

Article

Whole Genome Sequencing of A(H3N2) Influenza Viruses Reveals Variants Associated with Severity during the 2016–2017 Season

Bruno Simon ^{1,2}, Maxime Pichon ^{1,2,3}, Martine Valette ^{1,2,3}, Gwendolyne Burfin ³, Mathilde Richard ⁴, Bruno Lina ^{1,2,3} and Laurence Josset ^{1,2,3,*}

¹ Virpath, CIRI, Univ Lyon, Inserm U1111 CNRS UMR5308, ENS, UCBL, Faculté de Médecine Lyon Est, 69372 Lyon, France; sib0.smb@gmail.com (B.S.); Maxime.PICHON@chu-poitiers.fr (M.P.); martine.valette@chu-lyon.fr (M.V.); bruno.lina@chu-lyon.fr (B.L.)

² Laboratoire de Virologie, IAI, CBN, Groupement Hospitalier Nord, Hospices Civils de Lyon, 69317 Lyon, France

³ Centre National des Virus Respiratoires, IAI, CBN, Groupement Hospitalier Nord, Hospices Civils de Lyon, 69317 Lyon, France; gwendolyne.burfin@chu-lyon.fr

⁴ Department of Viroscience, Erasmus MC, 3000 Rotterdam, The Netherlands; m.richard@erasmusmc.nl

* Correspondence: laurence.josset@chu-lyon.fr; Tel.: +33-(0)4-72-07-10-22

Received: 18 December 2018; Accepted: 22 January 2019; Published: 28 January 2019



Abstract: Influenza viruses cause a remarkable disease burden and significant morbidity and mortality worldwide, and these impacts vary between seasons. To understand the mechanisms associated with these differences, a comprehensive approach is needed to characterize the impact of influenza genomic traits on the burden of disease. During 2016–2017, a year with severe A(H3N2), we sequenced 176 A(H3N2) influenza genomes using next generation sequencing (NGS) for routine surveillance of circulating influenza viruses collected via the French national influenza community-based surveillance network or from patients hospitalized in the intensive care units of the University Hospitals of Lyon, France. Taking into account confounding factors, sequencing and clinical data were used to identify genomic variants and quasispecies associated with influenza severity or vaccine failure. Several amino acid substitutions significantly associated with clinical traits were found, including NA V263I and NS1 K196E which were associated with severity and co-occurred only in viruses from the 3c.2a1 clade. Additionally, we observed that intra-host diversity as a whole and on a specific set of gene segments increased with severity. These results support the use of whole genome sequencing as a tool for the identification of genetic traits associated with severe influenza in the context of influenza surveillance.

Keywords: influenza; NGS; severity; vaccination; epidemiology; quasispecies

1. Introduction

Seasonal influenza infections remain a considerable health and socioeconomic challenge, as vaccine uptake and effectiveness are still limited. The resulting morbidity and mortality have only marginally decreased in recent years, with annual epidemics still causing around 290,000 to 650,000 respiratory deaths worldwide [1]. This continuing burden leads to the constant adaptation of public health measures, including vaccination policies which remain the most potentially effective method to limit the impact of influenza [2]. However, current vaccines have moderate effectiveness, in part because of major and minor mismatches between vaccine and circulating influenza viruses, and weak immune responses elicited in the elderly. Severe influenza occurs most frequently in certain demographic groups, including the very young and the very old, but also those suffering from chronic

respiratory, cardiac, and metabolic diseases, as well as pregnant women and obese patients, as shown for the A(H1N1) pdm09 virus [3,4]. Antivirals that control the infection in some patients at risk of complications are available and usually recommended, yet resistant strains often appear in cases of prolonged treatments. In this context, influenza surveillance is mostly focused on studying antigenic variants to select vaccine strains [5] or resistance mutations to monitor antiviral effectiveness [6,7]. Many mutations in the influenza genome have been linked to virulence but clinical studies are scarce and have had, apparently, a minimal impact on clinical management, unlike what is routinely done in the case of bacterial infections [8]. This may be in relation to influenza pathogenesis that is dependent on complex and still-unpredictable mechanisms, including multigenic equilibrium and interactions within the viral population and with the host immunity, in the context of a fast-evolving virus with a highly error-prone replication. Moreover, influenza substitutions can only be considered within a specific viral background, as it was shown that their impact differed even among strains of the same subtype [6]. Therefore, global approaches considering virological and immunological parameters associated with clinical severity, as a whole, are better suited to investigate influenza pathogenesis [9,10].

