

# SCIENTIFIC REPORTS



OPEN

## Structures and functions linked to genome-wide adaptation of human influenza A viruses

Thorsten R. Klingen<sup>1</sup>, Jens Loers<sup>1</sup>, Stephanie Stanelle-Bertram<sup>3</sup>, Gülsah Gabriel<sup>3,4</sup> & Alice C. McHardy<sup>1,2</sup> 

Received: 2 November 2018

Accepted: 27 March 2019

Published online: 18 April 2019

Human influenza A viruses elicit short-term respiratory infections with considerable mortality and morbidity. While H3N2 viruses circulate for more than 50 years, the recent introduction of pH1N1 viruses presents an excellent opportunity for a comparative analysis of the genome-wide evolutionary forces acting on both subtypes. Here, we inferred patches of sites relevant for adaptation, i.e. being under positive selection, on eleven viral protein structures, from all available data since 1968 and correlated these with known functional properties. Overall, pH1N1 have more patches than H3N2 viruses, especially in the viral polymerase complex, while antigenic evolution is more apparent for H3N2 viruses. In both subtypes, NS1 has the highest patch and patch site frequency, indicating that NS1-mediated viral attenuation of host inflammatory responses is a continuously intensifying process, elevated even in the longtime-circulating subtype H3N2. We confirmed the resistance-causing effects of two pH1N1 changes against oseltamivir in NA activity assays, demonstrating the value of the resource for discovering functionally relevant changes. Our results represent an atlas of protein regions and sites with links to host adaptation, antiviral drug resistance and immune evasion for both subtypes for further study.

Infections with influenza A viruses cause short term respiratory infections with drastic health burdens and economic losses. Recurrent epidemics cause up to 650,000 deaths and 3 to 5 million cases of severe illness per year worldwide<sup>1,2</sup>. Two influenza subtypes are currently endemic in the human population: H3N2 influenza A (H3N2) viruses circulate since 1968 and H1N1 influenza A (pH1N1) viruses circulate since 2009<sup>3–6</sup>. Both lineages descend from pandemic outbreaks of reassortant viruses with segments of human and zoonotic origin<sup>5</sup>. The influenza genome consists of eight segments<sup>3</sup> and encodes fourteen known proteins<sup>7–10</sup>. Reassortant viruses inherit viral segments from different viruses co-infecting the same cell, which, in case a novel HA is introduced, meet a largely naïve human population and thus can spread widely, causing a pandemic<sup>3</sup>. For H3N2 viruses, the polymerase basic protein 2 (PB1) and hemagglutinin (HA) segments are of avian origin, with evidence of adaptive processes, and the remaining ones originate from the previously circulating human H2N2 subtype<sup>6,11</sup>. The pH1N1 virus arose from a reassortment of three porcine viruses<sup>5</sup>, creating a virus with some segments of prior avian origins (NA, M; and another lineage with PB2 and PA), some distantly related to a human lineage circulating until 1998 (HA, NP, NS) and some descending more recently from humans (PB1)<sup>5</sup>.

Human influenza A viruses are the prime example of a rapidly evolving pathogen engaging in a co-evolutionary arms-race with the host adaptive immune defenses. Due to their rapid evolution, the antigenic change of the major antigen to the humoral (B-cell) immune response, the surface protein HA, is monitored across the years<sup>2,12–16</sup>. Every couple of years, as susceptibility of the population becomes low due to prior infection or vaccinations, antigenically novel strains emerge and predominate in seasonal epidemics<sup>17</sup>. An antigenic match of vaccine viruses to the circulating viral population is achieved by regular updates of the vaccine composition<sup>17,18</sup>. Conversely, on the host side, upon infection or vaccination, specific antibodies are produced, recognizing B-cell epitopes (BCE) on HA, neuraminidase (NA) and matrix protein 2 (M2) and interfering with their function. In addition, cell-mediated immunity evokes T-cell responses against small peptides of internalized viral

<sup>1</sup>Department for Computational Biology of Infection Research, Helmholtz Center for Infection Research (HZI), Braunschweig, Germany. <sup>2</sup>German Center for Infection Research (DZIF), Braunschweig, Germany. <sup>3</sup>Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany. <sup>4</sup>University of Veterinary Medicine, Hannover, Germany. Correspondence and requests for materials should be addressed to A.C.M. (email: [Alice.McHardy@helmholtz-hzi.de](mailto:Alice.McHardy@helmholtz-hzi.de))

	pH1N1	sH3N2
PB2	1. 66, 62, 65	1. 67, 63
	2. 102, 104, 106	2. 141, 221
	3. 155, 156	3. 394, 398
	4. 184, 185	4. 451, 456, 460, 461, 463, 471, 475, 478, 480
	5. 251, 255	5. 613, 615
	6. 338, 340	6. 674, 676
	7. 351, 353	
	8. 394, 397	
	9. 411, 414	
	10. 441, 442	
	11. 451, 453	
	12. 385, 478, 480	
	13. 224, 524	
	14. 553, 615, 616	
	15. 660, 661	
	16. 547, 667	
	17. 673, 674, 676, 677	
PB1	1. 176, 179	1. 179, 181
	2. 216, 219	2. 374, 375
	3. 423, 469	3. 454, 459
	4. 525, 527, 530	4. 722, 743, 746, 751, 753
	5. 637, 640	
	6. 645, 652, 653	
	7. 753, 754	
PA	1. 20, 22	1. 225, 226
	2. 63, 65	2. 330, 332
	3. 231, 233	3. 348, 350
	4. 330, 332	4. 404, 407, 409
	5. 362, 363, 364	5. 421, 423
	6. 394, 395, 396	6. 552, 554
	7. 267, 269, 271, 272, 275, 400, 402	7. 712, 713
	8. 405, 407	
	9. 465, 485	
	10. 521, 369, 505	
	11. 322, 323, 325, 543	
	12. 552, 556	
HA1	1. 120, 121	1. 7, 9, 5, 6, 3
	2. 137, 139, 141, 143	2. 25, 31, 33
	3. 162, 163	3. 92, 94
	4. 128, 129, 149, 152, 185, 186, 190, 191	4. 121, 122, 124
	5. 195, 197	5. 157, 158, 159, 160
	6. 183, 215, 216	6. 188, 192, 193, 197, 198, 199
	7. 222, 223	7. 203, 212
	8. 203, 205, 232, 234, 238	8. 135, 137, 138, 140, 142, 144, 145, 186, 220, 221, 223, 225
	9. 112, 114, 116, 261, 171, 256, 258, 260	9. 229, 230
	10. 54, 82, 83, 84, 85, 86, 268, 270, 271, 272	10. 260, 261, 262
	11. 45, 47, 48, 277	11. 47, 48, 50, 51, 53, 272, 273
	12. 35, 295	12. 289, 291
HA2	NONE	1. 147, 149
NP	1. 61, 63, 66	1. 450, 451, 453
	2. 100, 101, 105	
	3. 119, 120	
	4. 128, 129, 130	
	5. 186, 189, 190, 260	
	6. 318, 373	
	7. 423, 425	
Continued		

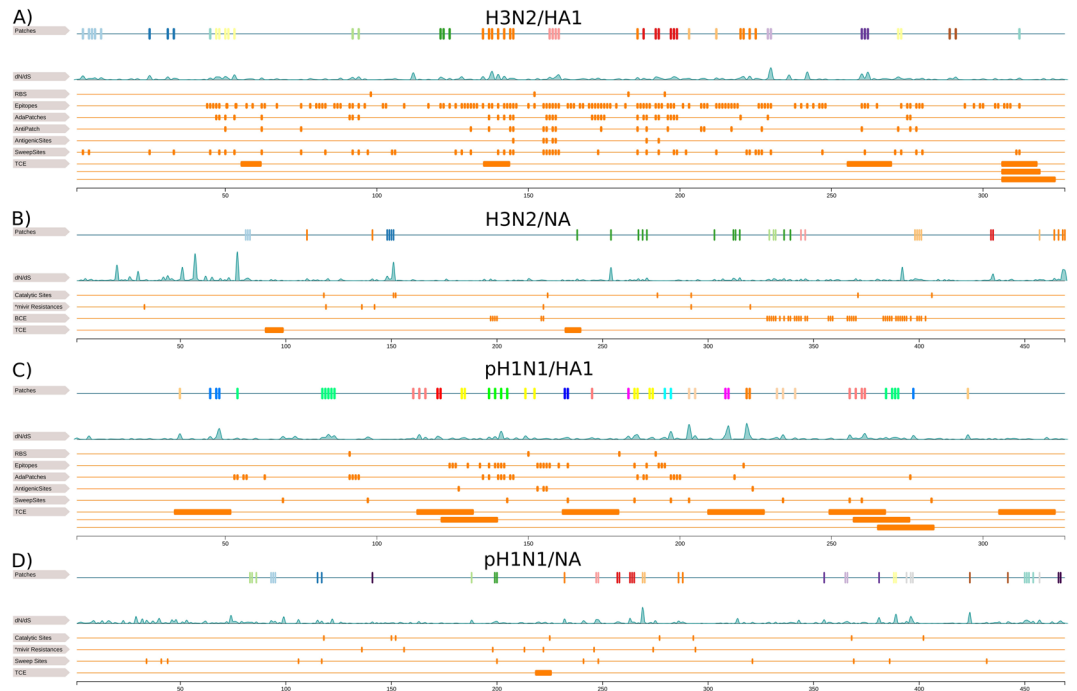
	pH1N1		sH3N2	
NA	1.	93, 94, 95	1.	<b>81, 82</b> , 83
	2.	115, <b>117</b>	2.	149, <b>150</b> , 151, 148
	3.	84, 86, 83, 188	3.	329, 331, <b>332</b>
	4.	199, <b>200</b>	4.	238, 254, <b>267</b> , 269, 271, 303, 312, 313, 315, 336, <b>339</b>
	5.	247, 248	5.	346, 344
	6.	257, 258, 263, <b>264</b> , 265	6.	434, 435
	7.	269, <b>270</b>	7.	398, <b>399</b> , 400, <b>401</b> , 457
	8.	232, 286, 288	8.	110, 141, <b>464</b> , 466, <b>468</b> , 469
	9.	365, 366		
	10.	355, 381		
	11.	388, 389		
	12.	424, 442		
	13.	450, 451, 452, 454		
	14.	394, 396, <b>397</b> , 457		
	15.	141, 466, 467		
M1	1.	30, 33		NONE
	2.	77, <b>80</b>		
M2	1.	25, 27, 28	1.	25, 26, 27, 28
	2.	39, 41, 42, 43, 44, 46, 48	2.	39, 43, 48, <b>51</b> , 55, <b>56</b> , 54
	3.	55, 54		
NS1	1.	<b>2</b> , 3, 4, 6	1.	22, 23, <b>26</b>
	2.	22, 23, 24, 25	2.	101, 103
	3.	<b>55</b> , 56, 60	3.	135, 74, 76, 78, 79, 80, 81, <b>82</b> , 84, 85, 86, 90, 91, 137
	4.	71, 72, 73, 75, 76, 78, <b>80</b> , 81, 85	4.	55, <b>56</b> , <b>59</b> , 60, 162, 164, 165, 167
	5.	94, 96, 98	5.	171, 174
	6.	101, 103	6.	197, 200
	7.	135, <b>90</b>	7.	122, <b>123</b> , 127, <b>129</b> , 150, 155, 157, 180, 182, 193, 194, 204, 206, 209
	8.	180, 181, 184, 185	8.	214, 220
	9.	109, 111, 112, 114, 115, <b>123</b> , <b>125</b> , 129, 150, 154, 155, 157, 159, 161, 162, 164, 166, 167, 194, 195	9.	224, 225, <b>226</b> , 227, 228, 230
	10.	171, 174, 200, 202, <b>205</b> , 206, 209		
	11.	212, 213		
NS2		NONE		NONE

