Mutation in Hemagglutinin Antigenic Sites in Influenza A pH1N1 Viruses from 2015–2019 in the United States Mountain West, Europe, and the Northern Hemisphere

Craig H. Decker †, Naomi Rapier-Sharman † and Brett E. Pickett *

Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT 84602, USA; craigdecker1@gmail.com (C.H.D.); naomi.rapier.sharman@gmail.com (N.R.-S.)
* Correspondence: brett_pickett@byu.edu
† These authors contributed equally to this work.

Abstract: H1N1 influenza A virus is a respiratory pathogen that undergoes antigenic shift and antigenic drift to improve viral fitness. Tracking the evolutionary trends of H1N1 aids with the current detection and the future response to new viral strains as they emerge. Here, we characterize antigenic drift events observed in the hemagglutinin (HA) sequence of the pandemic H1N1 lineage from 2015–2019. We observed the substitutions S200P, K147N, and P154S, together with other mutations in structural, functional, and/or epitope regions in 2015–2019 HA protein sequences from the Mountain West region of the United States, the larger United States, Europe, and other Northern Hemisphere countries. We reconstructed multiple phylogenetic trees to track the relationships and spread of these mutations and tested for evidence of selection pressure on HA. We found that the prevalence of amino acid substitutions at positions 147, 154, 159, 200, and 233 significantly changed throughout the studied geographical regions between 2015 and 2019. We also found evidence of coevolution among a subset of these amino acid substitutions. The results from this study could be relevant for future epidemiological tracking and vaccine prediction efforts. Similar analyses in the future could identify additional sequence changes that could affect the pathogenicity and/or infectivity of this virus in its human host.

Keywords: influenza virus; H1N1; hemagglutinin; HA; comparative genomics; virology; bioinformatics; phylogenetic tree; selection pressure

1. Introduction

The H1N1 subtype of Influenza A virus that was responsible for the 1918–1919 influenza pandemic killed between 50–100 million individuals [1]. More recently, the 2009 triple-reassortant swine H1N1 influenza A (pH1N1) virus was also associated with a pandemic with a much lower mortality rate [2]. Specifically, this recent pandemic was responsible for approximately 60 million infections and an estimated 12,469 deaths [3]. Although nearly a century elapsed between these two H1N1 pandemics, ongoing research efforts enabled a rapid response when the pH1N1 virus emerged. These past pandemics highlight the continued importance of sequence-based surveillance efforts to facilitate mutation tracking and evolutionary predictions that enable the development of more robust and effective vaccines against future influenza A isolates [4,5].

Influenza A virus (IAV) possesses several features that enable it to be a perennial threat to human health. The high error rate of the IAV RNA-dependent RNA polymerase [6], the reassortment of the eight genomic segments [7,8], the perpetual ability to infect humans and other hosts [9,10], and evasion of the host immune response [11]. IAV consists of eight genomic segments that traditionally code for at least ten proteins, with the potential for producing at least seven additional proteins [12]. The Hemagglutinin (HA) segment is 1698 bases in length, codes for a protein that is 565 amino acids long and is extremely

immunogenic. The HA protein is known to interact with the sialic acid on the surface of human cells, facilitating viral entry into host cells [13]. Consequently, the properties and overall structure of the HA sialic acid-binding region is highly conserved among IAVs to facilitate entry into the host cell [14,15]. In contrast, epitope regions in the HA protein are constantly under positive selection pressure from the host adaptive immune system [16], which is detected by observing higher-than-expected mutation rates at antigenic sites [17].

It is imperative to continue performing sequencing and surveillance in all regions of the world in an effort to reduce the risk that one or more advantageous mutations spread and become predominant, as documented with the pH1N1 [18–20]. Since the emergence of the 2009 pH1N1 triple-reassortant virus, a range of mutation and case-rate tracking studies have been conducted in many countries worldwide, including Australia [21,22], China [23], India [24,25], New Zealand [21], Singapore [21], the United Kingdom [26], the United States [27], and Zambia [28].

Through the combination of surveillance efforts with high-throughput sequencing technologies, the influenza research community has meticulously tracked viral protein changes across subtypes and lineages. Early efforts identified at least three epitope regions in the HA protein [29,30]. In 2010, Maurer-Stroh et al. noted an E391K mutation in the pH1N1 HA sequence that destabilized the protein by altering salt-bridge interactions [31]. Sakabe et al. demonstrated that the D127E, K142N, and D222G HA mutations in pH1N1 were necessary to adapt to a mouse host [32]. Ginting et al. showed that the combination of the neuraminidase H274Y mutation and the T82K, K141E, and R189K HA mutations resulted in oseltamivir resistance and increased virulence, thus augmenting the danger of the emergent 2009 H1N1 pandemic strain [33]. An analysis of H1N1 mutations concluded that HA substitutions improved protein stability that was lost after the prior emergence of seven mutations at functional regions [18].
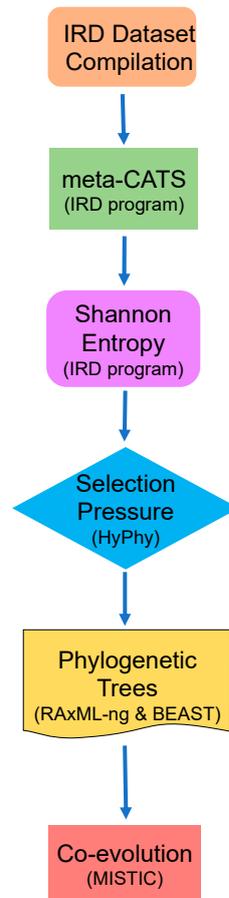
The aim of this comparative genomics study was to identify sequence substitutions in the HA protein that emerged between 2015–2019. Specifically, those that play a key role in the evolutionary trajectory of IAV across diverse geographic regions. While prior studies have evaluated and compared H1N1 mutational trends during this time period [34], they have not compared viruses collected in the United States Mountain West to other regions and have included both covariance and selection-pressure analyses.

2. Materials and Methods
2.1. Sequence Datasets

A bioinformatics workflow was designed to facilitate the analysis of HA sequence data from 2015–2019 (Figure 1). HA sequence data for the selected geographical areas were retrieved from the Influenza Research Database (www.fludb.org, accessed on 1 March 2020; IRD) [35]. The search criteria consisted of HA coding sequences from H1N1 viruses collected between 2015 and 2019 from avian, human, and swine hosts. Additional search parameters were applied to construct four sets of sequences based on the geographical location of sample collection (Figure 2), including: (A) Mountain West (MW) dataset consisting of HA sequences from Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming; (B) USA dataset with HA sequences from Alabama, Arizona, Colorado, Georgia, Idaho, Louisiana, Mississippi, Montana, Nevada, New Mexico, Texas, Utah, Vermont, Washington state, and Wyoming; (C) European dataset with HA sequences from Belgium, Czech Republic, Denmark, France, Germany, Italy, Russia, Spain, and the United Kingdom; as well as (D) a Northern Hemisphere HA sequence set consisting of the three collections described above as well as representative sequences from Japan and Canada. Sequences that contained unreasonably large numbers of insertions/deletions or other errors were manually excluded from our datasets and subsequent analysis. Maps highlighting the regions included in each dataset were made using MapChart [36].

2015–2019 H1N1 Basic Analytical Diagram

IRD Dataset
Compilation

meta-CATS
(IRD program)

Shannon
Entropy
(IRD program)

Selection
Pressure
(HyPhy)

Phylogenetic
Trees
(RAxML-ng & BEAST)

Co-evolution
(MISTIC)

Figure 1. A visual representation of the analytical workflow implemented in the current study.

2.2. Sequence Alignment and Variant Identification

Each of these four nucleotide-sequence sets was first aligned with MAFFT version 7.471 [37] using default parameters, followed by visualization and manual inspection with JalView version 2.11.1.0 [38]. Computational translation of the nucleotide sequences was performed to generate amino acid sequences that were subsequently realigned with MAFFT. The amino acid alignments were then used as input to the metadata-driven comparative analysis tool for sequences (meta-CATS) [39]. Briefly, this algorithm performs a chi-square statistical analysis on aligned sequences to identify positions that display a statistically significant skew in the distribution of residues and the associated metadata (e.g., geographical or temporal point of isolation). Amino acid positions were mapped to the HA protein using the A/California/04/2009 strain as the reference sequence. Positions identified by meta-CATS as significant (p-value < 0.05) were included in subsequent analyses and indicated a potential association between one or more sequence variations and a given metadata attribute, such as time of virus isolation.

The Shannon entropy for each alignment was calculated using the integrated IRD tool to quantify the prevalence of amino acid substitutions across each flu season in each dataset [40]. This provided a quantitative value for overall sequence diversity. The sequence feature variant type component of the IRD was queried to find annotated functions associated with each position containing an amino acid substitution [41]. The meta-CATS and entropy data were used to track the progress of the variants in structural, functional, and/or epitope regions through geographic areas.

Figure 2. Geographical regions represented by each dataset. (**A**) Mountain West dataset; (**B**) USA dataset; (**C**) European dataset; (**D**) Northern Hemisphere dataset.