Recently, the improved accessibility of next generation sequencing technologies has allowed their use in contexts closer to routine clinical practice such as in University hospital centers and national surveillance networks. This has led to considering each infection as being caused by a swarm of nearly identical viruses (or quasispecies) rather than by an individual strain. For influenza, quasispecies were found to have an important role in the pathophysiology of infections and to be involved in virulence and intra-host and inter-host spread [11–15]. Several beneficial characteristics of quasispecies at the host scale have been reported [11]. For example, it was shown that minor resistance variants pre-exist any treatment despite lower fitness and that, at least in vitro, a cooperation between variants for an improved global fitness exists [16,17]. A(H3N2) viruses are considered to evolve faster than other subtypes [18,19]; in recent years they have been found to have high clade diversity and increased morbidity and mortality [20,21]. During the 2016–2017 influenza epidemic, the majority of influenza viruses circulating in the northern hemisphere were A(H3N2) viruses from clade 3c.2a. In France, the dominance of A(H3N2) viruses was associated with an early season of moderate intensity, but with a high severity particularly among the elderly in a context of low vaccine coverage and sub-optimal vaccine effectiveness [22]. Low vaccine effectiveness was related to a hemagglutinin mutation in the 2016–2017 egg-adapted H3N2 (clade 3c.2a) vaccine strain A/HongKong/4801/2014 that altered its antigenicity [23]. As a consequence, there is a growing interest in performing in-depth genomic analysis of these H3N2 viruses, including at the quasispecies level.

We describe herein the genomic characteristics of Influenza A (H3N2) clinical strains recovered via the French national influenza surveillance network or from patients hospitalized in the intensive care units of the University Hospitals of Lyon, France. Several substitutions were significantly associated with severe cases and/or ineffective vaccination, some of them only in specific backgrounds. Individually, minor variants did not appear to be associated with any clinical traits, whereas the intra-host diversity as a whole and on a specific limited set of gene segments increased with severity.

2. Materials and Methods

2.1. Ethics Statement

This is a retrospective study, requiring no additional samples from patients as they were obtained during routine diagnosis. This study is considered non-interventional, as clinical and therapeutic data collection was performed retroactively from a clinical files database, respecting patient confidentiality. This study was approved by the ethics committee of the Lyon University Hospitals (Hospices Civils de Lyon, France), on 3 May, 2017.

2.2. Clinical Samples

Samples included in this study were recovered between May 2016 and April 2017 via the French national influenza surveillance network or from patients hospitalized in Lyon University Hospitals. This cohort was designed to investigate molecular evolution of influenza viruses and studied herein in a secondary analysis. The French national influenza surveillance network recruits patients with influenza-like illnesses in ambulatory care and in University hospitals. Only data from the southern half of France were included. In the Lyon University hospitals systematic surveillance was set up for influenza cases hospitalized in the intensive care units and for influenza in vaccinated patients. When several samples were retrieved and/or sequenced for a patient, only the earliest sampled after the onset of symptoms was included. For the present study, only treatment-naïve patients infected by the A(H3N2) strains for which we obtained at least partial sequences for each genomic segment were included. Clinical data were retrieved from questionnaires filled out by the healthcare professionals who took the samples and from computerized clinical files for the surveillance network and Lyon University Hospitals, respectively. All patients had sufficient clinical data reported to classify them as having either mild or severe influenza. Clinical criteria leading to classification of patients as severe included admission to intensive care for respiratory complications, need for continuous supplemental oxygen or hypoxemia, neurological signs (seizures, fainting with loss of consciousness, evocative imagery, or lumbar puncture) or any type of shock. Severe cases without any major respiratory signs were subclassified depending on their main symptomatology (Table 1).

Table 1. Included patients baseline demographics and clinical and sequenced sample characteristics.

		Patient Characteristics	Mild Outcome	Severe Outcome
Baseline Demographics	Number of patients (% of total)		97 (55%)	79 (45%)
	Age in median years (range)		34 (0–91)	73 (1–97)
	Sex	Male	48	36
		Female	48	43
	Origin	Lyon University Hospitals	6	65
Surveillance network		91	14	
Sample Characteristics	Sample type	NPA	2	6
		NS	95	49
		TBA	0	13
		BAL	0	11
	Median time since onset of symptoms, days (range)		1 (0–5)	3 (0–8)
Viral load—median cycle threshold (range)		19.7 (14.9–30)	23.3 (17–35.3)	
Clinical Characteristics	Vaccinated against Influenza for the current season		21	7
	Severe outcome risk factor *		17	65
	Main severity component	Respiratory		70
		Neurological	-	5
Multiple organ failure			4	

* Risk factors included age >65 years, immunodeficiency, ongoing cancer, chronic pulmonary, cardio-vascular, or neuro-muscular disease, BMI >35, diabetes, and ongoing pregnancy. NPA: Nasopharyngeal aspirate; NS: Nasal swab; TBA: Tracheobronchial aspirate; BAL: Bronchoalveolar Lavage.