**Table 1.** Patches and residues under positive selection for all analyzed influenza proteins. Overview of all patches and patch sites for each protein of pH1N1 and H2N3 viruses. Sweep-related sites are marked in bold.

proteins, representing T-cell epitopes (TCE), which are digested and presented on the cell surface by the major histocompatibility complex (MHC). The innate immune system also recognizes pathogen-associated molecular patterns (PAMPs) as determinants of viruses and activates cellular antiviral responses, like activation of interferon (IFN)-induced proteins, to counteract viral replication<sup>19,20</sup>. In addition, human influenza viruses evolve towards resistance against antivirals<sup>21</sup> and further adaptations to their host after initial establishment occur<sup>5</sup>.

Evolutionary processes such as adaptation to specific environmental conditions can be studied on the level of genes, individual sites or even protein regions<sup>15,16,22–29</sup>. Various methods include biophysical, geometrical and evolutionary features to describe protein sites relevant for viral evolution<sup>30</sup>. A widely used method determines the rates of non-synonymous to synonymous changes ( $dN/dS$  ratio) in a phylogeny<sup>13,31</sup>. A significant excess of  $dN$  to  $dS$ , or  $dN/dS > 1$ , provides evidence for positive selection, assuming that synonymous changes are neutral. This indicates that adaptation is taking place and that the changes of the respective genetic elements have led to a more favorable phenotype<sup>32</sup>. Methods measuring  $dN/dS$  should best be applied to analyze selection across, not within populations, and temporal influenza data may be considered as a series of independently evolving populations<sup>33</sup>.

While H3N2 viruses have been circulating for more than 50 years and presumably do not require more adaptation to the human host, the recent introduction of pH1N1 viruses presents an excellent opportunity for a comparative analysis of the genome-wide evolutionary forces acting on both subtypes. We here sought to determine protein regions with signs of positive selection in both subtypes, and correlate them with known functional properties, to get a better understanding of the forces shaping the viral genome-wide evolution. We studied all available longitudinal sequence data for H3N2 and pH1N1 viruses collected since 1968 and 2009, respectively, across all eleven available protein structures. Though  $dN/dS$  values have been analyzed already for sites<sup>34</sup> or entire proteins of H3N2<sup>35</sup> and pH1N1<sup>36</sup>, to our knowledge this is the first comparative and compressive study of the structures and functions linking to the genome-wide adaptation of both circulating influenza A subtypes.



**Figure 1.** Linear representation of surface proteins. This figure illustrates the overlap of patch sites with known functional sites in H3N2/HA1 (A) H3N2/NA (B) pH1N1/HA1 (C) and pH1N1/NA (D). In each subfigure, the first row is a linear representation of patches and the second row shows the  $dN/dS$  distribution of the protein. In (A), we compare with the RBS<sup>44</sup>, BCE sites<sup>40,41</sup>, “adapatch” sites<sup>22</sup>, “antigenic patch” sites<sup>23</sup>, key antigenic sites<sup>42</sup>, sweep-related sites<sup>16</sup> and TCE sites<sup>34</sup>. In (B), we compare with the catalytic sites<sup>65</sup>, oseltamivir and zanamivir resistance sites<sup>21</sup>, BCE sites and TCE sites<sup>34</sup>. In (C), we compare with the RBS<sup>49</sup>, BCE sites<sup>45,46</sup>, “adapatch” sites<sup>22</sup>, key antigenic sites<sup>47</sup>, sweep-related sites<sup>16</sup>, TCE sites<sup>34</sup>. In (D), we compare with catalytic sites<sup>65,66</sup>, oseltamivir and zanamivir resistance sites<sup>21</sup>, sweep-related sites<sup>16</sup> and TCE sites.

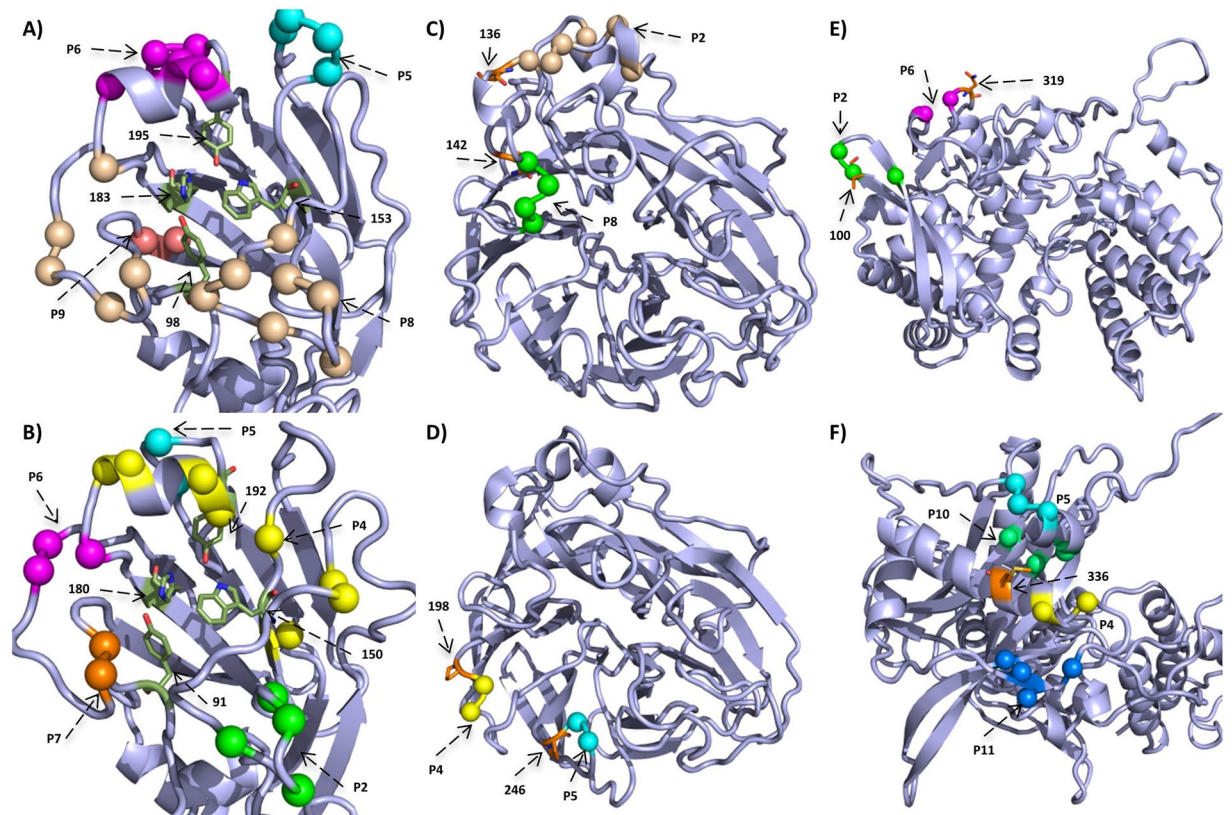
## Results

**Global trends.** We calculated  $dN/dS$  values for all proteins of pH1N1 and H3N2 viruses (Supplementary Fig. 1). Notably, the HA1 subunit, historically considered the main driver of adaptation of seasonal influenza viruses<sup>12,13</sup>, did not show the most evidence for positive selection. It was ranked fifth for both subtypes when jointly analyzing all time periods, as in agreement with another study of pH1N1<sup>36</sup> and partly (NA, NS1 and M2 among the top four) in agreement for H3N2 studies on smaller data sets<sup>34,35</sup>. Instead, NA had the highest mean  $dN/dS$  value for both subtypes, followed by M2, NS1 and NS2. NP and the PB1 proteins had the lowest mean  $dN/dS$  value in both subtypes (Supplementary Table 3). We confirmed this finding by re-calculating the  $dN/dS$  statistics using the Suzuki-Gojobori counting approach implemented in HyPhy SLAC<sup>37</sup> (Supplementary Table 3).

The Kolmogorov–Smirnov-test (KS-test;  $H_0$ :  $dN/dS$  distribution of pH1N1 protein is smaller than  $dN/dS$  distribution of H3N2 protein) showed that  $dN/dS$  distributions of pH1N1 proteins were significantly larger than those of H3N2 proteins, except for HA2, M2 and NS1 (Supplementary Table 4). This was especially apparent for the polymerase proteins PB2, PB1 and PA, which are of known importance for host adaptation<sup>38</sup>. We also compared the  $dN/dS$  distribution protein-wise between the pandemic and the post-pandemic phase. We followed the definition by Elderfield *et al.*<sup>39</sup> for the onset of the post-pandemic phase in the winter season 11 N and found that selection was acting more strongly during the pandemic than in the post-pandemic time, except for NA and HA2. Our findings agree with those by Su *et al.* (2015) (Supplementary Table 5).

Following the methodology in Klingen *et al.* (2018), we calculated the average number of sweep-related amino acid changes for pH1N1 and H3N2 viruses and compared the number of sweeps fixed over time across all eight segments (Supplementary Fig. 2). The number of sweep-related changes becoming fixed per year was greatest for HA1 and NA in both viruses but substantially increased in H3N2 relative to pH1N1 viruses, in line with a lesser need for antigenic immune escape for the latter in the early years of circulating in the human population. For M2 and NS2, it was the opposite case and all remaining proteins had similar rates.