2.3. Phylogenetic Tree Reconstruction

The Mountain West dataset was first evaluated with the 3seq algorithm [42]. This analysis determines whether recombinant sequences were present since such sites can bias downstream phylogenetic reconstructions and selection-pressure analyses [43,44]. Default settings were used to analyze 520 nonidentical HA CDS sequences, where both inferred segments had a minimum length of at least 100 nucleotides. Randomized Axelerated Maximum Likelihood-Next Generation (RAxML-NG) version 0.9.0 was run in a high-performance computing environment to generate bootstrapped maximum-likelihood phylogenetic trees for each aligned set of nucleotide sequences using the GTR-G model [45]. The Robinson–Foulds (RF) distances were used to measure the similarity of symmetric Maximum-Likelihood (ML) trees generated by RAxML-NG from the multiple sequence alignments. Relative RF distances were calculated [46] with smaller absolute and relative

distances representing more similarity between the generated tree and the best-scoring ML tree (Supplementary Table S1). Bootstrapped trees were then generated for each geographically distinct set of sequences and the RF distance was calculated.

2.4. Selection-Pressure Analysis

Prealigned sequences used for selection pressure were directly retrieved from the IRD website. Each set of aligned sequences was used to create a maximum-likelihood phylogenetic tree using PhyML software [47] on the IRD website with the HKY evolutionary model and all other default parameters. These IRD trees were used as seeds for the HyPhy analysis. HyPhy version 2.5.1 (MP) for Linux was run in a high-performance computing environment to detect positive or negative selection pressure on the aligned HA nucleotide sequences [48]. This analysis was performed using the mixed-effects model of evolution (MEME) [49], fixed-effects likelihood (FEL) [50], and single-likelihood ancestor counting (SLAC) algorithms. [50,51]

For each aligned sequence set, the selection-pressure analysis used the following procedure: phylogenetic tree files (in Newick format) and HA nucleic acid CDS aligned sequence files (in fasta format) were downloaded from IRD and the stop codons were manually trimmed from each alignment. Sequences that decreased the accuracy of the alignment were identified and manually excluded from this analysis. The trimmed sets of sequences were then run through HyPhy. The MEME algorithm was used to identify specific sites that were undergoing selection pressure [49]. The FEL method was used to test each site in the alignment for an overall evolutionary rate [50]. The SLAC algorithm generated a theoretical common ancestor coding sequence and then used it to calculate the synonymous and nonsynonymous substitution rates [50]. The output of all selection pressure algorithms was manually combined and compared to minimize false positives due to algorithm bias(es).

2.5. Bayesian Evolutionary Analysis

Bayesian Evolutionary Analysis by Sampling Trees (BEAST) v2.6.3 was used for Bayesian phylogenetic tree reconstruction. This Markov chain Monte Carlo (MCMC) re-sampling algorithm uses a Bayesian method to generate a posterior probability distribution that estimates the correctness of the tree based on the provided data. Aligned nucleotide files from each dataset (MW, USA, Euro, Northern Hemisphere) were converted from fasta to Nexus format to meet input requirements. BEAUti v.2.6.4 was used to generate the XML-based input file and a relaxed clock model was employed [52]. BEAST was run in a high-performance computing environment [53].

Effective sample size (ESS) is a measure of the quality of the BEAST analysis. In order to reach an ESS of more than 200—the commonly accepted BEAST quality indicator—the datasets were run with a chain-sampling frequency of 1000 for varying chain lengths [54]. The wide variation present in these large datasets required chain lengths of 100 to 450 million generations (Supplementary Table S2) to achieve acceptable ESS values, with values from 100–200 considered acceptable in some scenarios and values greater than 200 considered ideal. In cases where multiple tree files were generated for the same sequence set, they were combined with BEAST's postprocessing tool Logcombiner v2.6.3 and viewed with Tracer v1.7.2 [55]. Two of the resulting trees (MW and Euro) were strict consensus trees with 10% of states removed, while the other two (USA and Global) were majority-rule consensus trees with 50–60% of states removed due to the large number of sequences in the datasets [56]. An ESS of $\gtrsim$200 for three of the datasets was calculated by Tracer with the NH dataset having an ESS of 112 (Supplementary Table S2). Finally, the trees were combined to create a consensus tree using TreeAnnotator v2.6.3 and viewed with FigTree v1.4.4.

2.6. Coevolutionary Sequence Analysis

The Mutual Information Server to Infer Coevolution (MISTIC) was used to calculate coevolution, or covariance, between pairs of aligned amino acid positions in the HA

protein [57]. This method uses a mutual information (MI) algorithm to identify residues that coevolve or co-vary with others in the aligned sequences. The output was summarized in tabular form and as a circos plot displaying the MI coevolution network.

3. Results

3.1. Basic Analytical Design

For each dataset, we followed the same bioinformatics analytical workflow (Figure 1). First, we gathered all available H1N1 sequence samples that met our search criteria in the IRD from the geographical region being studied (MW, USA, Euro, or NH; Figure 2, Table 1). We then reconstructed phylogenetic trees to determine the evolutionary relationships between H1N1 HAs over time. To better understand the sequence variations that play an important role in separating the phylogenetic clades, we next performed a meta-CATS analysis on these aligned sequences to identify amino acid substitutions that significantly differed between 2015 and 2019. This method incorporates a chi-square statistical approach, with the sequences from each region divided into two groups (pre-summer 2017 and post-summer 2017). This approach enabled us to identify which amino acid positions displayed significant "skew" in the distribution of residues before and after the summer of 2017. We specifically chose the summer in the Northern Hemisphere as the division, since it corresponds to a natural break between flu seasons. The meta-CATS method reports amino acid substitutions that significantly associate with the year of collection, and not when the amino acid substitutions occurred. We consequently calculated Shannon entropy values, as implemented in the IRD, for each alignment to determine when the amino acid substitutions occurred within our 2015–2019 temporal window. This method calculates the frequencies of amino acids present at each position across an alignment of flu sequences. For example, HA position 200 had 57% proline and 43% serine during the period between November 2016 to March 2017. Using the Shannon entropy values, we created tables to facilitate tracking of substitutions occurring at each position, when they occurred temporally, and their frequencies.

Table 1. Total number of samples in each dataset.

Mountain West	USA	Europe	Northern Hemisphere
933	1924	309	2389

3.2. 2015–2019. Mountain West H1N1 HA Sequences Fall into Two Distinct Clades with Specific Substitutions

We began our analysis of the Mountain West (MW) dataset by using the 3seq method to confirm that no recombinant sequences were present in this dataset. We then reconstructing maximum likelihood (Supplementary Figure S1) and Bayesian (Supplementary Figure S2) phylogenetic trees for this dataset, with the Bayesian tree having an ESS value of 291.6. These nucleotide tree reconstructions confirmed two primary clades, the first consisting of strains isolated in 2015–2016 and the second consisting of strains collected in 2018–2019. Given this phylogenetic signal, we next used meta-CATS to identify sequence positions that most contributed to the phylogenetic topology. This analysis identified 56 statistically significant amino acid positions that differentiated strains collected between the 2015–2016 Northern Hemisphere flu seasons (i.e., pre-summer 2017), and the 2017–2018 Northern Hemisphere flu seasons (i.e., post-summer 2017) (Supplementary Table S3). Sorting these positions by their *p*-values and subsequent manual review led us to prioritize three amino acid substitutions that are all located in functional and epitope regions of the HA protein: K147N (HA1 130), P154S (HA1 137), and S200P (HA1 183). We also identified positions 159 (HA1 142) and 233 (HA1 216), which are both in functional regions, as having strong *p*-values with some potential statistical skew introduced by low numbers of certain amino acid residues at these positions. We also identified eight additional positions that are located primarily in known immunogenic regions, including 62 (HA1 45), 177 (HA1 160),

179 (HA1 162), 190 (HA1 173), 252 (HA1 235), 277 (HA1 260), 299 (HA1 282), and 313 (HA1 297). We then used the HyPhy software to determine whether these codons were subjected to selection pressure. We observed positive selection pressure at position 200 as well as the eight positions in immune epitope regions. In contrast, we observed no pressures at positions 147, 154, 159, or 233 (Supplementary Table S4).

3.3. USA, Euro, Northern Hemisphere Sampling Results Overview

After completing this process, we wanted to confirm our findings while minimizing the contribution of sampling bias from sequences collected in the US Mountain West region. We consequently decided to evaluate progressively larger datasets from more broad geographical regions (e.g., United States, Europe, and the Northern Hemisphere). Due to the large number of 2015–2019 H1N1 samples available in the United States, it was intractable to analyze all samples. To ameliorate this difficulty, in addition to the sequences from the MW dataset, we included representative samples from a variety of states across the United States, Europe, and the Northern Hemisphere.

Specifically, we wanted to determine whether the K147N, P154S, and/or S200P substitutions, and to a lesser extent the substitutions at positions 159 and 233, could be detected before and after the summer of 2017 in each of the larger sets of sequences. We decided to focus on the substitutions at positions 147, 154, and 200 since they are located within functional regions that were not solely characterized as immune epitopes. We consequently excluded the eight positions in known epitopes, as well as positions 159 and 233, since we observed a low number of counts for a subset of amino acid residues at these latter two positions, which adversely skewed statistics and did not justify a more in-depth temporal analysis.

