G.D., Blanton, L., D'Mello, T., Kniss, K., Smith, S., Mustaquim, D., Steffens, C., Dhara, R., Cohen, J., Chaves, S.S., Bresee, J., Wallis, T., Xu, X., Abd Elal, A.I., Gubareva, L., Wentworth, D.E., Katz, J., Jernigan, D., Brammer, L., Centers for Disease Control and Prevention (CDC), 2015. Influenza activity - United States, 2014-15 season and composition of the 2015-16 influenza vaccine. MMWR. Morb. Mortal. Wkly. Rep. 64, 583–90.

Bedford, T., Riley, S., Barr, I.G., Broor, S., Chadha, M., Cox, N.J., Daniels, R.S., Gunasekaran, C.P., Hurt, A.C., Kelso, A., Klimov, A., Lewis, N.S., Li, X., McCauley, J.W., Odagiri, T., Potdar, V., Rambaut, A., Shu, Y., Skepner, E., Smith, D.J., Suchard, M. a., Tashiro, M., Wang, D., Xu, X., Lemey, P., Russell, C. a., 2015. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. Nature 523, 217–20.

Broberg, E., Snacken, R., Adlhoch, C., Beauté, J., Galinska, M., Pereyaslov, D., Brown, C., Penttinen, P., 2015. Start of the 2014/15 influenza season in Europe: drifted influenza A(H3N2) viruses circulate as dominant subtype. Euro Surveill. 20, 1–5.

Bruns, K., Studtrucker, N., Sharma, A., Fossen, T., Mitzner, D., Eissmann, A., Tessmer, U., Röder, R., Henklein, P., Wray, V., Schubert, U., 2007. Structural characterization and oligomerization of PB1-F2, a proapoptotic influenza A virus protein. J. Biol. Chem. 282, 353–363.

CDC, 2016a. Flu Symptoms & Complications [WWW Document]. www.CDC.gov. URL https://www.cdc.gov/flu/about/disease/complications.htm (accessed 1.1.16).

CDC, 2016b. People at High Risk of Developing Flu–Related Complications [WWW Document]. www.CDC.gov. URL https://www.cdc.gov/flu/about/disease/high_risk.htm (accessed 1.1.16).

Chambers, B.S., Parkhouse, K., Ross, T.M., Alby, K., Hensley, S.E., 2015. Identification of Hemagglutinin Residues Responsible for H3N2 Antigenic Drift during the 2014–2015 Influenza Season. Cell Rep. 1–6.

Chen, K.-F., Blyn, L., Rothman, R.E., Ramachandran, P., Valsamakis, A., Ecker, D., Sampath, R., Gaydos, C. a, 2011a. Reverse transcription polymerase chain reaction and electrospray ionization mass spectrometry for identifying acute viral upper respiratory tract infections. Diagn. Microbiol. Infect. Dis. 69, 179–86.

Chen, K.-F., Rothman, R.E., Ramachandran, P., Blyn, L., Sampath, R., Ecker, D.J., Valsamakis, A., Gaydos, C. a, 2011b. Rapid identification viruses from nasal pharyngeal aspirates in acute viral respiratory infections by RT-PCR and electrospray ionization mass spectrometry. J. Virol. Methods 173, 60–6.

Donadeu, M., Lightowlers, M.W., Fahrion, A.S., Kessels, J., Abela-Ridder, B., 2015. Review of the 2014-2015 influenza season in the northern hemisphere. Relev. Epidemiol. Hebd. 90, 281–96.

Epperson, S., Blanton, L., Kniss, K., Mustaquim, D., Steffens, C., Wallis, T., Dhara, R., Leon, M., Perez, A., Chaves, S.S., Elal, A.A., Gubareva, L., Xu, X., Villanueva, J., Bresee, J., Cox, N., Finelli, L., Brammer, L., 2014. Influenza activity - United States, 2013-14 season and composition of the 2014-15 influenza vaccines. MMWR. Morb. Mortal. Wkly. Rep. 63, 483–90.

Fischer, W.A., King, L.S., Lane, A.P., Pekosz, A., 2015. Restricted replication of the live attenuated influenza A virus vaccine during infection of primary differentiated human