To investigate this further, we identified clusters (patches) of sites under positive selection on the protein structure based on  $dN/dS$  values and protein structure models. We used a revised methodology based on our prior work described in Tusche *et al.* (2012) and Kratsch *et al.* (2016) that now includes a fully automated parameter optimization, consideration of buried and exposed protein sites and direct inclusion of  $dN/dS$  measurements (Methods). We applied the novel method to all proteins in both subtypes for which we have a suitable homology model and sufficient evidence for positive selection within the regions of the available partial structures (Table 1). The number of patches and of sites clustered into patches for pH1N1 viruses exceeded those for H3N2 viruses, indicating that more regions of pH1N1 are currently suboptimally adapted to the human host. This was despite of H3N2 data covering a longer time span (1968–2016; though data largely originates from after 1999) than for

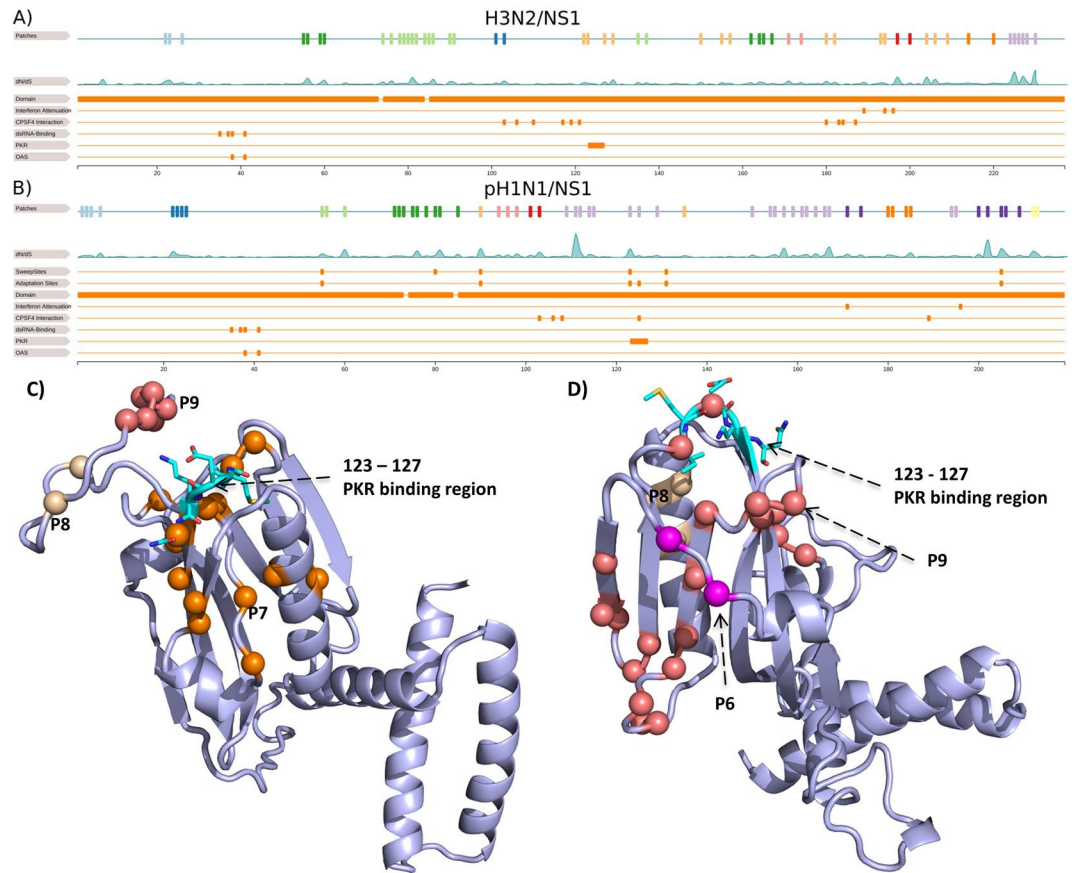


**Figure 2.** Structural representation of selected patches. We mapped selected patches on the corresponding protein models shown as cartoon in light blue with interesting residues shown as sticks and colored by atom-type (green and orange for carbon, red for oxygen and blue for nitrogen). Information about protein models is listed in Supplementary Table 1 and 2. The structures in (A,B) depict the head region of H3N2/HA1 and pH1N1/HA1, respectively, and show the RBS<sup>44,49</sup> and surrounding patches. In (C,D), we show views on the H3N2/NA and pH1N1/NA, respectively, and highlight patches that likely extend resistance sites (orange). In pH1N1/NP patch 2 and 6 are close to known mammalian adaptation sites 100<sup>16,70,71</sup> and site 319 (orange)<sup>72</sup>, respectively (E). The structure of pH1N1/PA is shown in (F) and highlights adaptation site 336 (orange) that increases polymerase activity in mammalian cells and neighboring patches 4, 5, 10 and 11.

pH1N1 (2009–2016). Specifically, for the polymerase subunits PB2 (17. vs.6), PB1 (7 vs.4) and PA (12 vs.7), pH1N1 viruses had more patches than H3N2, as well as for NP (7 vs.1) and NA (15 vs.8). There were fewer differences for NS1 (11 vs.9 patches) and M2 (3 vs.2), indicating a similar degree of positive selection acting on both proteins in both subtypes. We inferred two patches for pH1N1/M1 and no patches for H3N2/M1 and for NS2 (both subtypes). Only for the major surface antigen HA we detected fewer patches in pH1N1 viruses than in H3N2 viruses (12 vs.13 for HA1 and 0 vs. 1 for HA2), in agreement with pH1N1 experiencing less pressure to change antigenically and escape accumulating immunity in the host population over the past decade than H3N2 viruses (see below).

**Adaptive immune evolution.** We investigated which patches are located in regions that act as BCE or TCE and thus play a role in evading the host's immune response. In addition to extracellular regions of HA and NA, the M2 ion channel can be partly recognized as a BCE, too<sup>34</sup>. TCE epitopes exist in intracellular regions of HA, NA and M2<sup>34</sup>.

Of the H3N2/HA1 sites, 15% percent (52 sites) are located in a patch. Antigenic or BCE definitions for H3N2/HA1 were originally proposed by Wiley *et al.*<sup>40</sup> and Wiley and Skehel<sup>41</sup>. Subsequent studies refined key regions or sites responsible for antigenic drift in HA1<sup>14,23,42,43</sup>. Thus, we also compared H3N2 patch sites with “antigenic patch” sites<sup>23</sup>, “adapatch” sites<sup>22</sup>, key antigenic sites<sup>42</sup>, sweep-related sites<sup>16</sup> and TCE sites<sup>34</sup> (Fig. 1A). Eight of twenty-three antigenic patch sites are included in patches (34.7%). Of the seven key antigenic sites by Koel *et al.* (2013), sites 155, 156 and 189 were excluded from a patch. These findings underline that antigenic alterations cannot exclusively explain positive selection<sup>14,16</sup>. 65.8% of adapatch sites overlap with the patch sites of this study. In turn, thirty (57.7%) sites were newly clustered into patches. Half of the sweep-related sites from Klingen *et al.* match patch sites. While sweep-related sites are likely under positive selection and may become fixed in the circulating viral population, they do not necessarily have large  $dN/dS$  values, which would require multiple changes at a given site for detection. Of patches clustering on the globular head of HA1, patches 6, 8 and 9 are vicinal to



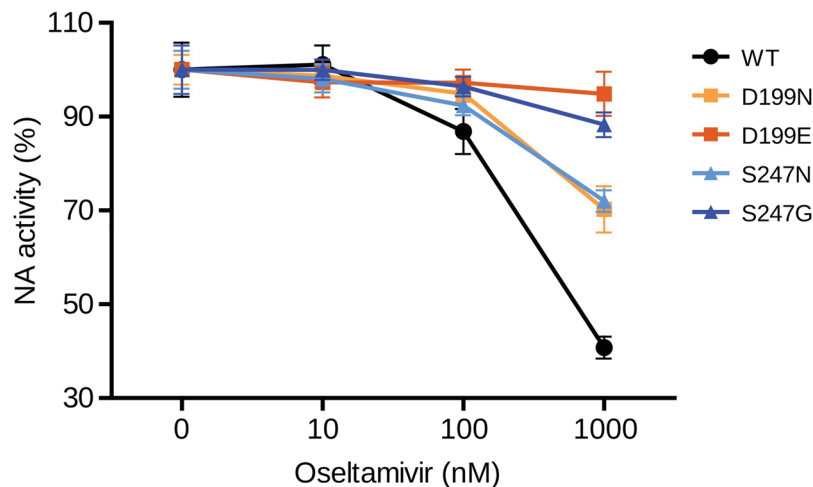
**Figure 3.** Structural and linear representation of NS1 patches. The NS1 models for H3N2 in (A) and pH1N1 in (B) are shown as cartoon in light blue with the PKR binding region<sup>51,61,62</sup> shown as sticks and colored by atom-type (cyan for carbon, red for oxygen and blue for nitrogen). We highlight patches in the surrounding region of the PKR binding region with a major focus on patch 7 in H3N2 and patch 9 in pH1N1. In (C,D), the first row is a linear representation of patches and the second row shows the  $dN/dS$  distribution of H3N2/NS1 and pH1N1/NS1. The third and fourth row in (D) show sweep-related sites<sup>16</sup> and adaptations sites<sup>57</sup> in pH1N1/NS1. Remaining rows in both subfigures depict the protein domains<sup>52,53</sup>, interferon attenuation sites<sup>52,59,60</sup>, CPSF30 interaction sites<sup>51,55,56</sup>, dsRNA binding sites<sup>7,63</sup>, PKR<sup>51,61,62</sup> and OAS binding sites<sup>51,63</sup>. Information about protein models is listed in Supplementary Tables 1 and 2.

receptor binding sites (RBS) 98, 153 and 195<sup>44</sup> (Fig. 2A). Altogether, the overlap of patch sites with BCE sites was ~70% and of patch sites with TCE sites was only ~20%.

Epitopes in HA1 for pH1N1 viruses were originally defined for seasonal H1N1 viruses by Caton *et al.*<sup>45</sup>. As the pH1N1 replaced the seasonal H1N1 virus, Matsuzaki *et al.*<sup>46</sup> redefined the former epitopes by mapping them onto the pH1N1/HA1 protein. We match patches in pH1N1/HA1 with “adapatch” sites<sup>22</sup>, key antigenic sites<sup>47</sup>, sweep-related sites<sup>16</sup>, TCE sites<sup>48</sup> and BCE sites<sup>46</sup> (Fig. 1C). In total, 16% percent of pH1N1/HA1 sites are located in patches. Fifteen of twenty-five “adapatch” sites overlap with novel patch sites (60%) and forty of fifty-five novel patch sites are no adapatch sites (~70%). 81% of sweep-related sites are also patch sites, which is a larger overlap compared to H3N2/HA1 data. Koel *et al.* (2015) described five key antigenic positions (127, 153, 155, 156 and 224) for pH1N1/HA1, of which none is located in a patch. However, most of them are in the vicinity of patch 4, which also surrounds RBS 150 and 192<sup>49</sup> (Fig. 2B). For pH1N1/HA1 we observe almost identical overlaps of patch sites with TCE sites (~20%) and BCE sites (~26%). The low percentage of TCE sites in patches for pH1N1/HA1 is similar to H3N2/HA1 results, but, the overlap with BCE sites is lower compared to H3N2 data.