We therefore created new datasets (Figure 2, Table 1) by sampling US states outside of the Mountain West region, nine countries from Europe, and additional Northern Hemisphere countries. We then reran meta-CATS on these datasets to identify any significant substitutions that were detected at these positions. We observed that many of the same 56 sites that contained significant changes in the MW set maintained their statistical significance in the three other datasets (Supplementary Table S3). Specifically, we observed 70, 48, and 161 sites that contained significant changes between pre-summer 2017 and post-summer 2017 in the United States, Europe, and NH datasets, respectively. We also observed no selective pressures at positions 147, 154, 159, or 233. Position 200 had detectable pressure by the FEL and SLAC algorithms in the MW dataset, and by the SLAC algorithm in the USA dataset (Supplementary Table S4). The maximum-likelihood and Bayesian phylogenetic trees for the USA (Supplementary Figures S3 and S4), Euro (Supplementary Figures S5 and S6), and NH (Supplementary Figures S7 and S8) datasets continued to show well-separated clades in the 2015–2016 and in the 2018–2019 temporal periods. We observed that the ESS values for the more divergent datasets were lower than expected, which is not surprising given the high mutation rate for influenza viruses. Specifically, the ESS values for the USA, European, and Northern Hemisphere trees were 197.8, 607.9, and 112, respectively.

3.4. Selection Pressure

Our selection-pressure analysis across the various sequence sets identified 27 HA residues that underwent either positive or negative selection pressure in at least one regional dataset (Table 2; Supplementary Table S4). We observed that position 200 underwent detectable selection pressure only in the United States and Mountain West datasets. We also detected statistically significant evidence selection occurring at the eight immune epitope positions, including positions 62, 177, 179, 190, 252, 277, 299, and 313. In contrast, we found that positions 147, 154, 159, and 233 were not detected by this analysis.

Table 2. Significant results from each HyPhy selection pressure algorithm by codon position, with the MW, US, EU, and NH datasets represented in blue, lavender, red, and yellow (respectively).

MW MEME	MW FEL	MW SLAC	US MEME	US FEL	US SLAC	EU MEME	EU FEL	EU SLAC	NH MEME	NH FEL	NH SLAC
3			3	3	3				3		
	6	6	6	6	6						
			11	11	11	11	11	11	11	11	11
36											
						57					
							60				
65			65								
			86								
90			90								
			137						137		
145											
	146										
	156	156		156	156						
											158
180	180										
200	200	200			200						
				203	203		203	203	203	203	203
						273					
			278	278	278						
289											
			472								
505											
						513			513		
						524	524				
545						545	545		545	545	545
						550	550			550	
						566					

3.5. Temporal Prevalence of Mutation across Geographical Areas

We next wanted to determine a more precise evolutionary timeline of the non-epitope positions, or those with skewed *p*-values, by calculating the frequency of amino acid residues that were at positions 147, 154, and 200 (Tables 3–5). To accomplish this, we generated Shannon Entropy values for each amino acid position across the multiple sequence sets. We observed that the S200P mutation was relatively uncommon in Europe before the 2017–2018 flu season, with a frequency of 28.13% between 2015 and 2017 (Table 5). In contrast, the frequency of this substitution increased during the 2018–2019 flu season, reaching a frequency of 93.51% in Europe towards the end of the 2019 flu season. These quantitative results provide additional insight into the increased prevalence of the S200P substitution across each dataset in our study to eventually become the dominant allele. Subsequent visualization of these substitutions in a three-dimensional HA structure confirms their respective location and relevant biological functions (Supplementary Figures S9 and S10).

Table 3. Evolutionary timeline of hemagglutinin K147N variant by flu season and dataset, with the MW, US, EU, and NH datasets represented in blue, lavender, red, and yellow (respectively).

	MW		US		EU		NH	
Year	%K	%N	%K	%N	%K	%N	%K	%N
Summer 2015	100% (5/5)	0% (0/5)	100% (11/11)	0% (0/11)	100% (25/25)	0% (0/25)	100% (37/37)	0% (0/37)
Winter 2015–2016	100% (297/297)	0% (0/297)	100% (568/569)	0% (0/569)	100% (64/64)	0% (0/64)	100% (665/666)	0% (1/666)
Summer 2016	100% (34/34)	0% (0/34)	100% (59/59)	0% (0/59)	100% (4/4)	0% (0/4)	100% (66/66)	0% (0/66)
Winter 2016–2017	100% (69/69)	0% (0/69)	100% (143/143)	0% (0/143)	90% (9/10)	10% (1/10)	99% (153/154)	1% (1/154)
Summer 2017	100% (12/12)	0% (0/12)	96% (25/26)	0% (0/26)	NA *	NA *	96% (25/26)	4% (1/26)
Winter 2017–2018	100% (122/122)	0% (0/122)	100% (316/317)	0% (0/317)	% (53/53)	0% (0/53)	100% (469/470)	0% (1/470)
Summer 2018	97% (28/29)	3% (1/29)	93% (42/45)	0% (0/45)	100% (1/1)	0% (0/1)	94% (47/50)	6% 3/50
Winter 2018–2019	91% (292/322)	8% (26/322)	90% (595/658)	9% (56/658)	95% (116/122)	5% (6/122)	91% (722/791)	8% 62/791
Summer 2019	44% (19/43)	56% (24/43)	55% (53/96)	44% (42/96)	100% (30/30)	0% (0/30)	67% (86/129)	33% 42/129

K: lysine; N: asparagine; winter season includes October through March; Summer season includes April through September. MW: Mountain West region: US: USA region; EU: Europe region; NH: Northern Hemisphere region. * NA values were recorded for absence of sequences during specific time periods.

Table 4. Evolutionary timeline of hemagglutinin P154S variant by flu season and dataset, with the MW, US, EU, and NH datasets represented in blue, lavender, red, and yellow (respectively).

	MW		US		EU		NH	
Year	%P	%S	%P	%S	%P	%S	%P	%S
Summer 2015	100% (5/5)	0% (0/5)	100% (11/11)	0% (0/11)	60% (15/25)	32% (8/25)	70% (26/37)	24% (9/37)
Winter 2015–2016	99% (293/297)	1% (4/297)	99% (563/569)	1% (6/569)	84% (54/64)	9% (6/64)	98% (650/666)	2% (12/666)
Summer 2016	91% (31/34)	9% (3/34)	95% (56/59)	5% (3/59)	50% (2/4)	25% (1/4)	92% (61/66)	6% (4/66)
Winter 2016–2017	100% (69/69)	0% (0/69)	99% (142/143)	1% (1/143)	40% (4/10)	60% (6/10)	95% (147/154)	5% (7/154)
Summer 2017	100% (12/12)	0% (0/12)	100% (26/26)	0% (0/26)	NA *	NA *	100% (26/26)	0% (0/26)
Winter 2017–2018	98% (120/122)	2% (2/122)	98% (311/317)	2% (6/317)	87% (46/53)	13% (7/53)	95% (447/470)	5% (22/470)
Summer 2018	97% (28/29)	3% (1/29)	93% (42/45)	7% (3/45)	0% (0/1)	100% (1/1)	92% (46/50)	8% (4/50)
Winter 2018–2019	75% (240/322)	25% (80/322)	76% (503/658)	23% (153/658)	94% (115/122)	6% (7/122)	80% (629/791)	20% (160/791)
Summer 2019	93% (40/43)	7% (3/43)	91% (87/96)	9% (9/96)	100% (30/30)	0% (0/30)	93% (120/129)	7% (9/129)

P: proline; S: serine; winter season includes October through March; Summer season includes April through September. MW: Mountain West region: US: USA region; EU: Europe region; NH: Northern Hemisphere region. * NA values were recorded for absence of sequences during specific time periods.

Table 5. Evolutionary timeline of hemagglutinin S200P variant by flu season and dataset, with the MW, US, EU, and NH datasets represented in blue, lavender, red, and yellow (respectively).

	MW		US		EU		NH	
Year	%S	%P	%S	%P	%S	%P	%S	%P
Summer 2015	100% (5/5)	0% (0/5)	100% (11/11)	0% (0/11)	64% (16/25)	36% (9/25)	73% (27/37)	27% (10/37)
Winter 2015–2016	99% (295/297)	1% (2/297)	100% (567/569)	0% (2/569)	84% (54/64)	16% (10/64)	98% (654/666)	2% (12/666)
Summer 2016	91% (31/34)	9% (3/34)	95% (56/59)	5% (3/59)	50% (2/4)	50% (2/4)	92% (61/66)	8% (5/66)
Winter 2016–2017	43% (30/69)	57% (39/69)	55% (79/143)	45% (64/143)	20% (2/10)	80% (8/10)	53% (82/154)	47% (72/154)
Summer 2017	83% (10/12)	17% (2/12)	81% (21/26)	19% (5/26)	NA *	NA *	81% (21/26)	19% (5/26)
Winter 2017–2018	25% (31/122)	75% (91/122)	19% (61/317)	81% (256/317)	75% (40/53)	25% (13/53)	33% (153/470)	67% (316/470)
Summer 2018	14% (4/29)	86% (25/29)	16% (7/45)	84% (38/45)	0% (0/1)	100% (1/1)	18% (9/50)	82% (41/50)
Winter 2018–2019	4% (13/322)	96% (309/322)	3% (22/658)	97% (636/658)	7% (9/122)	93% (113/122)	4% (31/791)	96% (760/791)
Summer 2019	5% (2/43)	95% (41/43)	3% (3/96)	97% (93/96)	3% (1/30)	97% (29/30)	3% (4/129)	97% (125/129)

P: proline; S: serine; winter season includes October through March; Summer season includes April through September. MW: Mountain West region: US: USA region; EU: Europe region; NH: Northern Hemisphere region. * NA values were recorded for absence of sequences during specific time periods.