nasal epithelial cells. Vaccine 33, 4495–4504.
Flannery, B., Zimmerman, R.K., Gubareva, L. V, Garten, R.J., Chung, J.R., Nowalk, M.P., Jackson, M.L., Jackson, L.A., Monto, A.S., Ohmit, S.E., Belongia, E.A., McLean, H.Q., Gaglani, M., Piedra, P.A., Mishin, V.P., Chesnokov, A.P., Spencer, S., Thaker, S.N., Barnes, J.R., Foust, A., Sessions, W., Xu, X., Katz, J., Fry, A.M., 2016. Enhanced genetic characterization of influenza A(H3N2) viruses and vaccine effectiveness by genetic group, 2014-2015. J. Infect. Dis. 2015, 1–23.
Grantham, M.L., Stewart, S.M., Lalime, E.N., Pekosz, A., 2010. Tyrosines in the influenza A virus M2 protein cytoplasmic tail are critical for production of infectious virus particles. J. Virol. 84, 8765–76.
Huzly, D., Korn, K., Bierbaum, S., Eberle, B., Falcone, V., Knöll, A., Steininger, P., Panning, M., 2016. Influenza A virus drift variants reduced the detection sensitivity of a commercial multiplex nucleic acid amplification assay in the season 2014/15. Arch. Virol. 161, 2417–2423.
Jeng, K., Massire, C., Zembower, T.R., Deyde, V.M., Gubareva, L. V, Hsieh, Y.-H., Rothman, R.E., Sampath, R., Penugonda, S., Metzgar, D., Blyn, L.B., Hardick, J., Gaydos, C. a, 2012. Monitoring seasonal influenza A evolution: rapid 2009 pandemic H1N1 surveillance with an reverse transcription-polymerase chain reaction/electro-spray ionization mass spectrometry assay. J. Clin. Virol. 54, 332–6.
Luk, J., Gross, P., Thompson, W.W., 2001. Observations on mortality during the 1918 influenza pandemic. Clin. Infect. Dis. 33, 1375–8.
Matrosovich, M., Matrosovich, T., Carr, J., Roberts, N.A., Klenk, H., 2003. Overexpression of the alpha-2,6-sialyltransferase in MDCK cells increases influenza virus sensitivity to neuraminidase inhibitors. J. Virol. 77, 8418–25.
McCauley, J.W., Daniels, R., Lin, Y.P., Xiang, Z., Gregory, V., Whittaker, L., Halai, C., Cross, K., Rattigan, A., Ermetal, B., Dai, M., 2015. Report prepared for the WHO annual consultation on the composition of influenza vaccine for the Northern Hemisphere 2015/16. WHO Influ. Cent. London.
McCown, M.F., Pekosz, A., 2005. The influenza A virus M2 cytoplasmic tail is required for infectious virus production and efficient genome packaging. J. Virol. 79, 3595–605.
Morens, D.M., Taubenberger, J.K., Harvey, H.A., Memoli, M.J., 2010. The 1918 influenza pandemic: lessons for 2009 and the future. Crit. Care Med. 38, e10-20.
Mostafa, A., Abdelwhab, E.M., Slanina, H., Hussein, M.A., Kuznetsova, I., Schüttler, C.G., Ziebuhr, J., Pleschka, S., 2016. Phylogenetic analysis of human influenza A/H3N2 viruses isolated in 2015 in Germany indicates significant genetic divergence from vaccine strains. Arch. Virol. 161, 1505–15.
Pebody, R.G., Warburton, F., Ellis, J., Andrews, N., Thompson, C., Wissmann, B. Von, Green, H.K., 2015. Low effectiveness of seasonal influenza vaccine in preventing laboratory-confirmed influenza in primary care in the United Kingdom : 2014 / 15 mid – season results. Euro Surveill 20, 1–7.
Ramanathan, M., Lee, W.-K., Dubin, M.G., Lin, S., Spannhake, E.W., Lane, A.P., 2009. Sinonasal epithelial cell expression of toll-like receptor 9 is decreased in chronic rhinosinusitis with polyps. Am. J. Rhinol. 21, 110–6.
Reed, L., Muench, H., 1938. A simple method of estimating fifty per cent endpoints. Am. J.