Patch sites of H3N2/NA match 27% reported BCE sites<sup>50</sup>. While the overlap of patch sites with TCE for H3N2/NA is ~2%, there is no overlap for pH1N1/NA with known TCE sites, indicating that there was little immune evasion of TCR of NA (Fig. 1B,D). We could not investigate for H3N2/M2 the overlap of BCE and TCE sites with patch sites, because the former are located in regions not covered by our homology model (Supplementary Table 2).

**Attenuation of innate immune responses.** NS1 suppresses multiple antiviral host responses and mediates viral replication. Interferon synthesis is suppressed by binding the cleavage and polyadenylation specificity factor (CPSF30) and interfering with IFN- $\beta$  pre-mRNA synthesis<sup>51</sup>. 22.3% (Fifty-three sites) of protein sites in H3N2 viruses and 27.4% (sixty sites) of protein sites of pH1N1 viruses in NS1 are clustered in a patch, which

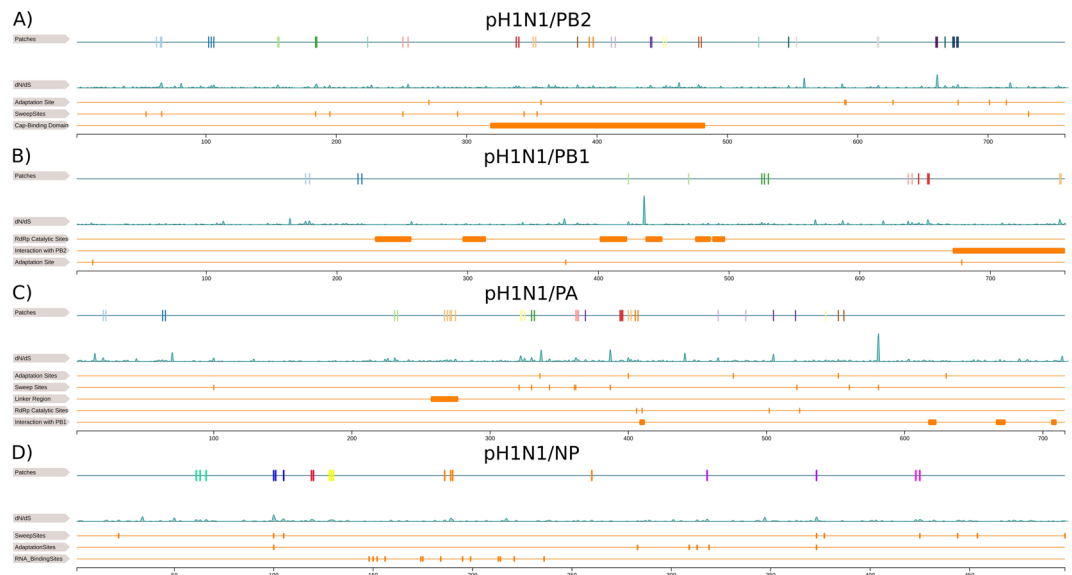


**Figure 4.** NA activity of pH1N1 viruses in the presence of oseltamivir. NA activity of wt and four NA mutant pH1N1 viruses (D199N, D199E, S247N and S247G) was determined by a fluorescence-based assay using the fluorogenic substrate 4-MU-NANA in the presence of indicated concentrations oseltamivir or mock treated. The mock-treated controls of each virus were set to 100%. Error bars are presented as standard deviation. Shown are the results of three independent experiments performed in triplicates.

is the largest number for all proteins (Supplementary Table 6). In both subtypes, most patches occur in the C-terminal effector domain (85-end of protein), while the N-terminal RNA-binding domain (1–73) is almost conserved across subtypes<sup>52</sup>. A flexible linker region of 12 conserved residues connects both distinct domains<sup>53</sup>. In both subtypes, a large patch spans the linker region but excludes sites 69 and 77 that are critical to maintain dimerization<sup>54</sup>. In H3N2 viruses, changes at sites 103 and 106<sup>55</sup> and in the binding pocket (110, 117, 119, 121, 180, 183, 184 and 187), which interacts with the second and third zinc finger domain of CPSF30<sup>56</sup>, link to post-transcriptional inhibition of antiviral IFN-stimulation<sup>51</sup>. Of these sites, patch 2 includes site 103 and patch 7 is next to sites 119, 121 and 183, and includes site 180 (Fig. 3A). In addition, the sites 108, 125 and 189 are relevant for CPSF30 binding in pH1N1 viruses<sup>55,56</sup>. In pH1N1, they overlap with patch 6 (site 103), patch 9 (site 125) and are adjacent to patch 9 (sites 106 and 108) (Fig. 3B). Clark *et al.* (2017) showed that substitutions at sites 55, 90, 123, 125, 131 and 205 inhibit host gene expression primarily by inhibiting CPSF30<sup>57</sup>, representing an adaptation of pH1N1 to humans<sup>16,57</sup>. Site 125 was also described as a host adaptation site in Selman *et al.*<sup>58</sup>. Of these sites, site 55 is part of patch 3, site 90 of patch 7, site 123 and 125 of patch 9 and 205 of patch 10 (Fig. 3B). NS1 also interferes with the expression of transcription factor IRF3 and NF- $\kappa$ B, which initiate IFN transcription<sup>59</sup>. The suppression of IFN responses is mediated via site 196 that blocks the IFN cascade when the amino acid E is present<sup>59</sup>. Site 196 is proximal to patch 6 in H3N2 and patch 9 in pH1N1. In H3N2 viruses, sites 189 and 194 affect interferon responses<sup>60</sup> and in pH1N1 viruses, changes in site 171 reduce host gene expression<sup>52</sup>. Site 194 is included in patch 7, while the pH1N1 specific site 171 lies in patch 10. A different viral strategy is directly binding of NS1 region 123–127 to protein kinase PKR, suppressing its activation and downstream antiviral responses<sup>51,61,62</sup>. This region overlaps with patch 9 in pH1N1/NS1 and patch 7 in H3N2/NS1 (Fig. 3C,D). Thus, patch 9 in pH1N1/NS1, which has a high average  $dN/dS$  value ( $\sim 1.96$ ), and patch 7 in H3N2/NS1, with a high average  $dN/dS$  value ( $\sim 1.2$ ), may be relevant for antagonizing IFN responses via CPSF30-binding and suppression of protein kinase PKR in both subtypes. Patches in the N-terminal RNA-binding domain do not overlap with functionally relevant sites such as dsRNA-binding sites (35, 37, 38 and 41)<sup>7,63</sup> and sites that inhibit IFN induced 2', 5'-Oligoadenylate Synthetase (OAS) (38 and 41), which both prevent NF- $\kappa$ B activation and IFN- $\beta$  induction<sup>51,63</sup>.

**Resistance evolution.** The ion channel of M2 is targeted by the antivirals amantadine and rimantadine, which are used to treat influenza infections<sup>21</sup>. We found a strong link between patch sites and antiviral resistance in the M2 protein. Known amantadine resistance sites 26, 27, 30, 31, 34 and 38, are surrounded by or part (site 27) of patch 1 and are also in the vicinity of patch 2 in pH1N1<sup>21</sup>. In H3N2, patch 1 includes amantadine resistance sites (26 and 27). Rimantadine resistance sites 40, 41, 42 and 44 are included or in the vicinity of patch 2 in pH1N1 and H3N2. Thus, there are similar trends for pH1N1 and H3N2 patches in M2, and all patch sites are either part of or next to known resistance sites, indicating their relevance for resistance development.

Currently, both drugs are not recommended anymore, because of increasing resistance in circulating viruses<sup>64</sup>. Alternatives are oseltamivir and zanamivir that inhibit NA sialidase, preventing viral detachment from the host cell<sup>21</sup>. In both subtypes, two patches are linked to known antiviral resistance sites, providing evidence for positive selection against these two drugs in NA for seasonal influenza viruses (2 out of 8 in H3N2 and 2 out of 15 in pH1N1): resistances against the anti-NA drugs are known to be linked to sites in NA 33, 119, 136, 142, 292, 320 and 222 in H3N2 viruses and 136, 156, 198, 213, 222, 246, 274 and 294 in pH1N1 viruses<sup>21</sup>. In H3N2/NA, resistance site 136 is in the vicinity of patch 2 and 142 is next to patch 8 (Fig. 2C). Resistance sites 198 and 246 are next to patch 4 and patch 5 in pH1N1/NA, respectively (Fig. 2D). In order to assess whether the predicted sites confer resistance against NA inhibitors, such as oseltamivir, we experimentally investigated positions 199 and site



**Figure 5.** Linear representation of pH1N1 RNP proteins. This figure illustrates the overlap of patch sites with known functional sites in pH1N1 RNP proteins PB2 (A) PB1 (B) PA (C) and NP (D). In each subfigure, the first row is a linear representation of patches and the second row shows the  $dN/dS$  distribution of the protein. For PB2 (A), we compare with adaptation sites<sup>38</sup>, sweep-related sites<sup>16</sup>, cap binding domain<sup>38</sup>. For PB1 (B), we compare with RdRp catalytic sites<sup>38</sup>, interactions sites with PB2<sup>38,77</sup> and adaptation sites<sup>78,79</sup>. For PA (C), we compare with adaptation sites<sup>38</sup>, sweep-related sites<sup>16</sup>, the linker region of PA<sup>38,76</sup>, RdRp catalytic sites<sup>38</sup> and interactions sites with PB1<sup>38,76</sup>. For NP (D), we compare with sweep-related sites<sup>16</sup>, adaptation sites<sup>16,70–72</sup> and RNA binding sites<sup>69</sup>.

247 of patch 4 and patch 5, respectively, for a role in generating oseltamivir resistances in NA activity assays. We show that amino acid substitutions in NA from D to N at position 199 and from S to N at position 247 considerably reduced the inhibitory effect of oseltamivir up to 30% neuraminidase activity upon treatment with 100 nm oseltamivir (Fig. 4). The ability of these mutations to confer oseltamivir resistance was further increased upon treatment with 1000 nm oseltamivir concentrations, further reducing NA activity to 5% and 12%, respectively.

In pH1N1/NA, patches 2 and 7, which also include the sweep-related site 117 and 248<sup>16</sup>, respectively, and patch 9 overlap or are vicinal to the active center of NA<sup>65–67</sup>. For H3N2/NA, conversely, this was the case for patch 2, which includes active site 151.