3.6. Additional Amino Acid Positions with Significant Temporal Changes

One mutation of particular interest is the nonsynonymous S200P substitution. The meta-CATS analysis for the NH dataset calculated a *p*-value of 9.308×10^{-237} for the S200P amino acid substitution, which is located within the sialic acid-binding domain of HA.

The K147N substitution was detected in only 13.8% of the Mountain West sequences by the 2018–2019 influenza season, but still achieved a *p*-value of 8.323×10^{-13}. In contrast, over the summer of 2018, we calculated that 6% of the H1N1 sequences in the Northern Hemisphere dataset had a lysine-to-asparagine substitution at this position (Table 3), with a *p*-value of 3.404×10^{-21}.

The substitution at position 154 underwent a partial change in our datasets from proline to serine during 2015–2019 (Table 4), with a *p*-value of 3.404×10^{-21} in the NH dataset. We observed that the partial P154S substitution first gained traction in the European

dataset during the 2016–2017 (60%) and 2017–2018 (13%) flu seasons, then emerged in the Mountain West (25%) and the USA (23%) datasets in the 2018–2019 flu season.

Positions 159 and 233 had significant *p*-values in the Northern Hemisphere comparison (0.0016 and 6.28×10^{-40}, respectively). However, as mentioned above, these positions displayed biased *p*-values due to the skew in the observed amino acid residues at these two positions. Due to the expected change in epitope regions over time, we did not calculate the temporal change of the eight epitope regions that we identified above.

3.7. Coevolution

We next wanted to determine whether any compensatory mutations facilitated the three substitutions described above. We consequently performed a coevolution analysis using the public MISTIC server. We observed that residues 147, 154, and 200 had large mutual information (MI) values with multiple other residues (Table 6; Supplementary Table S5), which indicates that they may have coevolved with other positions as a cluster. We generated a circos plot to visualize the large number of coevolving residues (Supplementary Figure S11). Additionally, we observed several distinct large MI values for residues coevolving with positions 147 and 154. Since the largest MI values have the highest rates of coevolution, these changes can be assumed to be non-independent. Although we did not observe strong quantitative evidence of direct coevolution between most pairwise combinations of our five selected non-epitope residues, we did observe that the forty-first best-scoring coevolution result was between positions 147 and 233. We did observe strong coevolution values for seven of the eight epitope positions, including 62, 177, 190, 252, 277, 299, and 313.

Table 6. Mutual information coevolution for the 20 highest-scoring pairs of residues *.

1° Residue Position	1° Residue	2° Residue Position	2° Residue	Mutual Information Value **
181	S	312	I	1175.141846
91	S	181	S	1163.426392
91	S	312	I	1159.654785
200	**S**	**312**	**I**	**714.461975**
91	**S**	**200**	**S**	**710.796936**
181	**S**	**200**	**S**	**661.189087**
146	**N**	**202**	**T**	**606.855347**
146	**N**	**277**	**N**	**527.068542**
62	**R**	**315**	**I**	**462.919434**
299	**P**	**315**	**I**	**450.615295**
202	**T**	**277**	**N**	**435.101715**
62	**R**	**299**	**P**	**423.069916**
154	**P**	**468**	**N**	**382.397064**
147	**K**	**313**	**H**	**374.515015**
154	**P**	**190**	**V**	**370.027618**
252	**E**	**537**	**V**	**356.437103**
147	**K**	**177**	**K**	**323.149689**
177	**K**	**313**	**H**	**316.047058**
177	**K**	**233**	**T**	**316.031616**
421	I	523	E	307.183044

* Bold text emphasizes rows containing at least one position from the current work (e.g., 62, 147, 154, 177, 190, 200, 202, 233, 252, 277, 299, and 313). ** Larger MI values signify higher levels of coevolution.

3.8. Amino Acid Positions Having Substantial Coevolution, Selection Pressure, and/or Temporal Changes

We then examined the remaining 53 of 56 amino acid positions in HA that had FDR-corrected *p*-values < 0.05 (Supplemental Table S6) as well as functions annotated from the literature. We filtered this original list to include those that were most likely to drive a sustained change over time. Specifically, we focused on positions located in either functional or protein structural regions rather than epitope regions—primarily since nearly

all positions in the HA protein have already been characterized as immune epitopes. This analysis identified two positions, 159 and 233, which had significant meta-CATS *p*-values. Residue 233 displayed strong evidence of coevolution (Table 5), while position 159 showed some evidence of coevolution (Supplementary Table S5).

We also observed other positions in functional and/or structural regions that had been identified by at least two of the algorithms. Positions that significant Bonferroni-adjusted meta-CATS *p*-values as well as good coevolution values include 147, 154, 202, and 233. Those positions with significant adjusted meta-CATS *p*-values and significant HyPhy *p*-values include 3, 6, 155, and 203. Lastly, positions with significant adjusted meta-CATS *p*-values, acceptable coevolution values, and significant HyPhy results include 179, 190, and 200 (Supplemental Table S6).

Lastly, we detected amino acid positions in known human immune epitope regions that had significant adjusted meta-CATS *p*-values within our temporal period, significant HyPhy *p*-values for selection pressure, and measurable coevolution values in at least one comparison. Specifically, these positions include 62, 177, 179, 190, 252, 277, 299, and 313.

4. Discussion

The goal of this study was to identify sequence substitutions that contributed to the observed separation of pH1N1 HA protein sequences from the Mountain West region of the United States and other regions around the world into two distinct phylogenetic clades. We then confirmed that these two clades are present in multiple sets of sequences from diverse geographic regions across the Northern Hemisphere. Subsequent sequence analyses for selection pressure and coevolution identified amino acid positions 200, 154, and 147 as strongly contributing to this divergence, with these positions playing roles in protein function, structure, and/or immune epitopes. We identified positions 159 and 233 as having a measurable contribution, and other positions with a smaller contribution to sequence divergence in IAV in the Northern Hemisphere between 2015–2019.

To our knowledge, applying a computational selection-pressure analysis to pH1N1 datasets from various geographical regions during this temporal period is novel. The advantage of using the three chosen selection-pressure algorithms is their ability to detect different facets of a complex system; with each method having unique capabilities to identify relevant codons undergoing selection. Similarly, the application of a coevolution method further supported a more in-depth analysis of the residues at positions 147, 154, 200, and 233. The lack of statistical significance for the S200P substitution in the European and Northern Hemisphere datasets could be due to various factors, including the increased prevalence of codons that produce this substitution in the European and Northern Hemisphere populations.

Our statistical analysis of amino acid sequences identified HA position 200 as relevant to the clade separation, which has been characterized previously [34]. Specifically, residue 200 has been shown to be located within the highly conserved sialic acid-binding site [58–60]. Position 200 is also located within a B-cell epitope in mice [61], rabbits [61], humans [62], as well as within a T-cell epitope [63]. Whether or not position 200 tests positive or negative as a T-cell epitope likely depends on the exact sequence surrounding position 200 in the strain in question and which HLA allele is present in the host [64–67].

Typically, HA functional pockets undergo relatively low levels of mutation [68], making the emergent S200P substitution somewhat surprising. The substitution of proline for any other amino acid has been shown to potentially impact the folding, secondary structure, and/or stereochemistry of a protein due to its ring structure [69]. Interestingly, Al Khatib et al. (2018) discovered that S200P increases binding avidity between HA and sialic acid [34]. The potentially large effect of the S200P substitution on the structure of the highly conserved binding pocket likely justifies additional experimental investigation.

Previous studies have also noted the presence of the S200P mutation in pH1N1 HA proteins [70]. The S200P substitution first appeared in Uganda during late 2010/early 2011 [71] and in the United States in 2010 [34]. Based on this evidence of S200P emerging in

the early 2010s, it is possible that the surge of S200P that occurred during 2015–2019 was a second wave, accompanied by compensatory mutations elsewhere in the protein that may have increased its fitness [72]. Researchers observed a similar surge in China from 2017–2019 [73].

HA position 154 is located within the highly conserved sialic acid binding site, which is located near position 200 in three-dimensional space. This position partially underwent a change from proline to serine during the summer of 2017 [74]. This previous observation at least partially supports the possibility that P154S coevolved with the S200P mutation and may compensate to optimize function after the emergence of the former change.