Epidemiol. 27, 493–497.
Rothman, R.E., Yang, S., Hardick, J., Gaydos, C. a, 2012. Harnessing genomic approaches for infectious disease diagnosis in emergency medicine: getting closer to prime time. Ann. Emerg. Med. 60, 621–3.
Skowronski, D.M., Chambers, C., Sabaiduc, S., De Serres, G., Winter, A.L., Dickinson, J.A., Krajden, M., Gubbay, J.B., Drews, S.J., Martineau, C., Eshaghi, A., Kwindt, T.L., Bastien, N., Li, Y., 2016. A perfect storm: Impact of genomic variation and serial vaccination on low influenza vaccine effectiveness during the 2014-2015 season. Clin. Infect. Dis. 63, 21–32.
Skowronski, D.M., Hottes, T.S., Janjua, N.Z., Purych, D., Sabaiduc, S., Chan, T., De Serres, G., Gardy, J., McElhaney, J.E., Patrick, D.M., Petric, M., 2010. Prevalence of seroprotection against the pandemic (H1N1) virus after the 2009 pandemic. CMAJ 182, 1–6.
Stucker, K.M., Schobel, S.A., Olsen, R.J., Hodges, H.L., Lin, X., Halpin, R.A., Fedorova, N., Stockwell, T.B., Tovchigrechko, A., Das, S.R., Wentworth, D.E., Musser, J.M., 2015. Haemagglutinin mutations and glycosylation changes shaped the 2012/13 influenza A(H3N2) epidemic, Houston, Texas. Euro Surveill. 20, 1–15.
Tang, Y.-W., Lowery, K.S., Valsamakis, A., Schaefer, V.C., Chappell, J.D., White-Abell, J., Quinn, C.D., Li, H., Washington, C. a, Cromwell, J., Giamanco, C.M., Forman, M., Holden, J., Rothman, R.E., Parker, M.L., Ortenberg, E. V, Zhang, L., Lin, Y.-L., Gaydos, C. a, 2013. Clinical accuracy of a PLEX-ID flu device for simultaneous detection and identification of influenza viruses A and B. J. Clin. Microbiol. 51, 40–5.
Zhou, B., Donnelly, M.E., Scholes, D.T., St George, K., Hatta, M., Kawaoka, Y., Wentworth, D.E., 2009. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza a viruses. J. Virol. 83, 10309–13.
Zhu, X., McBride, R., Nycholat, C.M., Yu, W., Paulson, J.C., Wilson, I. a., 2012. Influenza virus neuraminidases with reduced enzymatic activity that avidly bind sialic acid receptors. J. Virol. 86, 13371–13383.

Figures & Tables with Legends

Table 1. JHMI Clinical and demographic parameters by viral genotype.

Clinical/Demographics parameter		Total JHH population N =94	Virus		p-value
			NAg-F2H N= 34	**NAg+F2P** N =20	
Age					
	<5	16% (15/94)	12% (4/34)	15% (3/20)	0.6436
	5-17	21% (20/94)	32% (11/34)	15% (3/20)	
	18-49	16% (15/94)	18% (6/34)	15% (3/20)	
	50-64	10% (9/94)	9% (3/34)	10% (2/20)	
	65+	37% (35/94)	29% (10/34)	45% (9/20)	
Race					
	White	46% (43/94)	38% (13/34)	55% (11/20)	0.5142
	Black	29% (27/94)	29% (10/34)	30% (6/20)	
	Asian	11% (10/94)	9% (3/34)	5% (1/20)	
	Other	15% (14/94)	15% (5/34)	5% (1/20)	
Gender					
	Male	54% (51/94)	53% (18/34)	40% (8/20)	0.4083
	Female	46% (43/94)	47% (16/34)	60% (12/20)	
Pregnancy					
	Pregnant	5% (2/43)	29%(4/14)	20% (2/10)	0.6077
Vaccination status					
	Vaccinated	73% (24/33)	40% (4/10)	86% (6/7)	0.134
	Unknown	65% (61/94)	71% (24/34)	65% (13/20)	0.765
Disease severity					
	Pneumonia	15% (14/92)	9% (3/32)	20% (4/20)	0.4076
	Supplemental oxygen	29% (27/93)	32% (9/28)	44% (7/16)	0.5227
	Intubation	3% (3/93)	0% (0/22)	0% (0/4)	1

ICU	10% (9/94)	9% (3/34)	10% (2/20)	1
Death	5% (5/94)	3% (1/34)	10% (2/20)	0.5477
Hospital				
Inpatient	39% (37/94)	32% (11/34)	50% (10/20)	0.2527
Length of stay	5 (1-83)	7 (1-14)	5 (1-13)	0.4406
Symptoms				
Fever	81% (72/89)	91% (31/34)	71% (12/17)	0.09865
Cough	84% (74/88)	78% (25/32)	100% (19/19)	0.03701*
Headache	37% (19/52)	43% (9/21)	40% (4/10)	1
Sore throat	40% (17/42)	44% (8/18)	38% (3/8)	1
Runny nose	80% (45/56)	86% (18/21)	73% (11/15)	0.4178
Body aches	85% (23/27)	91% (10/11)	83% (5/6)	1
Shortness of breath	52% (31/60)	38% (8/21)	73% (11/15)	0.04893*
Wheezing	45% (22/49)	35% (7/20)	67% (8/12)	0.1437
Chills	35% (16/46)	40% (8/20)	33% (3/9)	
Fatigue	78% (42/54)	60% (12/14)	33% (3/9)	0.6818
Nausea/vomiting	29% (19/65)	9% (2/22)	21% (3/14)	0.3566
Co-morbidities				
None	41% (39/94)	50% (17/34)	25% (5/20)	0.136
One	24% (23/94)	24% (8/34)	50% (10/20)	
Multiple	34% (32/94)	26% (9/34)	25% (5/20)	
Specific co-morbidities				
Chronic lung disease	29% (27/93)	24% (8/34)	37% (7/19)	0.351
Cardiovascular disease	18% (17/93)	12% (4/34)	26% (5/19)	0.255
Renal disorder	14% (13/92)	12% (4/34)	11% (2/18)	1
Hepatic disorder	0% (0/93)	0% (0/34)	0% (0/19)	1