**Host adaptation of the RNP proteins.** Reassortant influenza viruses with segments of animal influenza viruses that newly establish in the human population, such as pH1N1, likely require further adaptive changes for fine-tuning the efficiency of replicating and spreading within the human host. NP and the polymerase proteins have important roles in host adaptation<sup>38,68</sup>. Notably, NP had entirely different patch sites in pH1N1 than in H3N2. In H3N2, NP is highly conserved with only one patch under positive selection (450, 451, 453), which does not appear in pH1N1. In pH1N1/NP, the RNA binding pocket surrounds patch 5, indicating a role in fine-tuning the binding of RNA molecules (Fig. 5D)<sup>69</sup>. We also found two patches close to known mammalian adaptation sites: site 100 in patch 2<sup>16,70,71</sup> and site 319 next to patch 6<sup>72</sup>. Site 319 is also in proximity to sites 373 and 100 implicated in selective sweeps (Fig. 5D). The region including patch 2 and patch 6 thus may well be implicated in human adaption in influenza A viruses (Fig. 2E), especially, because they have the largest average  $dN/dS$  value (~1.8) among all NP patches (Supplementary Table 7). Further, of three NP sites (100, 283, 313) relevant for escape recognition by intracellular restriction factor MxA, 313 is located in a beta-sheet adjacent to patch 6 and in the vicinity of patch 1<sup>73,74</sup>. Patches 3 and 4 are located in a region of NP with no known relevance for host adaptation.

Alterations in the polymerase complex proteins PB2, PB1 and PA can determine the host range and confer the ability to infect and replicate in humans. In PB2, site 627 changes when adapting to mammalian hosts and an E to K substitution at this site enhances polymerase activity in influenza A viruses in mammals, while site 701 enhances the pathogenicity and transmissibility of pH1N1 viruses<sup>38</sup>. In pH1N1 viruses, patch 17 includes the adaptation site 677 and patch 6 lies next to adaptation site 357 (Fig. 5A). Different from H3N2, for which we could not link patches to adaptation sites, likely because H3N2 is already adapted to the human host, pH1N1 viruses show signs for further host adaptation in PB2. When comparing patch sites with the nine sweep-related sites for pH1N1, three are located within a patch (66 in patch 1, 184 in patch 4 and 251 in patch 5) and one is localized in proximity to a patch (354 to patch 7) (Fig. 5A). Remarkably, there are seven patches (6 to 12) in pH1N1 viruses and two patches (3 and 4) in H3N2 viruses within the cap-binding domain (318–483)<sup>38</sup>, indicating that viral adaptation is related to the establishment and the enhancement of transcription activity of PB2.

The PA subunit includes known host adaptation sites 336, 400, 476, 552 and 630, of which we could link site 400 to patch 7, 552 to patch 12 and 336 to patch 4, 5, 10 and 11 in pH1N1 and 552 to patch 6 in H3N2<sup>38</sup>.



Interestingly, in pH1N1/PA, site 336 is the central residue of an adaptive cluster including sweep-related sites 321 (proximal to patch 11, average  $dN/dS$  value ( $\sim 1.7$ )), 330 (patch 4, average  $dN/dS$  value ( $\sim 1.4$ )), 361 (patch 5, average  $dN/dS$  value ( $\sim 1.2$ )) and 362 (proximal to patch 5) (Figs 2F and 3C)<sup>16,75</sup>. In addition, patch 10, which excludes a sweep-related site, is vicinal to site 336 and has an average  $dN/dS$  value of  $\sim 1.9$  (Fig. 2F). Other than for H3N2 viruses, for which we detected a single patch next to 336 (patch 2), an intensified signal of positive selection in this region in pH1N1 viruses emphasizes its relevance for further viral adaptation to the human population. For pH1N1, patch 11 suggest further adaptations of polymerase activity, as it is located next to sweep-related site 321, which is known to increase viral polymerase activity<sup>16,39</sup>. The linker region of PA (257–277) is crucial in the PA and PB1 interaction<sup>38,76</sup> and includes patch 7 in pH1N1 viruses, but none in H3N2 viruses. This patch has a very strong signal with a  $f_{p_7}$  value of 1.0 (Supplementary Table 7). As these segments have different evolutionary origins in the reassortant pH1N1 viruses (PB1 originating from a human H3N2; PA of avian origin, both intermediate in swine), this indicates a further fine-tuning of their interactions. We furthermore detected three patches (patch 7, patch 9 and patch 10) in pH1N1 viruses and one (patch 4) in H3N2 viruses in proximity to the RNA-dependent RNA polymerase activity sites (406, 410, 502 and 524)<sup>38</sup>.

The polymerase subunit PB1 has the fewest number of patches of the polymerase subunits. Patch 4 in H3N2 viruses and patch 7 in pH1N1 viruses overlap with the C-terminal region (671–757), which maintains tight inter-subunit contact with PB2 (1–35)<sup>38,77</sup>. Also for pH1N1, both segments have different evolutionary origins in the reassortant viruses, with PB2 originating more recently from an avian lineage circulating intermittently in swine, indicating a fine-tuning of their interactions<sup>38</sup>. Of three known host adaptation sites (13, 375 and 678), merely patch 2 in H3N2 includes site 375, while no patch of pH1N1 included these sites, which include typical mammalian adaptation residues<sup>78,79</sup>. As pH1N1/PB1 previously circulated in the human population for 30 years and re-entered the human population after circulating for 10 years in swine, the segment is likely already adapted to mammals and humans<sup>5</sup>. The centrally located RNA-dependent RNA polymerase activity motifs of PB1 are highly conserved<sup>38</sup>, with patch 3 in H3N2 viruses and in pH1N1 viruses located in proximity (Fig. 5B).

## Discussion

It is well established that the surface antigens of human influenza A viruses evolve under positive selection to escape immune recognition, in an ongoing co-evolutionary arms-race with the human adaptive immune response. However, much less is known about the relevance of other proteins in this co-evolutionary arms-race and which other factors shape the evolutionary trajectories of these viruses. There is emerging evidence that the pH1N1 virus since 2009 also has acquired changes in improving its adaptation to the human host<sup>15,39,71</sup>. The recent introduction of the pH1N1 virus into the human population allows studying the evolutionary forces acting on these viruses in comparison to the H3N2 viruses, which have circulated for five decades. To analyze these effects, we searched for protein regions under positive selection from all available historic sequences and protein structures for both influenza A subtypes with a refined computational method, including both structural proximity and the strength of site-wise positive selection into the inference process. While there are studies focusing on selective measurements per protein sites<sup>34–36</sup>, to our knowledge, this study is the first to describe protein regions relevant for adaptation of influenza viruses for almost all viral proteins and a comparative analysis of their functional importance. Although reassortment events are prevalent in influenza viruses and shape their genome-wide reticulate evolution<sup>35,80</sup>, this process is not relevant for the analysis, as we inferred phylogenies and underlying selection for each gene locus separately.

Contrary to expectations, the HA1 subunit of HA does not have the largest number of patches or sites under positive selection and was ranked behind NA, M2, NS1 and NS2. Most BCE sites overlap with patch sites in H3N2 viruses ( $\sim 70\%$ ), while the patches in pH1N1/HA1 marginally overlap with BCE sites. This indicates, as expected, that antigenic evolution over the past decade was not pronounced in pH1N1 viruses, in agreement with a single vaccine strain update for pH1N1 viruses from A/California/7/2009 to A/Michigan/45/2015 being recommended by the WHO for the use in the 2017 southern hemisphere influenza season<sup>81</sup>. However, high rates of selective sweeps indicate that positive selection has a strong effect on both viruses on the major surface proteins. When comparing to the “adapatch” sites we identified in Tusche *et al.* (2012), the  $F_1$ -score for detecting epitope sites for H3N2 with our improved method was identical. The improved approach trades some precision against sensitivity, with an increase in overall detected patch sites, patches and frequency of sites  $dN/dS > 1$  (76% vs. 35%) in identified patches. Another difference was that for subtype pH1N1, the data covered a substantially longer time period (2009–2017 vs. 2009–2011) and we used a protein model generated from a pH1N1 strain (Supplementary Table 1).

Notably, NS1 is covered most densely with patches of all proteins for both subtypes, suggesting that viral attenuation of innate immune responses is an important, continuous process in the evolution of human influenza viruses, even decades after the establishment of the H3N2 virus in the human host. This is in line with classic evolutionary theory, postulating that in co-evolutionary arms-races, both players evolve towards a mitigated state over time causing less severe infections<sup>82,83</sup>. However, recent results have challenged this paradigm, showing the alarming rate of causalities of rabbits infected with myxoma viruses<sup>84</sup>, raising the possibility that the current evolutionary trajectory of down tuning host innate immune defenses for human influenza viruses may also eventually result in an escalation of viral virulence uncontrollable by host immune defenses<sup>84</sup>. For pH1N1, patches 3, 7, 9 and 10 include known changes linking to substantial attenuation of innate immune responses, by restoring NS1-mediated general gene expression inhibition and resulting in less severe inflammatory response after influenza infections<sup>57</sup>. Patches 2 and 7 in H3N2 and patches 6 and 9 in pH1N1 link to CPSF30-binding that initiates inhibition of antiviral IFN-stimulation<sup>55,56</sup>. Further, IFN regulation is altered by amino acid changes at site 196 and 171 in pH1N1<sup>52,59</sup>, as well as 196 and 189 in H3N2<sup>59,60</sup>, which overlap with patch 6 and 7 in H3N2, as well as patch 9 and 10 in pH1N1. Other sites in patch 9 in pH1N1/NS1 and patch 7 in H3N2/NS1 link to binding host

PKR, a critical component of IFN-stimulated defenses<sup>51</sup>. In conclusion, particularly patch 9 in pH1N1/NS1 and patch 7 in H3N2/NS1 may be important for antagonizing IFN responses via CPSF30-binding and other mechanisms in both subtypes.