Our observation that large frequencies of S200P substitutions are often accompanied by P154S substitutions is interesting since the latter is located in a T-cell epitope [75,76] a B-cell epitope [62] the sialic acid binding site [77]. The P154S mutation has been shown to allow efficient escape from antibodies targeting the Ca2 antigenic region around residue 154 [60]. The P154S HA mutation, which appeared to decrease viral replication of oseltamivir-resistant H1N1 strains in ferrets [78], emerged as early as 2010 in Germany [79].

The 10% K147N shift during the summer of 2018 represents a mutation that may justify continued investigation and surveillance. We found this substitution interesting because it occurs within the functional Lysine fence [80]. This sequence plays a role in stabilizing the virus–host membrane interaction during entry into a new host cell. To our knowledge, it is unknown whether this substitution affects the function of the lysine fence. Additionally, HA position 147 is located within known B-cell [62] and T-cell [63,76,81–85] epitopes. This antigenic shift has been associated with immune escape [60]. Previous passaging experiments of the A/California/07/09 isolate in the presence of a novel neuraminidase inhibitor gave rise to a subpopulation of viruses that coded for the K147N substitution, suggesting a potential selective advantage for HA proteins containing the K147N substitution [86].

Our identification of positions 159 (HA1 142) and 233 (HA1 216) is of particular interest. Specifically, the former is found in the conserved Lysine fence region [80,87], which has been shown to be unique to pH1N1 strains [87]. This infers that it would not be subjected to selection pressure. Its lack of substantial coevolution suggests that this position evolved independently of others and may be at least a partial driver of the antigenic drift that we observed before and after the summer of 2017. Position 233 is located in the receptor-binding site [88]. The lack of selection pressure and the detected coevolution at this site indicates that changes at this site may have been accompanied by other substitutions across that HA protein with an overall preference for sequence conservation. It is likely that the five sites identified by this study substantially contributed to the continued virulence of Influenza virus in 2017. Although positions 159 and 233 had highly significant *p*-values, the statistics were likely at least somewhat skewed due to the presence of low numbers of residues (<3) in at least one of the groups of sequences that were compared.

Positions 3, 6, 154, 155, 202, 203, 179, and 190 had attributes of being important to the divergence of IAV; however, we are unable to determine whether they are likely to be drivers of the antigenic drift or simply co-occurring passenger substitutions. Additional experiments in the wet lab are needed to better quantify whether they affect viral fitness and/or pathogenesis.

Amino acid positions 62, 177, 179, 190, 252, 277, 299, and 313 are located within known human B-cell and/or T-cell epitope regions. Interestingly, position 179 lies within the well-characterized S_a epitope region first identified by Caton et al. [29,89]. Although it is not surprising to identify regions undergoing selection and coevolution as epitope regions, such substitutions could hint at which epitope regions changed to facilitate continued infection of various hosts in our selected temporal period.

Phylogenetic tree reconstructions have been generated previously for IAV strains collected during our target temporal period [58,90]. Our observation that Influenza A viruses undergo recombination rarely has been reported previously [42,91]. The phylogenetic analyses that we performed showed a distinct separation of clades between strains

isolated in the Northern Hemisphere before and after 2017. The fact that our trees obtained good ESS values across multiple datasets supports that influenza A viruses evolve under a molecular clock [92–94]. We suspect that the large shift in the prevalence of strains that contained these three substitutions occurred between the 2016–2017 and 2017–2018 Northern Hemisphere seasons. A subset of the strain metadata only specified the year of isolation, which prevents us from providing a more precise time when this separation of clades became readily apparent. We expect that future analyses with additional sequence data and in-depth metadata will improve our understanding of when this phenomenon occurred.

5. Conclusions

Our phylogenetic results show that a substantial antigenic drift event occurred before and after 2017 in the pH1N1 HA protein. Comparative sequence analysis showed significant changes occurring at HA positions 147, 154, 159, 200, and 233 after the summer of 2017 in the Northern Hemisphere. A subset of these positions was found to undergo selection pressure and to coevolve as a cluster, which likely improved overall fitness. We expect these findings could be relevant to ongoing evolutionary studies, as well as future vaccine-strain prediction efforts.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes13050909/s1, Table S1: Table of RAxML quality scores and metrics: Robinson–Foulds distances and bootstrapping convergence; Table S2: Table of BEAST phylogenetic tree quality scores and metrics; Table S3: Table of meta-CATS results show amino acid positions that significantly differ between pre-2017 and post-2017 using chi-square algorithm on geographically diverse datasets; Table S4: Table of combined meta-CATS and HyPhy results show codons predicted to undergo selection pressure and significantly differ among geographically diverse datasets; Table S5: Table of MISTIC coevolution values for each dataset; Table S6: Table of summarized results from all algorithms across all residues; Figure S1: Figure of MW tree (RAxML-NG); Figure S2: Figure of MW Tree (BEAST); Figure S3: Figure of USA tree (RAxML-NG); Figure S4: Figure of USA tree (BEAST); Figure S5: Figure of Euro tree (RAxML-NG); Figure S6: Figure of Euro tree (BEAST); Figure S7: Figure of Northern Hemisphere tree (RAxML-NG); Figure S8: Figure of Northern Hemisphere tree (BEAST); Figure S9: Figure of bird's-eye view of hemagglutinin dimer; Figure S10: Figure of side view of hemagglutinin dimer; Figure S11: Figure to visualize coevolution results as a circos plot.