	Endocrine/metabolic disorder	25% (23/93)	15% (5/34)	21% (4/19)	0.7059
	Neurological disorder	20% (19/93)	21% (7/34)	11% (2/19)	0.4626
	Hematological disorder	3% (3/93)	3% (1/34)	0% (0/19)	1
	HIV/AIDS	3% (3/93)	3% (1/34)	0% (0/19)	1
	Obesity (BMI >30)	24% (11/46)	13% (2/15)	13% (2/16)	1

Influenza	WGS JHMI	WGS WPAFB
Type A	86% (66/77)	88% (14/16)
Subtype H3N2	84% (65/77)	88% (14/16)
Clade determined	77% (59/77)	88% (14/16)
3C.2	0% (0/59)	0% (0/14)
3C.2a	93% (55/59)	57% (8/14)
3C.2b	0% (0/59)	0% (0/14)
3C.3	0% (0/59)	7% (1/14)
3C.3a	5% (3/59)	14% (2/14)
3C.3b	2% (1/59)	21% (3/14)

Table 2. Influenza virus classification by WGS. A total of 77 JHMI and 16 WPAFB nasal swab samples were sequenced and influenza virus type, subtype, and clade were determined.

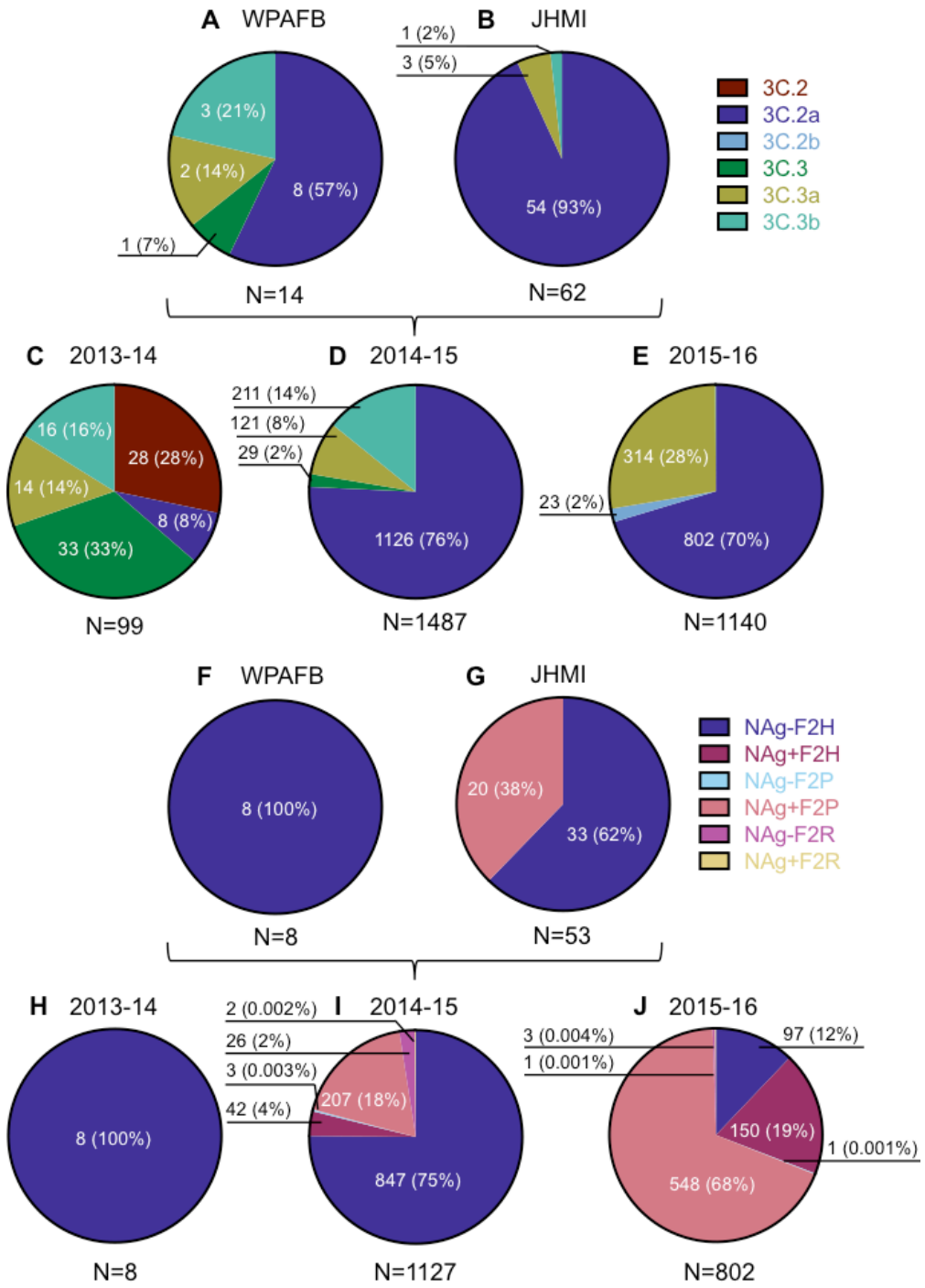

Figure 1. Characterization of human isolates by clade and genotype. The number of virus isolates that belong to clades and genotypes are reported according to sequence only. The prevalence of each clade is represented in pie charts A-E. The prevalence of each 3C.2a genotype, as a percentage of clade 3C.2a isolates, is represented in charts F-J. Pie charts A and F represent isolates from WPAFB, charts B and G represent isolates from JHMI. EpiFlu isolates from the 2013-14 season are represented in C and H and those from the 2015-16 season are represented in E and J. The pie charts for 2014-15 include both EpiFlu sequences and the sequences from this study.