Overall, we found significantly higher mean  $dN/dS$  values in most pH1N1 proteins in comparison to H3N2. This was despite the longer time period for which sequence data was available for H3N2 viruses (1968–2016) than for pH1N1 viruses (2009–2016), further suggesting that pH1N1 viruses are not as well adapted to the human population yet as H3N2 viruses are. This is especially in agreement with elevated  $dN/dS$  values in the pandemic vs. post-pandemic phase for pH1N1 viruses. Accordingly, the number of patches found in PB2, PB1, PA and NP found for pH1N1 viruses exceeds those for H3N2 viruses, and their locations correspond to evolutionary stable areas in H3N2 proteins. For instance, while in H3N2 viruses, NP is largely conserved with only a single patch; in pH1N1 seven patches were found. Patch 2 and 6 are particularly interesting for host adaptation, as they are near known mammalian adaptation sites 100 and 319, respectively<sup>16,70–72</sup>. For PA, the neighboring region around mammalian adaptation site 336 that confers selective advantage to promote polymerase activity in human<sup>75</sup> overlaps with patches 4, 5, 10 and 11 in pH1N1 viruses, indicating their relevance for human host adaptation. Also for PB2 of pH1N1, patches linking to mammalian adaptation sites (patch 17 and 6) and located in proximity to the cap-binding domain were found, different from H3N2, indicating that recent adaptation of pH1N1 is related to enhanced transcription activity<sup>38</sup>. PB1 of pH1N1 is the protein descending from a recent human lineage, and fewer patches than in the other polymerase subunits were found. Specifically, one patch in the C-terminal region was identified, likely to adjust the interaction with the novel version of PB2 in the reassortant lineage<sup>38</sup>, as PB2 descends from an avian lineage circulating in swine before.

For M2 in both subtypes, we found patches (1 and 2) that include known resistance sites against amantadine and rimantadine, respectively, and multiple currently undescribed sites likely linking to resistances. Similarly, oseltamivir and zanamivir resistances in NA could be linked to patch 2 and 8 in H3N2 and patch 4 and 5 in pH1N1, respectively. In NA activity assays, we could show the relevance of substitutions in patch site 199 and patch site 247 in pH1N1 in conferring resistance to oseltamivir. Patches in NA without a clear link to resistances or immune responses could be relevant for host adaptation, i.e. for establishing its catalytic function in the human host, to establish cleavage activity on an altered substrate range, as pH1N1 viruses had many more patches than H3N2 viruses.

The collection of patches and included sites provide an atlas of genetic elements linked to multiple factors predominantly influencing the evolutionary trajectories of human influenza A viruses. We demonstrate the value by confirming the resistance-conferring phenotype of changes at two such sites. The collection provides a rich resource of candidate sites and markers for viral host adaptation, immune evasion, resistance generation and attenuation of host immune responses, which could lead to an improved understanding of the underlying biological processes and more effective monitoring of phenotypes of circulating viral strains. The software and the patch collection are provided at <https://github.com/hzi-bifo/PatchDetection><sup>85</sup>.

## Methods

**Structure modeling.** We analyzed the polymerase proteins PB2, PB1 and PA, the HA subunits 1 and 2, NP, NA and the splice-variants M1 and M2, as well as NS1 and NS2 for influenza subtypes H3N2 and pH1N1, respectively. Protein structures for HA for H3N2 viruses and NA for pH1N1 viruses were extracted from the RSCB database<sup>86</sup>. We modeled missing or incomplete protein structures with the homology modeling tool MODELLER using the A/Aichi/2/1968(H3N2) strain for H3N2 and the A/California/04/2009(H1N1) strain for pH1N1 as target sequences (Supplementary Tables 1 and 2)<sup>87</sup>. The sequence identity between the target sequence and the sequence of the homology models ranges from 67% (PB1) to 99% (HA1) (Supplementary Tables 1 and 2). We were able to generate complete protein models for all polymerase subunits (both subtypes), HA and NS1 (both pH1N1) and partial protein models for the remaining proteins (Supplementary Tables 1 and 2).

**Phylogenetic analysis.** For each protein, we downloaded the amino acid sequences and the corresponding coding sequences from the NCBI flu database<sup>88</sup>. We removed identical coding sequences, which would accumulate on the same leaf in a phylogenetic tree because they contribute no additional information and removing significantly speeds up the calculation of alignments. No additional filtering was applied to sequences. The following steps were applied for each protein, individually:

We generated a multiple sequence alignment for amino acid sequences using muscle<sup>89</sup>. Subsequent removal of positions with more than 80% gaps using trimAL ensured a consistent numbering<sup>90</sup>. To speed up the analysis, we applied pal2nal to the protein alignment and the coding sequences to generate a multiple codon alignment<sup>91</sup>. Based on the multiple codon alignment, we inferred a phylogeny using FastTree with the GTR-model, which allows fast and precise inferences<sup>92,93</sup>. For each codon in the coding sequence and its corresponding amino acid, we reconstructed ancestral character states assigned to internal nodes using the Fitch algorithm with accelerated transformation (ACCTRAN) such that each internal node contains an intermediate sequence state of codon sequence<sup>94</sup>.

To measure positive selection, we calculate  $dN/dS$  ratios for every amino acid position considering each codon in the phylogenetic tree, separately<sup>31,95</sup>. To calculate the  $dN/dS$  statistics, we apply the tool described in Munch *et al.* (2017). Codon sequence changes are mapped onto branches in the phylogeny and are classified as synonymous or non-synonymous mutations. A synonymous mutation does not change the amino acid while a non-synonymous mutation does<sup>13</sup>. Starting from the root of the tree, we traverse down to the leaves and count position-wise every change that occurs between two nodes on the branches. For each site, we divide the amount of non-synonymous changes by the amount of synonymous changes normalized by the amount of comparisons to get the  $dN/dS$  ratio. To reduce the influence of sequencing artifacts, we exclude terminal branches to count changes supported by at least two viral strains but did not control for lab adaptation effects described in McWhite *et al.*<sup>96</sup>.

**PatchDetection.** The patch detection method is based on the methods from Tusche *et al.* (2012) and Kratsch *et al.* (2016). It includes a graph cut algorithm combining spatial information of structure models with  $dN/dS$  measurements. Following the original versions, we create a graph with nodes  $n$  representing protein sites that are connected to all neighboring residues within a radius of  $\delta$ . The edges are weighted by a distance function  $e^{-dist(m,n)}$  where  $dist(m,n)$  calculates the euclidean distance between the  $C_\alpha$ -atoms of a residue pair  $n$  and  $m$ . Nodes with a close spatial proximity therefore have a strong connection. We add two additional nodes to the graph, the positive selection node ( $pos$ ) and the negative selection node ( $neg$ ). Each  $n$  is connected to  $pos$  with  $P(n)$  and to  $neg$  with  $\bar{P}(n)$ . Sites are separated into a positive or a negative selection set by a minimum graph cut approach. The cut divides the graph into two halves by minimizing the sum  $C$  of weights from edges connecting these halves with a cost function:

$$C = \left( \sum_{n \in pos} \sum_{\substack{m \in neg \\ m \in neg(n)}} e^{-dist(m,n)} \right) + \beta \left( \sum_{n \in pos} \bar{P}(n) + \sum_{n \in neg} P(n) \right) \quad (1)$$

In a final step, sites in the positive selection set are merged into patches together with neighboring sites within a distance of  $\delta$ .

In the new formulation, we adjusted the graph-cut function to directly include  $dN/dS$  values. We set  $P(n) = dN/dS$  and  $\bar{P}(n) = (1 - P(n))$  to weigh the positive selection node ( $pos$ ) and the negative selection node ( $neg$ ), balancing positively-selected sites against sites not under positive selection. We applied a radius  $\delta$  in the stable interval between 7 and 8.5 Å; (see approach Supplementary Table 2 in Kratsch *et al.* (2016)). We include buried and exposed sites to account for macromolecular changes in the core of the protein. To determine  $\beta$ , we start at 1, incrementally increase  $\beta$  and calculate patches at each step until patches do not change anymore within 100 steps. Having reached saturation at  $\beta^{max}$ , we set  $\beta = \beta^{max} * 0.5$  to balance between both extreme cases, in which either the distance function dominates at  $\beta = 0$  or the  $dN/dS$  term dominates at  $\beta^{max}$ .

We evaluate the robustness of each patch by subsampling protein sites. This provides statistical support to estimate the stability of each patch regarding its site content. We define a set of patches for a protein as  $P = \{p_1, p_2, \dots, p_n\}$ , where  $n$  is the total number of patches and  $p_i$  a set of sites under positive selection in the  $i$ -th patch. We generate  $N = 1,000$  samples by randomly removing 10% of sites from the protein structure and define  $P_j = \{p_{j,1}, p_{j,2}, \dots, p_{j,k}\}$  as the  $k$  patches that we re-calculate on the  $j$ -th sample. For each patch  $p_i \in P$ , we perform the following steps: For each sample  $j$ , we define the patch  $p_{max} \in P_j$ , which has the maximum number of overlapping sites with  $p_i$ , and calculate the number of true positives as  $TP_{p_i} = |p_i \cap p_{max}|$ , false negatives as  $FN_{p_i} = |p_i \setminus p_{max}|$  and false positives as  $FP_{p_i} = |p_{max} \setminus p_i|$ . Sites that were removed by the sampling process are omitted for calculation of  $TP_{p_i}$ ,  $FN_{p_i}$  and  $FP_{p_i}$ . We use the standard definitions for recall, precision and F-score<sup>97</sup>. The criterion  $f_{p_i} = \frac{|p_{j,k} \mid |p_i \cap p_{j,k}| \geq 2|}{N}$  reflects how stable a patch appears in the samples. When the value approaches zero, the patch tends to disappear, a value of one indicates that the patch is stable and a value larger than one indicates an unstable patch that breaks apart in the subsamples. In addition, we provide the average  $dN/dS$  value per patch.

**NA activity assays.** The NA activity of pH1N1 influenza A virus (A/Hamburg/NY1580/2009) was measured by a fluorescence-based assay using the fluorogenic substrate 2'-(4-Methylumbelliferyl)- $\alpha$ -D-N-acetylneuraminic acid (4-MU-NANA, Sigma-Aldrich) which is cleaved by the viral NA enzyme to release the fluorescent product 4-methylumbelliferone (4-MU)<sup>98</sup>. Therefore, recombinant wildtype (wt) or NA mutant (D199N, D199E, S247N, S247G) pH1N1 viruses were diluted to an equivalent NA activity and pre-incubated with 10, 100 or 1000 nM Oseltamivir (Tamiflu, Roche) or mock treated with 1xPBS for 15 min at 37 °C. The NA enzyme activity was initiated by adding 40  $\mu$ M 4-MU-NANA in calcium-TBS (6.8 mM CaCl<sub>2</sub>, 0.85% NaCl, 0.02 M Tris; pH 7.3). After 30 min at 37 °C, the reaction was stopped by the addition of 100  $\mu$ l of 0.1 M glycine buffer (pH 10.7) containing 25% ethanol. The fluorescence of released 4-MU was determined with a Safire 2 multi-plate reader (Tecan) using excitation and emission wavelengths of 355 nm and 460 nm, respectively. The mean fluorescence signal of the substrate without virus was subtracted as background from the signals obtained in the other wells. The specific NA activity was expressed in percentage.