References

1. Morens, D.M.; Fauci, A.S. The 1918 Influenza Pandemic: Insights for the 21st Century. *J. Infect. Dis.* **2007**, *195*, 1018–1028. [CrossRef] [PubMed]
2. Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team; Dawood, F.S.; Jain, S.; Finelli, L.; Shaw, M.W.; Lindstrom, S.; Garten, R.J.; Gubareva, L.V.; Xu, X.; Bridges, C.B.; et al. Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans. *N. Engl. J. Med.* **2009**, *360*, 2605–2615. [CrossRef] [PubMed]
3. Shrestha, S.S.; Swerdlow, D.L.; Borse, R.H.; Prabhu, V.S.; Finelli, L.; Atkins, C.Y.; Owusu-Edusei, K.; Bell, B.; Mead, P.S.; Biggerstaff, M.; et al. Estimating the Burden of 2009 Pandemic Influenza A (H1N1) in the United States (April 2009-April 2010). *Clin. Infect. Dis.* **2011**, *52* (Suppl. S1), S75–S82. [CrossRef] [PubMed]
4. Henritzi, D.; Petric, P.P.; Lewis, N.S.; Graaf, A.; Pessia, A.; Starick, E.; Breithaupt, A.; Strebelow, G.; Luttermann, C.; Parker, L.M.K.; et al. Surveillance of European Domestic Pig Populations Identifies an Emerging Reservoir of Potentially Zoonotic Swine Influenza A Viruses. *Cell Host Microbe* **2020**, *28*, 614–627.e6. [CrossRef]
5. Rambo-Martin, B.L.; Keller, M.W.; Wilson, M.M.; Nolting, J.M.; Anderson, T.K.; Vincent, A.L.; Bagal, U.R.; Jang, Y.; Neuhaus, E.B.; Davis, C.T.; et al. Influenza A Virus Field Surveillance at a Swine-Human Interface. *mSphere* **2020**, *5*, e00822-19. [CrossRef]
6. Parvin, J.D.; Moscona, A.; Pan, W.T.; Leider, J.M.; Palese, P. Measurement of the Mutation Rates of Animal Viruses: Influenza A Virus and Poliovirus Type 1. *J. Virol.* **1986**, *59*, 377–383. [CrossRef]
7. Wille, M.; Tolf, C.; Avril, A.; Latorre-Margalef, N.; Wallerström, S.; Olsen, B.; Waldenström, J. Frequency and Patterns of Reassortment in Natural Influenza A Virus Infection in a Reservoir Host. *Virology* **2013**, *443*, 150–160. [CrossRef]
8. Nelson, M.I.; Detmer, S.E.; Wentworth, D.E.; Tan, Y.; Schwartzbard, A.; Halpin, R.A.; Stockwell, T.B.; Lin, X.; Vincent, A.L.; Gramer, M.R.; et al. Genomic Reassortment of Influenza A Virus in North American Swine, 1998-2011. *J. Gen. Virol.* **2012**, *93*, 2584–2589. [CrossRef]
9. Valkenburg, S.A.; Rutigliano, J.A.; Ellebedy, A.H.; Doherty, P.C.; Thomas, P.G.; Kedzierska, K. Immunity to Seasonal and Pandemic Influenza A Viruses. *Microbes Infect.* **2011**, *13*, 489–501. [CrossRef]
10. Clohisey, S.; Baillie, J.K. Host Susceptibility to Severe Influenza A Virus Infection. *Crit. Care* **2019**, *23*, 303. [CrossRef]
11. Chen, X.; Liu, S.; Goraya, M.U.; Maarouf, M.; Huang, S.; Chen, J.-L. Host Immune Response to Influenza A Virus Infection. *Front. Immunol.* **2018**, *9*, 320. [CrossRef] [PubMed]
12. Vasin, A.V.; Temkina, O.A.; Egorov, V.V.; Klotchenko, S.A.; Plotnikova, M.A.; Kiselev, O.I. Molecular Mechanisms Enhancing the Proteome of Influenza A Viruses: An Overview of Recently Discovered Proteins. *Virus Res.* **2014**, *185*, 53–63. [CrossRef] [PubMed]
13. Suzuki, Y.; Ito, T.; Suzuki, T.; Holland, R.E.; Chambers, T.M.; Kiso, M.; Ishida, H.; Kawaoka, Y. Sialic Acid Species as a Determinant of the Host Range of Influenza A Viruses. *J. Virol.* **2000**, *74*, 11825–11831. [CrossRef] [PubMed]
14. Du, R.; Cui, Q.; Rong, L. Competitive Cooperation of Hemagglutinin and Neuraminidase during Influenza A Virus Entry. *Viruses* **2019**, *11*, 458. [CrossRef]
15. Fujioka, Y.; Nishide, S.; Ose, T.; Suzuki, T.; Kato, I.; Fukuhara, H.; Fujioka, M.; Horiuchi, K.; Satoh, A.O.; Nepal, P.; et al. A Sialylated Voltage-Dependent Ca^{2+} Channel Binds Hemagglutinin and Mediates Influenza A Virus Entry into Mammalian Cells. *Cell Host Microbe* **2018**, *23*, 809–818.e5. [CrossRef] [PubMed]
16. Duvvuri, V.R.S.K.; Duvvuri, B.; Cuff, W.R.; Wu, G.E.; Wu, J. Role of Positive Selection Pressure on the Evolution of H5N1 Hemagglutinin. *Genom. Proteom. Bioinform.* **2009**, *7*, 47–56. [CrossRef]
17. Stray, S.J.; Pittman, L.B. Subtype- and Antigenic Site-Specific Differences in Biophysical Influences on Evolution of Influenza Virus Hemagglutinin. *Virol. J.* **2012**, *9*, 91. [CrossRef]
18. Castelán-Vega, J.A.; Magaña-Hernández, A.; Jiménez-Alberto, A.; Ribas-Aparicio, R.M. The Hemagglutinin of the Influenza A(H1N1)Pdm09 is Mutating towards Stability. *Adv. Appl. Bioinform. Chem.* **2014**, *7*, 37–44. [CrossRef]
19. Moore, K.A.; Ostrowsky, J.T.; Mehr, A.J.; Osterholm, M.T. CEIRS Pandemic Planning Committee Influenza Response Planning for the Centers of Excellence for Influenza Research and Surveillance: Science Preparedness for Enhancing Global Health Security. *Influenza Other. Respir. Viruses* **2020**, *14*, 444–451. [CrossRef]
20. Spackman, E.; Cardona, C.; Muñoz-Aguayo, J.; Fleming, S. Successes and Short Comings in Four Years of an International External Quality Assurance Program for Animal Influenza Surveillance. *PLoS ONE* **2016**, *11*, e0164261. [CrossRef]
21. Barr, I.G.; Cui, L.; Komadina, N.; Lee, R.T.; Lin, R.T.; Deng, Y.; Caldwell, N.; Shaw, R.; Maurer-Stroh, S. A New Pandemic Influenza A(H1N1) Genetic Variant Predominated in the Winter 2010 Influenza Season in Australia, New Zealand and Singapore. *Eurosurveillance* **2010**, *15*, 19692. [CrossRef] [PubMed]
22. Fielding, J.; Higgins, N.; Gregory, J.; Grant, K.; Catton, M.; Bergeri, I.; Lester, R.; Kelly, H. Pandemic H1N1 Influenza Surveillance in Victoria, Australia, April–September, 2009. *Eurosurveillance* **2009**, *14*, 19368. [CrossRef] [PubMed]
23. Kang, M.; Zhong, H.; He, J.; Rutherford, S.; Yang, F. Using Google Trends for Influenza Surveillance in South China. *PLoS ONE* **2013**, *8*, e55205. [CrossRef] [PubMed]
24. Potdar, V.A.; Chadha, M.S.; Jadhav, S.M.; Mullick, J.; Cherian, S.S.; Mishra, A.C. Genetic Characterization of the Influenza A Pandemic (H1N1) 2009 Virus Isolates from India. *PLoS ONE* **2010**, *5*, e9693. [CrossRef]
25. Jones, S.; Nelson-Sathi, S.; Wang, Y.; Prasad, R.; Rayen, S.; Nandel, V.; Hu, Y.; Zhang, W.; Nair, R.; Dharmaseelan, S.; et al. Evolutionary, Genetic, Structural Characterization and Its Functional Implications for the Influenza A (H1N1) Infection Outbreak in India from 2009 to 2017. *Sci. Rep.* **2019**, *9*, 14690. [CrossRef]

26. Elderfield, R.A.; Watson, S.J.; Godlee, A.; Adamson, W.E.; Thompson, C.I.; Dunning, J.; Fernandez-Alonso, M.; Blumenkrantz, D.; Hussell, T.; MOSAIC Investigators; et al. Accumulation of Human-Adapting Mutations during Circulation of A(H1N1)Pdm09 Influenza Virus in Humans in the United Kingdom. *J. Virol.* **2014**, *88*, 13269–13283. [CrossRef]

27. Brammer, L.; Blanton, L.; Epperson, S.; Mustaquim, D.; Bishop, A.; Kniss, K.; Dhara, R.; Nowell, M.; Kamimoto, L.; Finelli, L. Surveillance for Influenza during the 2009 Influenza A (H1N1) Pandemic-United States, April 2009-March 2010. *Clin. Infect. Dis.* **2011**, *52* (Suppl. S1), S27–S35. [CrossRef]

28. Theo, A.; Liwewe, M.; Ndumba, I.; Mupila, Z.; Tambatamba, B.; Mutemba, C.; Somwe, S.W.; Mwinga, A.; Tempia, S.; Monze, M. Influenza Surveillance in Zambia, 2008–2009. *J. Infect. Dis.* **2012**, *206* (Suppl. S1), S173–S177. [CrossRef]

29. Caton, A.J.; Brownlee, G.G.; Yewdell, J.W.; Gerhard, W. The Antigenic Structure of the Influenza Virus A/PR/8/34 Hemagglutinin (H1 Subtype). *Cell* **1982**, *31*, 417–427. [CrossRef]

30. Lee, A.J.; Das, S.R.; Wang, W.; Fitzgerald, T.; Pickett, B.E.; Aevermann, B.D.; Topham, D.J.; Falsey, A.R.; Scheuermann, R.H. Diversifying Selection Analysis Predicts Antigenic Evolution of 2009 Pandemic H1N1 Influenza A Virus in Humans. *J. Virol.* **2015**, *89*, 5427–5440. [CrossRef]

31. Maurer-Stroh, S.; Lee, R.T.C.; Eisenhaber, F.; Cui, L.; Phuah, S.P.; Lin, R.T. A New Common Mutation in the Hemagglutinin of the 2009 (H1N1) Influenza A Virus. *PLoS Curr.* **2010**, *2*, RRN1162. [CrossRef] [PubMed]

32. Sakabe, S.; Ozawa, M.; Takano, R.; Iwastuki-Horimoto, K.; Kawaoka, Y. Mutations in PA, NP, and HA of a Pandemic (H1N1) 2009 Influenza Virus Contribute to Its Adaptation to Mice. *Virus Res.* **2011**, *158*, 124–129. [CrossRef] [PubMed]

33. Ginting, T.E.; Shinya, K.; Kyan, Y.; Makino, A.; Matsumoto, N.; Kaneda, S.; Kawaoka, Y. Amino Acid Changes in Hemagglutinin Contribute to the Replication of Oseltamivir-Resistant H1N1 Influenza Viruses. *J. Virol.* **2012**, *86*, 121–127. [CrossRef] [PubMed]

34. Al Khatib, H.A.; Al Thani, A.A.; Yassine, H.M. Evolution and Dynamics of the Pandemic H1N1 Influenza Hemagglutinin Protein from 2009 to 2017. *Arch. Virol.* **2018**, *163*, 3035–3049. [CrossRef]

35. Zhang, Y.; Aevermann, B.D.; Anderson, T.K.; Burke, D.F.; Dauphin, G.; Gu, Z.; He, S.; Kumar, S.; Larsen, C.N.; Lee, A.J.; et al. Influenza Research Database: An Integrated Bioinformatics Resource for Influenza Virus Research. *Nucleic Acids Res.* **2017**, *45*, D466–D474. [CrossRef]

36. MapChart: Create Your Own Custom Map. Available online: https://mapchart.net/index.html (accessed on 4 August 2021).

37. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]

38. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2–a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef]

39. Pickett, B.E.; Liu, M.; Sadat, E.L.; Squires, R.B.; Noronha, J.M.; He, S.; Jen, W.; Zaremba, S.; Gu, Z.; Zhou, L.; et al. Metadata-Driven Comparative Analysis Tool for Sequences (Meta-CATS): An Automated Process for Identifying Significant Sequence Variations That Correlate with Virus Attributes. *Virology* **2013**, *447*, 45–51. [CrossRef]

40. Strait, B.J.; Dewey, T.G. The Shannon Information Entropy of Protein Sequences. *Biophys. J.* **1996**, *71*, 148–155. [CrossRef]