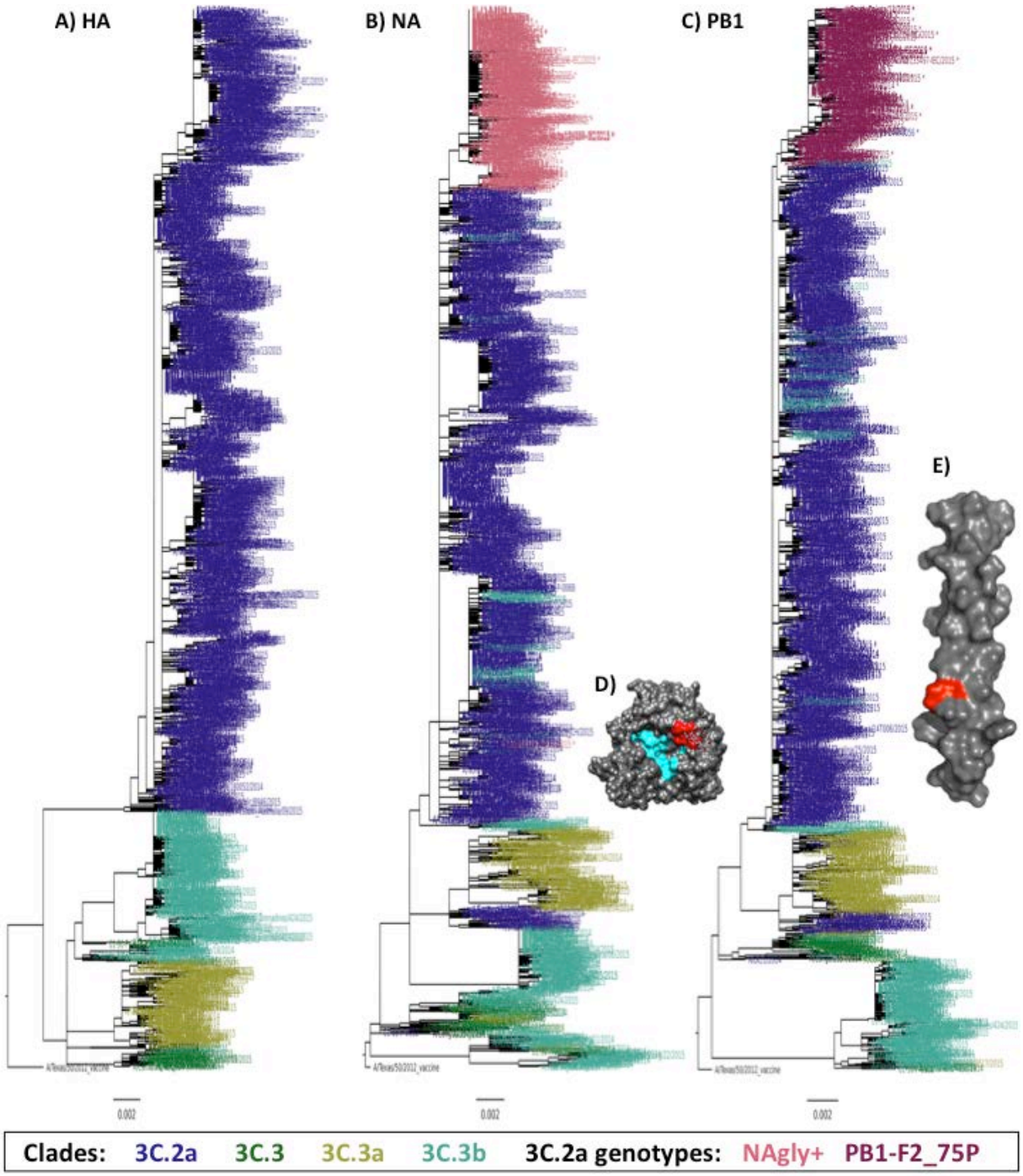

Figure 2. HA, NA, and PB1 phylogenetic trees and crystal structures. (A, B, C) Each tree was made with sequences from the gene segment indicated. The sequences included 1116 from the EpiFlu database that had full length nucleotide coverage for all open reading frames isolated from 10/01/2014 to 5/31/2015 plus 54 JHMI and 8 WPAFB isolates for a total of 1178 genomes. Phylogenetic trees were made using FastTree and images were made with FigTree. Branches are colored by clade (HA) or clade and genotype (NA and PB1) as indicated in the key. (D) X-ray

diffraction crystal structure of NA from A/Tanzania/205/2010 H3N2 (PDB 4GZP). The enzymatic site is colored cyan and the amino acids that encode the putative glycosylation site of NAg+ viruses are colored red. (E) Solution NMR structure of PB1-F2 C-terminal domain AA residues 50-87 from A/Puerto Rico/8/1934 (PDB 2HN8). Residue 75 is colored red. (D,E) Visualized using UCSF Chimera v1.11.1.