## Data Availability

The PatchDetection software, figures and tables from this manuscript, and all related data used in this publication are fully available under the Apache License 2.0 at <https://github.com/hzi-bifo/PatchDetection>.

## References

- WHO. *Fact sheet number 211*, (2018).
- Petrova, V. N. & Russell, C. A. The evolution of seasonal influenza viruses. *Nat Rev Microbiol* **16**, 47–60, <https://doi.org/10.1038/nrmicro.2017.118> (2018).
- Medina, R. A. & Garcia-Sastre, A. Influenza A viruses: new research developments. *Nat Rev Microbiol* **9**, 590–603, <https://doi.org/10.1038/nrmicro2613> (2011).
- Bouvier, N. M. & Palese, P. The biology of influenza viruses. *Vaccine* **26**(Suppl 4), D49–53 (2008).
- Garten, R. J. *et al.* Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* **325**, 197–201, <https://doi.org/10.1126/science.1176225> (2009).
- Scholtissek, C., Rohde, W., Von Hoyningen, V. & Rott, R. On the origin of the human influenza virus subtypes H2N2 and H3N2. *Virology* **87**, 13–20 (1978).
- Das, K., Aramini, J. M., Ma, L. C., Krug, R. M. & Arnold, E. Structures of influenza A proteins and insights into antiviral drug targets. *Nat Struct Mol Biol* **17**, 530–538, <https://doi.org/10.1038/nsmb.1779> (2010).

8. Jagger, B. W. *et al.* An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* **337**, 199–204, <https://doi.org/10.1126/science.1222213> (2012).
9. Wise, H. M. *et al.* Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. *PLoS Pathog* **8**, e1002998, <https://doi.org/10.1371/journal.ppat.1002998> (2012).
10. Wise, H. M. *et al.* A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. *J Virol* **83**, 8021–8031, <https://doi.org/10.1128/JVI.00826-09> (2009).
11. Lindstrom, S. E., Cox, N. J. & Klimov, A. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957–1972: evidence for genetic divergence and multiple reassortment events. *Virology* **328**, 101–119, <https://doi.org/10.1016/j.virol.2004.06.009> (2004).
12. Fitch, W. M., Bush, R. M., Bender, C. A. & Cox, N. J. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc Natl Acad Sci USA* **94**, 7712–7718 (1997).
13. Bush, R. M., Fitch, W. M., Bender, C. A. & Cox, N. J. Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Mol Biol Evol* **16**, 1457–1465 (1999).
14. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376, <https://doi.org/10.1126/science.1097211> (2004).
15. Steinbruck, L. & McHardy, A. C. Allele dynamics plots for the study of evolutionary dynamics in viral populations. *Nucleic Acids Res* **39**, e4, <https://doi.org/10.1093/nar/gkq909> (2011).
16. Klingen, T. R. *et al.* Sweep Dynamics (SD) plots: Computational identification of selective sweeps to monitor the adaptation of influenza A viruses. *Sci Rep* **8**, 373, <https://doi.org/10.1038/s41598-017-18791-z> (2018).
17. Klingen, T. R., Reimering, S., Guzman, C. A. & McHardy, A. C. In Silico Vaccine Strain Prediction for Human Influenza Viruses. *Trends in microbiology* **26**, 119–131, <https://doi.org/10.1016/j.tim.2017.09.001> (2018).
18. Morris, D. H. *et al.* Predictive Modeling of Influenza Shows the Promise of Applied Evolutionary Biology. *Trends in microbiology* **26**, 102–118, <https://doi.org/10.1016/j.tim.2017.09.004> (2018).
19. Iwasaki, A. & Pillai, P. S. Innate immunity to influenza virus infection. *Nature reviews. Immunology* **14**, 315–328, <https://doi.org/10.1038/nri3665> (2014).
20. Brubaker, S. W., Bonham, K. S., Zanoni, I. & Kagan, J. C. Innate immune pattern recognition: a cell biological perspective. *Annual review of immunology* **33**, 257–290, <https://doi.org/10.1146/annurev-immunol-032414-112240> (2015).
21. Hussain, M., Galvin, H. D., Haw, T. Y., Nutsford, A. N. & Husain, M. Drug resistance in influenza A virus: the epidemiology and management. *Infection and drug resistance* **10**, 121–134, <https://doi.org/10.2147/idr.s105473> (2017).
22. Tusche, C., Steinbruck, L. & McHardy, A. C. Detecting patches of protein sites of influenza A viruses under positive selection. *Mol Biol Evol* **29**, 2063–2071, <https://doi.org/10.1093/molbev/mss095> (2012).
23. Kratsch, C., Klingen, T. R., Mumken, L., Steinbruck, L. & McHardy, A. C. Determination of antigenicity-altering patches on the major surface protein of human influenza A/H3N2 viruses. *Virus Evol* **2**, vev025, <https://doi.org/10.1093/ve/vev025> (2016).
24. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends in ecology & evolution* **15**, 496–503 (2000).
25. Bhatt, S., Holmes, E. C. & Pybus, O. G. The genomic rate of molecular adaptation of the human influenza A virus. *Mol Biol Evol* **28**, 2443–2451, <https://doi.org/10.1093/molbev/msr044> (2011).
26. Neher, R. A. & Bedford, T. nextflu: real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics* **31**, 3546–3548, <https://doi.org/10.1093/bioinformatics/btv381> (2015).
27. Suzuki, Y. New methods for detecting positive selection at single amino acid sites. *J Mol Evol* **59**, 11–19, <https://doi.org/10.1007/s00239-004-2599-6> (2004).
28. Meyer, A. G. & Wilke, C. O. Geometric Constraints Dominate the Antigenic Evolution of Influenza H3N2 Hemagglutinin. *PLoS Pathog* **11**, e1004940, <https://doi.org/10.1371/journal.ppat.1004940> (2015).
29. Meyer, A. G. & Wilke, C. O. Integrating sequence variation and protein structure to identify sites under selection. *Mol Biol Evol* **30**, 36–44, <https://doi.org/10.1093/molbev/mss217> (2013).
30. Stray, S. J. & Pittman, L. B. Subtype- and antigenic site-specific differences in biophysical influences on evolution of influenza virus hemagglutinin. *Viral J* **9**, 91, <https://doi.org/10.1186/1743-422x-9-91> (2012).
31. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418–426 (1986).
32. Kosakovsky Pond, S. L. & Frost, S. D. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208–1222, <https://doi.org/10.1093/molbev/msi105> (2005).
33. Kryazhimskiy, S. & Plotkin, J. B. The population genetics of dN/dS. *PLoS genetics* **4**, e1000304, <https://doi.org/10.1371/journal.pgen.1000304> (2008).
34. Suzuki, Y. Natural selection on the influenza virus genome. *Mol Biol Evol* **23**, 1902–1911, <https://doi.org/10.1093/molbev/msl050> (2006).
35. Westgeest, K. B. *et al.* Genomewide analysis of reassortment and evolution of human influenza A(H3N2) viruses circulating between 1968 and 2011. *J Virol* **88**, 2844–2857, <https://doi.org/10.1128/jvi.02163-13> (2014).
36. Su, Y. C. *et al.* Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. *Nature communications* **6**, 7952, <https://doi.org/10.1038/ncomms8952> (2015).
37. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679, <https://doi.org/10.1093/bioinformatics/bti079> (2005).
38. Stubbs, T. M. & Te Velthuis, A. J. The RNA-dependent RNA polymerase of the influenza A virus. *Future virology* **9**, 863–876, <https://doi.org/10.2217/fvl.14.66> (2014).
39. Elderfield, R. A. *et al.* Accumulation of human-adapting mutations during circulation of A(H1N1)pdm09 influenza virus in humans in the United Kingdom. *J Virol* **88**, 13269–13283, <https://doi.org/10.1128/jvi.01636-14> (2014).
40. Wiley, D. C., Wilson, I. A. & Skehel, J. J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* **289**, 373–378 (1981).
41. Wiley, D. C. & Skehel, J. J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annual review of biochemistry* **56**, 365–394, <https://doi.org/10.1146/annurev.bi.56.070187.002053> (1987).
42. Koel, B. F. *et al.* Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* **342**, 976–979, <https://doi.org/10.1126/science.1244730> (2013).
43. Steinbruck, L. & McHardy, A. C. Inference of genotype-phenotype relationships in the antigenic evolution of human influenza A (H3N2) viruses. *PLoS Comput Biol* **8**, e1002492, <https://doi.org/10.1371/journal.pcbi.1002492> (2012).
44. de Graaf, M. & Fouchier, R. A. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *Embo j* **33**, 823–841, <https://doi.org/10.1002/embj.201387442> (2014).
45. 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* **31**, 417–427 (1982).
46. Matsuzaki, Y. *et al.* Epitope mapping of the hemagglutinin molecule of A/(H1N1)pdm09 influenza virus by using monoclonal antibody escape mutants. *J Virol* **88**, 12364–12373, <https://doi.org/10.1128/jvi.01381-14> (2014).
47. Koel, B. F. *et al.* Identification of amino acid substitutions supporting antigenic change of influenza A(H1N1)pdm09 viruses. *J Virol* **89**, 3763–3775, <https://doi.org/10.1128/jvi.02962-14> (2015).
48. Yang, J. *et al.* CD4<sup>+</sup> T cells recognize unique and conserved 2009 H1N1 influenza hemagglutinin epitopes after natural infection and vaccination. *Int Immunol* **25**, 447–457, <https://doi.org/10.1093/intimm/dxt005> (2013).