41. Noronha, J.M.; Liu, M.; Squires, R.B.; Pickett, B.E.; Hale, B.G.; Air, G.M.; Galloway, S.E.; Takimoto, T.; Schmolke, M.; Hunt, V.; et al. Influenza Virus Sequence Feature Variant Type Analysis: Evidence of a Role for NS1 in Influenza Virus Host Range Restriction. *J. Virol.* **2012**, *86*, 5857–5866. [CrossRef]

42. Lam, H.M.; Ratmann, O.; Boni, M.F. Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. *Mol. Biol. Evol.* **2018**, *35*, 247–251. [CrossRef] [PubMed]

43. Arenas, M.; Posada, D. The Effect of Recombination on the Reconstruction of Ancestral Sequences. *Genetics* **2010**, *184*, 1133–1139. [CrossRef] [PubMed]

44. Schierup, M.H.; Hein, J. Consequences of Recombination on Traditional Phylogenetic Analysis. *Genetics* **2000**, *156*, 879–891. [CrossRef] [PubMed]

45. Kozlov, A.M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference. *Bioinformatics* **2019**, *35*, 4453–4455. [CrossRef] [PubMed]

46. Robinson, D.F.; Foulds, L.R. Comparison of Phylogenetic Trees. *Math. Biosci.* **1981**, *53*, 131–147. [CrossRef]

47. Guindon, S.; Lethiec, F.; Duroux, P.; Gascuel, O. PHYML Online–A Web Server for Fast Maximum Likelihood-Based Phylogenetic Inference. *Nucleic Acids Res.* **2005**, *33*, W557–W559. [CrossRef]

48. Kosakovsky Pond, S.L.; Poon, A.F.Y.; Velazquez, R.; Weaver, S.; Hepler, N.L.; Murrell, B.; Shank, S.D.; Magalis, B.R.; Bouvier, D.; Nekrutenko, A.; et al. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol. Biol. Evol.* **2020**, *37*, 295–299. [CrossRef]

49. Murrell, B.; Wertheim, J.O.; Moola, S.; Weighill, T.; Scheffler, K.; Kosakovsky Pond, S.L. Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genet.* **2012**, *8*, e1002764. [CrossRef]

50. Kosakovsky Pond, S.L.; Frost, S.D.W. Not so Different after All: A Comparison of Methods for Detecting Amino Acid Sites under Selection. *Mol. Biol. Evol.* **2005**, *22*, 1208–1222. [CrossRef]

51. Weaver, S.; Shank, S.D.; Spielman, S.J.; Li, M.; Muse, S.V.; Kosakovsky Pond, S.L. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol. Biol. Evol.* **2018**, *35*, 773–777. [CrossRef]

52. Drummond, A.J.; Suchard, M.A.; Xie, D.; Rambaut, A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **2012**, *29*, 1969–1973. [CrossRef] [PubMed]

53. Bouckaert, R.; Vaughan, T.G.; Barido-Sottani, J.; Duchêne, S.; Fourment, M.; Gavryushkina, A.; Heled, J.; Jones, G.; Kühnert, D.; De Maio, N.; et al. BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **2019**, *15*, e1006650. [CrossRef] [PubMed]

54. Barido-Sottani, J.; Bošková, V.; Plessis, L.D.; Kühnert, D.; Magnus, C.; Mitov, V.; Müller, N.F.; PecErska, J.; Rasmussen, D.A.; Zhang, C.; et al. Taming the BEAST-A Community Teaching Material Resource for BEAST 2. *Syst. Biol.* **2018**, *67*, 170–174. [CrossRef]

55. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [CrossRef]

56. Jansson, J.; Shen, C.; Sung, W.-K. Improved Algorithms for Constructing Consensus Trees. In Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA, USA, 6–8 January 2013.

57. Simonetti, F.L.; Teppa, E.; Chernomoretz, A.; Nielsen, M.; Marino Buslje, C. MISTIC: Mutual Information Server to Infer Coevolution. *Nucleic Acids* **2013**, *41*, W8–W14. [CrossRef]

58. Boonnak, K.; Mansanguan, C.; Schuerch, D.; Boonyuen, U.; Lerdsamran, H.; Jiamsomboon, K.; Sae Wang, F.; Huntrup, A.; Prasertsopon, J.; Kosoltanapiwat, N.; et al. Molecular Characterization of Seasonal Influenza A and B from Hospitalized Patients in Thailand in 2018–2019. *Viruses* **2021**, *13*, 977. [CrossRef]

59. Ma, Y.; Liu, K.; Yin, Y.; Qin, J.; Zhou, Y.-H.; Yang, J.; Li, S.; Poon, L.L.M.; Zhang, C. The Phylodynamics of Seasonal Influenza A/H1N1pdm Virus in China Between 2009 and 2019. *Front. Microbiol.* **2020**, *11*, 735. [CrossRef]

60. Matsuzaki, Y.; Sugawara, K.; Nakauchi, M.; Takahashi, Y.; Onodera, T.; Tsunetsugu-Yokota, Y.; Matsumura, T.; Ato, M.; Kobayashi, K.; Shimotai, Y.; et al. Epitope Mapping of the Hemagglutinin Molecule of A/(H1N1)Pdm09 Influenza Virus by Using Monoclonal Antibody Escape Mutants. *J. Virol.* **2014**, *88*, 12364–12373. [CrossRef]

61. Price, J.V.; Jarrell, J.A.; Furman, D.; Kattah, N.H.; Newell, E.; Dekker, C.L.; Davis, M.M.; Utz, P.J. Characterization of Influenza Vaccine Immunogenicity Using Influenza Antigen Microarrays. *PLoS ONE* **2013**, *8*, e64555. [CrossRef]

62. Zhao, R.; Cui, S.; Guo, L.; Wu, C.; Gonzalez, R.; Paranhos-Baccalà, G.; Vernet, G.; Wang, J.; Hung, T. Identification of a Highly Conserved H1 Subtype-Specific Epitope with Diagnostic Potential in the Hemagglutinin Protein of Influenza A Virus. *PLoS ONE* **2011**, *6*, e23374. [CrossRef]

63. Richards, K.A.; Chaves, F.A.; Krafcik, F.R.; Topham, D.J.; Lazarski, C.A.; Sant, A.J. Direct Ex Vivo Analyses of HLA-DR1 Transgenic Mice Reveal an Exceptionally Broad Pattern of Immunodominance in the Primary HLA-DR1-Restricted CD4 T-Cell Response to Influenza Virus Hemagglutinin. *J. Virol.* **2007**, *81*, 7608–7619. [CrossRef] [PubMed]

64. Yang, J.; James, E.A.; Huston, L.; Danke, N.A.; Liu, A.W.; Kwok, W.W. Multiplex Mapping of CD4 T Cell Epitopes Using Class II Tetramers. *Clin. Immunol.* **2006**, *120*, 21–32. [CrossRef] [PubMed]

65. Richards, K.A.; Chaves, F.A.; Sant, A.J. The Memory Phase of the CD4 T-Cell Response to Influenza Virus Infection Maintains Its Diverse Antigen Specificity. *Immunology* **2011**, *133*, 246–256. [CrossRef] [PubMed]

66. Chow, I.-T.; James, E.A.; Tan, V.; Moustakas, A.K.; Papadopoulos, G.K.; Kwok, W.W. DRB1*12:01 Presents a Unique Subset of Epitopes by Preferring Aromatics in Pocket 9. *Mol. Immunol.* **2012**, *50*, 26–34. [CrossRef] [PubMed]

67. Babon, J.A.B.; Cruz, J.; Orphin, L.; Pazoles, P.; Co, M.D.T.; Ennis, F.A.; Terajima, M. Genome-Wide Screening of Human T-Cell Epitopes in Influenza A Virus Reveals a Broad Spectrum of CD4(+) T-Cell Responses to Internal Proteins, Hemagglutinins, and Neuraminidases. *Hum. Immunol.* **2009**, *70*, 711–721. [CrossRef] [PubMed]

68. Russell, R.J.; Gamblin, S.J.; Haire, L.F.; Stevens, D.J.; Xiao, B.; Ha, Y.; Skehel, J.J. H1 and H7 Influenza Haemagglutinin Structures Extend a Structural Classification of Haemagglutinin Subtypes. *Virology* **2004**, *325*, 287–296. [CrossRef] [PubMed]

69. Das, M.; Basu, G. Glycine Rescue of β-Sheets from Cis-Proline. *J. Am. Chem. Soc.* **2012**, *134*, 16536–16539. [CrossRef]

70. Vázquez-Pérez, J.A.; De La Rosa-Zamboni, D.; Vega-Sánchez, Á.E.; Gutiérrez-González, L.H.; Téllez-Navarrete, N.A.; Campos, F.; Guadarrama-Pérez, C.; Sandoval, J.L.; Castillejos-López, M.; Jiménez-Juárez, R.N.; et al. Amino Acid Changes in HA and Determinants of Pathogenicity Associated with Influenza Virus A H1N1pdm09 during the Winter Seasons 2015–2016 and 2016–2017 in Mexico. *Virus Res.* **2019**, *272*, 197731. [CrossRef]