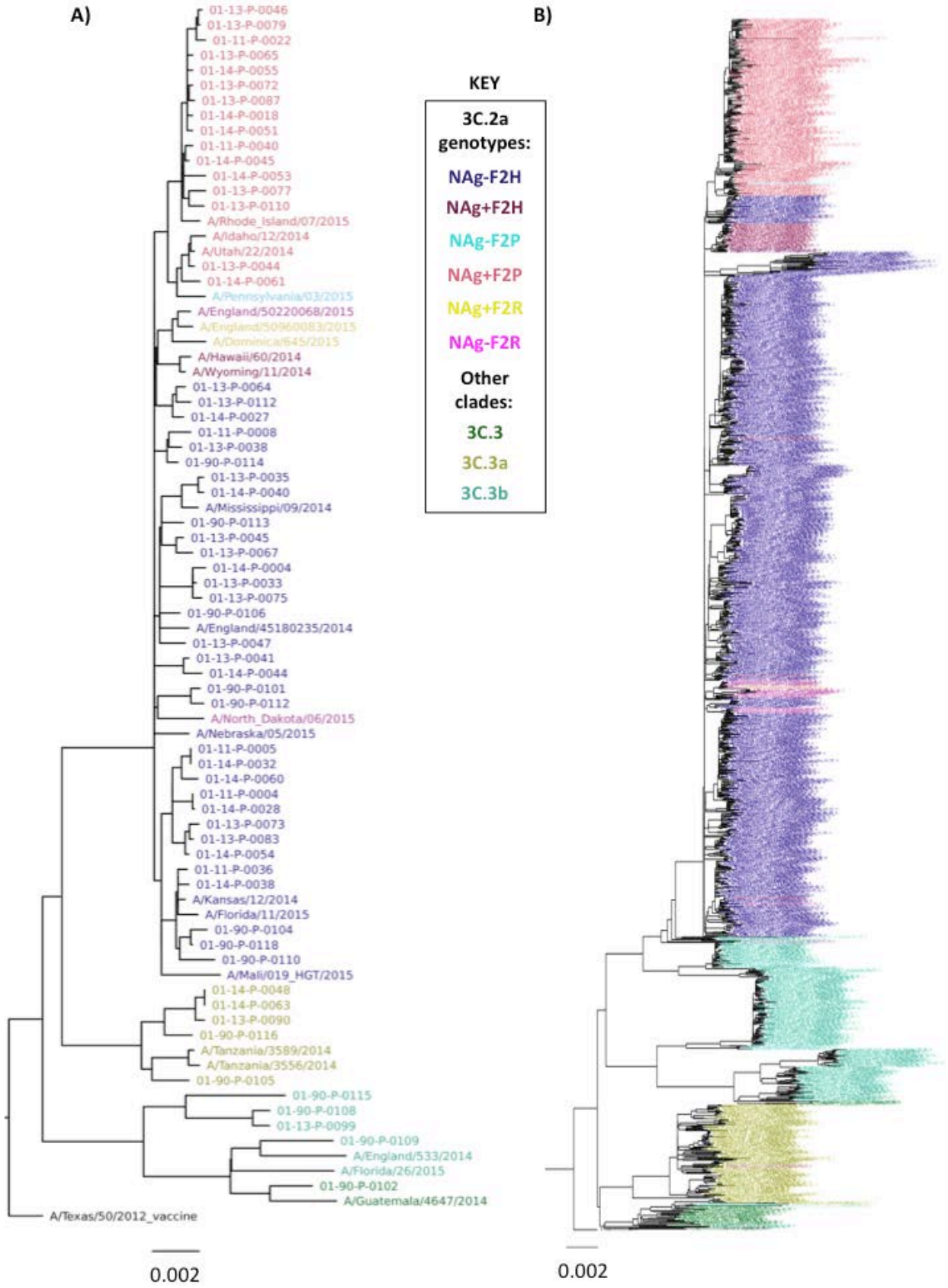
A)
B)
KEY
3C.2a genotypes:
NAg-F2H
NAg+F2H
NAg-F2P
NAg+F2P
NAg+F2R
NAg-F2R
Other clades:
3C.3
3C.3a
3C.3b
01-13-P-0046
01-13-P-0079
01-11-P-0022
01-13-P-0065
01-14-P-0055
01-13-P-0072
01-13-P-0087
01-14-P-0018
01-14-P-0051
01-11-P-0040
01-14-P-0045
01-14-P-0053
01-13-P-0077
01-13-P-0110
A/Rhode_Island/07/2015
A/Idaho/12/2014
A/Utah/22/2014
01-13-P-0044
01-14-P-0061
A/Pennsylvania/03/2015
A/England/50220068/2015
A/England/50960083/2015
A/Dominica/645/2015
A/Hawaii/60/2014
A/Wyoming/11/2014
01-13-P-0064
01-13-P-0112
01-14-P-0027
01-11-P-0008
01-13-P-0038
01-90-P-0114
01-13-P-0035
01-14-P-0040
A/Mississippi/09/2014
01-90-P-0113
01-13-P-0045
01-13-P-0067
01-14-P-0004
01-13-P-0033
01-13-P-0075
01-90-P-0106
A/England/45180235/2014
01-13-P-0047
01-13-P-0041
01-14-P-0044
01-90-P-0101
01-90-P-0112
A/North_Dakota/06/2015
A/Nebraska/05/2015
01-11-P-0005
01-14-P-0032
01-14-P-0060
01-11-P-0004
01-14-P-0028
01-13-P-0073
01-13-P-0083
01-14-P-0054
01-11-P-0036
01-14-P-0038
A/Kansas/12/2014
A/Florida/11/2015
01-90-P-0104
01-90-P-0118
01-90-P-0110
A/Mali/019_HGT/2015
01-14-P-0048
01-14-P-0063
01-13-P-0090
01-90-P-0116
A/Tanzania/3589/2014
A/Tanzania/3556/2014
01-90-P-0105
01-90-P-0115
01-90-P-0108
01-13-P-0099
01-90-P-0109
A/England/533/2014
A/Florida/26/2015
01-90-P-0102
A/Guatemala/4647/2014
A/Texas/50/2012_vaccine
0.002
0.002

Figure 3. Phylogenetic trees of concatemerized influenza virus genomes. (A) Sequences from the 54 JHMI, and 8 WPAFB isolates and some reference strains were concatenated and used to construct a phylogenetic tree. (B) A second phylogenetic tree of concatenated genes was generated using the 1178 2014-15 isolates used in Figure 2 plus all 2015-16 isolates from the EpiFlu database. Trees were made with the same methods for Figure 2. Branches are colored by clade or genotype as indicated in the key.