49. Yang, H., Carney, P. & Stevens, J. Structure and Receptor binding properties of a pandemic H1N1 virus hemagglutinin. *PLoS currents* **2**, Rrn1152, <https://doi.org/10.1371/currents.RRN1152> (2010).
50. Air, G. M., Els, M. C., Brown, L. E., Laver, W. G. & Webster, R. G. Location of antigenic sites on the three-dimensional structure of the influenza N2 virus neuraminidase. *Virology* **145**, 237–248 (1985).
51. Hale, B. G., Randall, R. E., Ortin, J. & Jackson, D. The multifunctional NS1 protein of influenza A viruses. *J Gen Virol* **89**, 2359–2376, <https://doi.org/10.1099/vir.0.2008/004606-0> (2008).
52. Platt, E. P., Ilyushina, N. A., Sheikh, F., Donnelly, R. P. & Ye, Z. Influenza virus NS1 protein mutations at position 171 impact innate interferon responses by respiratory epithelial cells. *Virus Res* **240**, 81–86, <https://doi.org/10.1016/j.virusres.2017.07.021> (2017).
53. Carrillo, B. *et al.* The influenza A virus protein NS1 displays structural polymorphism. *J Virol* **88**, 4113–4122, <https://doi.org/10.1128/jvi.03692-13> (2014).
54. Marc, D. Influenza virus non-structural protein NS1: interferon antagonism and beyond. *J Gen Virol* **95**, 2594–2611, <https://doi.org/10.1099/vir.0.069542-0> (2014).
55. Wang, B. X., Brown, E. G. & Fish, E. N. Residues F103 and M106 within the influenza A virus NS1 CPSF4-binding region regulate interferon-stimulated gene translation initiation. *Virology* **508**, 170–179, <https://doi.org/10.1016/j.virol.2017.05.009> (2017).
56. Das, K. *et al.* Structural basis for suppression of a host antiviral response by influenza A virus. *Proc Natl Acad Sci USA* **105**, 13093–13098, <https://doi.org/10.1073/pnas.0805213105> (2008).
57. Clark, A. M., Nogales, A., Martinez-Sobrido, L., Topham, D. J. & DeDiego, M. L. Functional Evolution of Influenza Virus NS1 Protein in Currently Circulating Human 2009 Pandemic H1N1 Viruses. *J Virol* **91**, <https://doi.org/10.1128/jvi.00721-17> (2017).
58. Selman, M., Dankar, S. K., Forbes, N. E., Jia, J. J. & Brown, E. G. Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerg Microbes Infect* **1**, e42, <https://doi.org/10.1038/emi.2012.38> (2012).
59. Kuo, R. L., Zhao, C., Malur, M. & Krug, R. M. Influenza A virus strains that circulate in humans differ in the ability of their NS1 proteins to block the activation of IRF3 and interferon-beta transcription. *Virology* **408**, 146–158, <https://doi.org/10.1016/j.virol.2010.09.012> (2010).
60. Nogales, A., Martinez-Sobrido, L., Topham, D. J. & DeDiego, M. L. NS1 Protein Amino Acid Changes D189N and V194I Affect Interferon Responses, Thermosensitivity, and Virulence of Circulating H3N2 Human Influenza A Viruses. *J Virol* **91**, <https://doi.org/10.1128/jvi.01930-16> (2017).
61. Min, J. Y., Li, S., Sen, G. C. & Krug, R. M. A site on the influenza A virus NS1 protein mediates both inhibition of PKR activation and temporal regulation of viral RNA synthesis. *Virology* **363**, 236–243, <https://doi.org/10.1016/j.virol.2007.01.038> (2007).
62. Li, S., Min, J. Y., Krug, R. M. & Sen, G. C. Binding of the influenza A virus NS1 protein to PKR mediates the inhibition of its activation by either PACT or double-stranded RNA. *Virology* **349**, 13–21, <https://doi.org/10.1016/j.virol.2006.01.005> (2006).
63. Wang, W. *et al.* RNA binding by the novel helical domain of the influenza virus NS1 protein requires its dimer structure and a small number of specific basic amino acids. *RNA (New York, N.Y.)* **5**, 195–205 (1999).
64. Deyde, V. M. *et al.* Surveillance of resistance to adamantanes among influenza A(H3N2) and A(H1N1) viruses isolated worldwide. *The Journal of infectious diseases* **196**, 249–257, <https://doi.org/10.1086/518936> (2007).
65. Colman, P. M., Hoyne, P. A. & Lawrence, M. C. Sequence and structure alignment of paramyxovirus hemagglutinin-neuraminidase with influenza virus neuraminidase. *J Virol* **67**, 2972–2980 (1993).
66. Maurer-Stroh, S., Ma, J., Lee, R. T., Sirota, F. L. & Eisenhaber, F. Mapping the sequence mutations of the 2009 H1N1 influenza A virus neuraminidase relative to drug and antibody binding sites. *Biology direct* **4**, 18; discussion 18, <https://doi.org/10.1186/1745-6150-4-18> (2009).
67. Abed, Y., Baz, M. & Boivin, G. Impact of neuraminidase mutations conferring influenza resistance to neuraminidase inhibitors in the N1 and N2 genetic backgrounds. *Antiviral therapy* **11**, 971–976 (2006).
68. Fodor, E. The RNA polymerase of influenza a virus: mechanisms of viral transcription and replication. *Acta Virol* **57**, 113–122 (2013).
69. Liu, C. L. *et al.* Using mutagenesis to explore conserved residues in the RNA-binding groove of influenza A virus nucleoprotein for antiviral drug development. *Sci Rep* **6**, 21662, <https://doi.org/10.1038/srep21662> (2016).
70. Otte, A. *et al.* Evolution of 2009 H1N1 influenza viruses during the pandemic correlates with increased viral pathogenicity and transmissibility in the ferret model. *Sci Rep* **6**, 28583, <https://doi.org/10.1038/srep28583> (2016).
71. Otte, A. *et al.* Adaptive Mutations That Occurred during Circulation of H1N1 Influenza Virus in the 2009 Pandemic Enhance Virulence in Mice. *J Virol* **89**, 7329–7337, <https://doi.org/10.1128/jvi.00665-15> (2015).
72. Gabriel, G., Herwig, A. & Klenk, H. D. Interaction of polymerase subunit PB2 and NP with importin alpha1 is a determinant of host range of influenza A virus. *PLoS Pathog* **4**, e11, <https://doi.org/10.1371/journal.ppat.0040011> (2008).
73. Deeg, C. M. *et al.* In vivo evasion of MxA by avian influenza viruses requires human signature in the viral nucleoprotein. *J Exp Med* **214**, 1239–1248, <https://doi.org/10.1084/jem.20161033> (2017).
74. Gotz, V. *et al.* Corrigendum: Influenza A viruses escape from MxA restriction at the expense of efficient nuclear vRNP import. *Sci Rep* **6**, 25428, <https://doi.org/10.1038/srep25428> (2016).
75. Bussey, K. A. *et al.* PA residues in the 2009 H1N1 pandemic influenza virus enhance avian influenza virus polymerase activity in mammalian cells. *J Virol* **85**, 7020–7028, <https://doi.org/10.1128/jvi.00522-11> (2011).
76. Guu, T. S., Dong, L., Wittung-Stafshede, P. & Tao, Y. J. Mapping the domain structure of the influenza A virus polymerase acidic protein (PA) and its interaction with the basic protein 1 (PB1) subunit. *Virology* **379**, 135–142, <https://doi.org/10.1016/j.virol.2008.06.022> (2008).
77. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355–360, <https://doi.org/10.1038/nature14008> (2014).
78. Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893, <https://doi.org/10.1038/nature04230> (2005).
79. Gabriel, G. *et al.* Differential polymerase activity in avian and mammalian cells determines host range of influenza virus. *J Virol* **81**, 9601–9604, <https://doi.org/10.1128/jvi.00666-07> (2007).
80. Holmes, E. C. *et al.* Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS biology* **3**, e300, <https://doi.org/10.1371/journal.pbio.0030300> (2005).
81. WHO. Recommended composition of influenza virus vaccines for use in the 2017 southern hemisphere influenza season. *Releve epidemiologique hebdomadaire* **91**, (469–484) (2016).
82. Burgess, H. M. & Mohr, I. Evolutionary clash between myxoma virus and rabbit PKR in Australia. *Proc Natl Acad Sci USA* **113**, 3912–3914, <https://doi.org/10.1073/pnas.1602063113> (2016).
83. Peng, C., Haller, S. L., Rahman, M. M., McFadden, G. & Rothenburg, S. Myxoma virus M156 is a specific inhibitor of rabbit PKR but contains a loss-of-function mutation in Australian virus isolates. *Proc Natl Acad Sci USA* **113**, 3855–3860, <https://doi.org/10.1073/pnas.1515613113> (2016).
84. Kerr, P. J. *et al.* Next step in the ongoing arms race between myxoma virus and wild rabbits in Australia is a novel disease phenotype. *Proc Natl Acad Sci USA* **114**, 9397–9402, <https://doi.org/10.1073/pnas.1710336114> (2017).
85. Klingen, T. R., Loers, J. & McHardy, A. C. PatchDetection data, <https://doi.org/10.5281/zenodo.1475963> (2018).
86. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).

87. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current protocols in bioinformatics* **54**, 5.6.1–5.6.37, <https://doi.org/10.1002/cpbi.3> (2016).
88. Bao, Y. *et al.* The influenza virus resource at the National Center for Biotechnology Information. *J Virol* **82**, 596–601, <https://doi.org/10.1128/jvi.02005-07> (2008).
89. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797, <https://doi.org/10.1093/nar/gkh340> (2004).
90. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973, <https://doi.org/10.1093/bioinformatics/btp348> (2009).
91. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, W609–612, <https://doi.org/10.1093/nar/gkl315> (2006).
92. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641–1650, <https://doi.org/10.1093/molbev/msp077> (2009).
93. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490, <https://doi.org/10.1371/journal.pone.0009490> (2010).
94. Fitch, W. M. & Farris, J. S. Evolutionary trees with minimum nucleotide replacements from amino acid sequences. *J Mol Evol* **3**, 263–278 (1974).
95. Munch, P. C., Stecher, B. & McHardy, A. C. EDEN: evolutionary dynamics within environments. *Bioinformatics* **33**, 3292–3295, <https://doi.org/10.1093/bioinformatics/btx394> (2017).
96. McWhite, C. D., Meyer, A. G. & Wilke, C. O. Sequence amplification via cell passaging creates spurious signals of positive adaptation in influenza virus H3N2 hemagglutinin. *Virus Evol* **2**, <https://doi.org/10.1093/ve/vew026> (2016).
97. Perry, J. W., Kent, A. & Berry, M. M. Machine literature searching X. Machine language; factors underlying its design and development. *American Documentation* **6**, 242–254, <https://doi.org/10.1002/asi.5090060411> (1955).
98. Matrosovich, M., Zhou, N., Kawaoka, Y. & Webster, R. The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *J Virol* **73**, 1146–1155 (1999).

## Acknowledgements

This work was supported by the Helmholtz foundation. The Heinrich Pette Institute, Leibniz Institute for Experimental Virology is supported by the Free and Hanseatic City of Hamburg and the Federal Ministry of Health. We are grateful to Prof. Thomas Krey for providing the homology models of the pH1N1 and H3N2 polymerase proteins. We thank Hanna Markowsky for excellent technical assistance in performing the NA activity assays.

## Author Contributions

A.C.M. conceived the study. A.C.M. and T.R.K. planned and coordinated the study. A.C.M., T.R.K. and J.L. designed the methodology. T.R.K. and J.L. maintained the data, implemented the software and created results. G.G. and S.S.B. performed wet lab experiments. A.C.M., J.L., T.R.K., G.G. and S.S.B. analyzed the results. T.R.K., J.L. and S.S.B. generated all visualizations. T.R.K., J.L. and A.C.M. wrote the original draft of this manuscript and all authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-42614-y>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019