71. Byarugaba, D.K.; Erima, B.; Millard, M.; Kibuuka, H.; Lkwago, L.; Bwogi, J.; Mimbe, D.; Kiconco, J.B.; Tugume, T.; Mworozi, E.A.; et al. Whole-Genome Analysis of Influenza A(H1N1)Pdm09 Viruses Isolated in Uganda from 2009 to 2011. *Influenza Other. Respir. Viruses* **2016**, *10*, 486–492. [CrossRef]

72. Anderson, C.S.; Ortega, S.; Chaves, F.A.; Clark, A.M.; Yang, H.; Topham, D.J.; DeDiego, M.L. Natural and Directed Antigenic Drift of the H1 Influenza Virus Hemagglutinin Stalk Domain. *Sci. Rep.* **2017**, *7*, 14614. [CrossRef]

73. Liu, B.; Wang, Y.; Liu, Y.; Chen, Y.; Liu, Y.; Cong, X.; Ji, Y.; Gao, Y. Molecular Evolution and Characterization of Hemagglutinin and Neuraminidase of Influenza A(H1N1)Pdm09 Viruses Isolated in Beijing, China, during the 2017–2018 and 2018–2019 Influenza Seasons. *Arch. Virol.* **2021**, *166*, 179–189. [CrossRef] [PubMed]

74. Yan, Y.; Ou, J.; Zhao, S.; Ma, K.; Lan, W.; Guan, W.; Wu, X.; Zhang, J.; Zhang, B.; Zhao, W.; et al. Characterization of Influenza A and B Viruses Circulating in Southern China During the 2017-2018 Season. *Front. Microbiol.* **2020**, *11*, 1079. [CrossRef] [PubMed]

75. Nayak, J.L.; Richards, K.A.; Chaves, F.A.; Sant, A.J. Analyses of the Specificity of CD4 T Cells during the Primary Immune Response to Influenza Virus Reveals Dramatic MHC-Linked Asymmetries in Reactivity to Individual Viral Proteins. *Viral. Immunol.* **2010**, *23*, 169–180. [CrossRef] [PubMed]

76. Cusick, M.F.; Wang, S.; Eckels, D.D. In Vitro Responses to Avian Influenza H5 by Human CD4 T Cells. *J. Immunol.* **2009**, *183*, 6432–6441. [CrossRef] [PubMed]

77. Al-Majhdi, F.N. Structure of the Sialic Acid Binding Site in Influenza A Virus: Hemagglutinin. *J. Biol. Sci.* **2007**, *7*, 113–122. [CrossRef]

78. Butler, J.; Hooper, K.A.; Petrie, S.; Lee, R.; Maurer-Stroh, S.; Reh, L.; Guarnaccia, T.; Baas, C.; Xue, L.; Vitesnik, S.; et al. Estimating the Fitness Advantage Conferred by Permissive Neuraminidase Mutations in Recent Oseltamivir-Resistant A(H1N1)Pdm09 Influenza Viruses. *PLoS Pathog.* **2014**, *10*, e1004065. [CrossRef]

79. Zell, R.; Groth, M.; Krumbholz, A.; Lange, J.; Philipps, A.; Dürrwald, R. Cocirculation of Swine H1N1 Influenza A Virus Lineages in Germany. *Viruses* **2020**, *12*, 762. [CrossRef]

80. Soundararajan, V.; Tharakaraman, K.; Raman, R.; Raguram, S.; Shriver, Z.; Sasisekharan, V.; Sasisekharan, R. Extrapolating from Sequence–the 2009 H1N1 "swine" Influenza Virus. *Nat. Biotechnol.* **2009**, *27*, 510–513. [CrossRef]

81. Herrera, M.T.; Gonzalez, Y.; Juárez, E.; Hernández-Sánchez, F.; Carranza, C.; Sarabia, C.; Guzman-Beltran, S.; Manjarrez, M.E.; Muñoz-Torrico, M.; Garcia-Garcia, L.; et al. Humoral and Cellular Responses to a Non-Adjuvanted Monovalent H1N1 Pandemic Influenza Vaccine in Hospital Employees. *BMC Infect. Dis.* **2013**, *13*, 544. [CrossRef]

82. Mozdzanowska, K.; Feng, J.; Eid, M.; Kragol, G.; Cudic, M.; Otvos, L.; Gerhard, W. Induction of Influenza Type A Virus-Specific Resistance by Immunization of Mice with a Synthetic Multiple Antigenic Peptide Vaccine That Contains Ectodomains of Matrix Protein 2. *Vaccine* **2003**, *21*, 2616–2626. [CrossRef]

83. Gerhard, W.; Haberman, A.M.; Scherle, P.A.; Taylor, A.H.; Palladino, G.; Caton, A.J. Identification of Eight Determinants in the Hemagglutinin Molecule of Influenza Virus A/PR/8/34 (H1N1) Which Are Recognized by Class II-Restricted T Cells from BALB/c Mice. *J. Virol.* **1991**, *65*, 364–372. [CrossRef] [PubMed]

84. Eisenlohr, L.C.; Gerhard, W.; Hackett, C.J. Acid-Induced Conformational Modification of the Hemagglutinin Molecule Alters Interaction of Influenza Virus with Antigen-Presenting Cells. *J. Immunol.* **1988**, *141*, 1870–1876. [PubMed]

85. Yang, J.; James, E.; Gates, T.J.; DeLong, J.H.; LaFond, R.E.; Malhotra, U.; Kwok, W.W. CD4+ T Cells Recognize Unique and Conserved 2009 H1N1 Influenza Hemagglutinin Epitopes after Natural Infection and Vaccination. *Int. Immunol.* **2013**, *25*, 447–457. [CrossRef] [PubMed]

86. Tai, S.-H.S.; Agafitei, O.; Gao, Z.; Liggins, R.; Petric, M.; Withers, S.G.; Niikura, M. Difluorosialic Acids, Potent Novel Influenza Virus Neuraminidase Inhibitors, Induce Fewer Drug Resistance-Associated Neuraminidase Mutations than Does Oseltamivir. *Virus Res.* **2015**, *210*, 126–132. [CrossRef] [PubMed]

87. Tse, H.; Kao, R.Y.T.; Wu, W.L.; Lim, W.W.L.; Chen, H.; Yeung, M.Y.; Woo, P.C.Y.; Sze, K.-H.; Yuen, K.-Y. Structural Basis and Sequence Co-Evolution Analysis of the Hemagglutinin Protein of Pandemic Influenza A/H1N1 (2009) Virus. *Exp. Biol. Med.* **2011**, *236*, 915–925. [CrossRef] [PubMed]

88. Jayaraman, A.; Pappas, C.; Raman, R.; Belser, J.A.; Viswanathan, K.; Shriver, Z.; Tumpey, T.M.; Sasisekharan, R. A Single Base-Pair Change in 2009 H1N1 Hemagglutinin Increases Human Receptor Affinity and Leads to Efficient Airborne Viral Transmission in Ferrets. *PLoS ONE* **2011**, *6*, e17616. [CrossRef]

89. Brownlee, G.G.; Fodor, E. The Predicted Antigenicity of the Haemagglutinin of the 1918 Spanish Influenza Pandemic Suggests an Avian Origin. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2001**, *356*, 1871–1876. [CrossRef]

90. Al Khatib, H.A.; Al Thani, A.A.; Gallouzi, I.; Yassine, H.M. Epidemiological and Genetic Characterization of PH1N1 and H3N2 Influenza Viruses Circulated in MENA Region during 2009–2017. *BMC Infect. Dis.* **2019**, *19*, 314. [CrossRef]

91. Boni, M.F.; de Jong, M.D.; van Doorn, H.R.; Holmes, E.C. Guidelines for Identifying Homologous Recombination Events in Influenza A Virus. *PLoS ONE* **2010**, *5*, e10434. [CrossRef]

92. Hasan, A.; Sasaki, T.; Phadungsombat, J.; Koketsu, R.; Rahim, R.; Ara, N.; Biswas, S.M.; Yonezawa, R.; Nakayama, E.E.; Rahman, M.; et al. Genetic Analysis of Influenza A/H1N1pdm Strains Isolated in Bangladesh in Early 2020. *Trop. Med. Infect. Dis.* **2022**, *7*, 38. [CrossRef]

93. Soli, R.; Kaabi, B.; Barhoumi, M.; Maktouf, C.; Ahmed, S.B.-H. Bayesian Phylogenetic Analysis of the Influenza-A Virus Genomes Isolated in Tunisia, and Determination of Potential Recombination Events. *Mol. Phylogenet. Evol.* **2019**, *134*, 253–268. [CrossRef] [PubMed]

94. Fusade-Boyer, M.; Pato, P.S.; Komlan, M.; Dogno, K.; Jeevan, T.; Rubrum, A.; Kouakou, C.K.; Couacy-Hymann, E.; Batawui, D.; Go-Maro, E.; et al. Evolution of Highly Pathogenic Avian Influenza A(H5N1) Virus in Poultry, Togo, 2018. *Emerg. Infect. Dis.* **2019**, *25*, 2287–2289. [CrossRef] [PubMed]