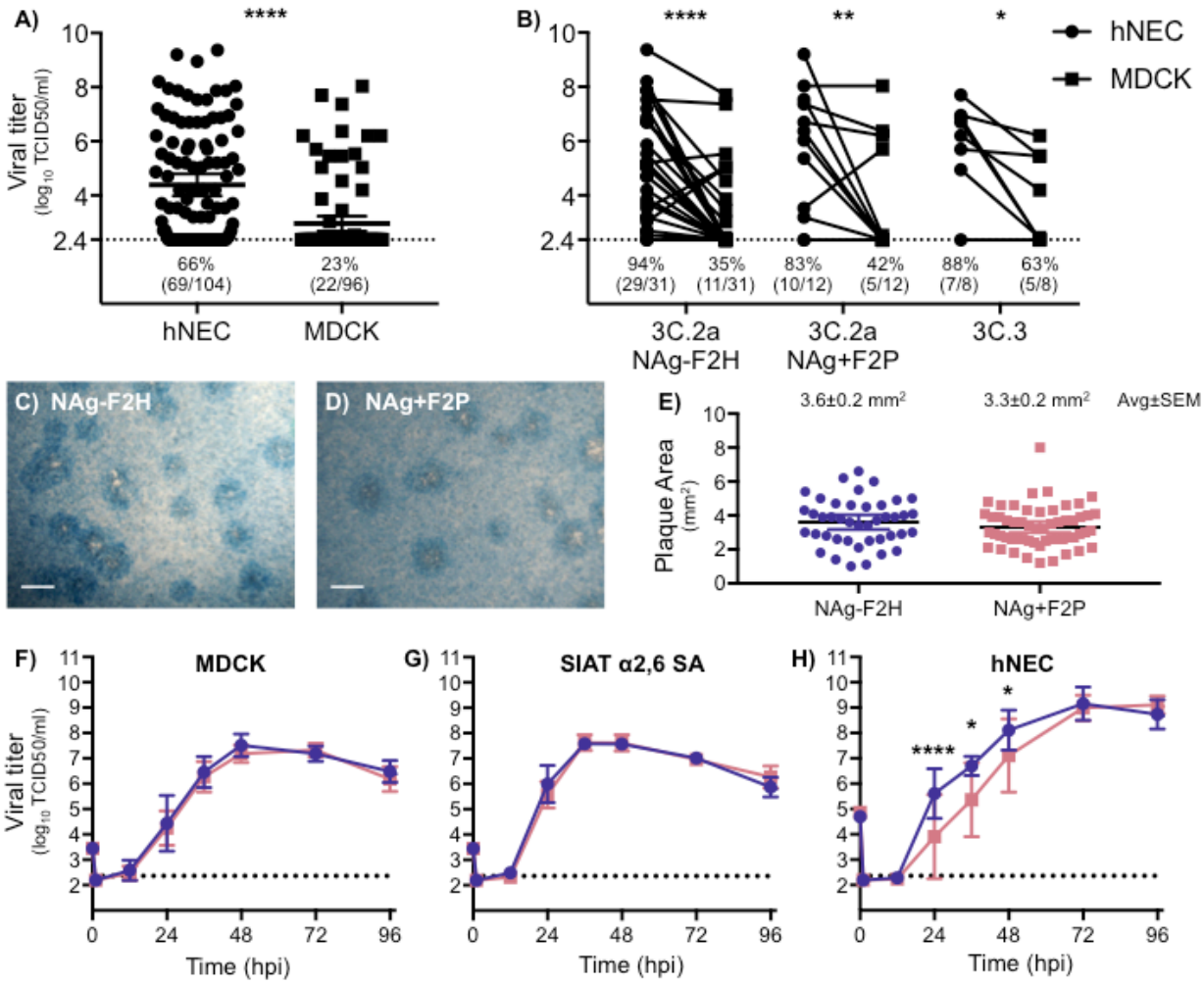

Figure 4. H3N2 virus isolation and characterization. (A) Virus isolation was attempted with 104 NPS specimens on hNEC cultures and 96 NPS specimen on MDCK cell cultures. Apical washes and supernatants were collected from hNEC and MDCK cell cultures, respectively, at 4 dpi and infectious virus titers determined by $TCID_{50}$. Lines represent means and error bars represent 95% CIs. Statistical significance was determined with a Wilcoxon matched pairs signed rank test. (B) Comparison of hNEC and MDCK 4dpi viral isolation titers was performed for three major genotypes. The percent and ratio of viruses that grew to levels above the limit of detection are noted below each dataset (A,B). Repeated Measures two-way ANOVA indicated that virus had no significant effect on growth and growth in MDCKs was significantly lower than that achieved in hNECs for all three virus genotypes. The limit of detection was 2.36 log_{10} $TCID_{50}$/ml. (C-E) Plaque assays were performed by inoculating MDCK cells and incubating for 3 days, under agarose, at 37°C. Representative images of plaques are shown for each virus (C,D). The area of over 40 plaques for each virus was determined (E) and analyzed using ImageJ. The average of plaque areas between the two viruses was determined by a two-tailed unpaired t-test. (F-H) Viruses isolated on hNECs and expanded once in MDCK cells were used to compare viral growth

in MDCK cell (F), MDCK-SIAT cell (G), and hNEC (H) cultures. Cultures were inoculated with MOI 0.001, 0.001, and 0.01, respectively. Supernatants or apical washes were collected at times indicated and infectious virus determined by $TCID_{50}$. N=8 wells per virus, per cell type, per time point. Error bars represent SD. Statistical significance was determined with a Repeated Measures two-way ANOVA with Bonferroni posttest correction.. *P<0.05, **P<0.01, ***P< 0.001, ****P<0.0001