2.3. RNA Extraction, Viral Load Determination, and Full-Genome Amplification

The following protocol was followed for all included samples, irrespective of source, sequencing date, or patient origin (Lyon University Hospital or French national influenza surveillance network). RNA was extracted from clinical respiratory samples conserved at $-20\text{ }^{\circ}\text{C}$ using Nuclisens EasyMag automated extraction platform (bioMérieux, Marcy-l’Etoile, France). In order to lower human and bacterial DNA content of the extract, nucleic acids were then incubated with $1\text{ }\mu\text{L}$ of Turbo DNase, $2\text{ }\mu\text{L}$ of Turbo DNase buffer (Life Technologies, Carlsbad, CA, USA), and $0.5\text{ }\mu\text{L}$ of RNasin plus RNase Inhibitor (Promega Corporation, Madison, WI, USA) for 90 min at $37\text{ }^{\circ}\text{C}$ before stopping the enzymatic

reaction through immediate purification by magnetic beads (0.5×; NucleoMag® NGS Clean-up and Size Select, Macherey–Nagel, Düren, Germany). Each segment was retro-transcribed and amplified using a published multi-segment reverse transcription protocol, optimized by adding 0.5 µl of RNasin plus RNase Inhibitor [24].

2.4. Illumina Sequencing

The DNA library was prepared using the commercial Nextera™ XT DNA library kit (Illumina, San Diego, CA, USA) consisting of enzymatic tagmentation, barcoding by PCR (12 cycles), and size-selection using magnetic beads. Then, the sequencing run was performed on a NextSeq500 using a mid-output cartridge resulting in 2×150 nucleotides paired-end reads. Fluorescent images were analyzed using the Illumina RTA1.8/SCS2.8 base-calling pipeline to obtain demultiplexed FASTQ-formatted sequence data.

2.5. Bioinformatic Analysis

As described previously, the sequencing files were cleaned from human reads using bmtagger (v1.1.0) with human genome database hg19 [25,26]. Adapters and low quality ends of reads (Phred score <20) were removed by cutadapt (v0.4.4), also removing reads smaller than 50 bases and the paired reads associated with them [27]. The BWA-MEM alignment tool (v0.7.15) was used to verify sequenced strain subtypes by hemagglutinin (HA) homology, and then to map the reads on the circulating A(H3N2) strains' common ancestor A/Perth/16/2009 [28]. This step generated consensus sequences that were used as the patients' own mapping reference in further steps. All read mapping by BWA-MEM was performed using default settings, except for gap opening and extension penalties which were set to 10 and 2, respectively, to increase indel stringency. Sorted mapped reads were manually inspected using FastQC (v0.11.5) to verify read quality. Variant calling files were generated with naive variant caller (Biomina Galaxy platform) reporting all necessary base metrics for each position on the A(H3N2) genome, ignoring low-quality bases [29]. A homemade vcf analysis python script applied filters to validate possible single nucleotide polymorphisms and generate end-user workable files. These filters comprised a quantitative strand bias (SB) score estimating an uncertainty for variant frequencies when compared to the consensus bases disequilibrium and a depth-adapted threshold. This depth-adapted threshold was calculated following a t-test distribution derived from the commonly-approved 100X threshold necessary to exclude heterozygote mutations and was verified empirically [25]. We determined minimal depth thresholds for variant calling as 2000 and 10,000 reads for variants above 5% and 1% frequencies, respectively, and a relative strand disequilibrium lower than the variant estimated frequency.

2.6. Statistical Analysis

Statistical differences between experimental groups were analyzed using chi-square tests for amino acid substitutions. The diversity analyses were performed using the Wilcoxon rank sum test comparing the count of minor variants on the nucleotide level for each segment. In each case, *p*-values < 0.05 were considered significant. Phylogenetic trees were generated using the Mega7 Neighbor-Joining method, as performed by the European center for disease prevention and control, and using the Poisson correction method and by considering deletions pairwise [30]. Reassortment events were identified using the Graph-incompatibility-based Reassortment Finder (GiRaF) (v1.02), considering an event as validated if three pairwise segment comparisons supported the hypothesis with a confidence >0.95 [31].

2.7. Accession Numbers

The Bioproject accession number for the raw sequences is SRPXXXXXPRJNA516781 and all genomic sequences used in this study are available on the GISAID databank.

3. Results

3.1. Sequencing Efficacy

Among the 176 samples, 155 whole A(H3N2) consensus genomes were successfully sequenced when considering only those with more than 95% coverage. Overall, sequencing efficacy varied as a function of segment size; shorter segments were sequenced to a greater depth. However, sequencing depth was constantly low on two thirds of segment 2 (coding for the polymerase basic protein 1, PB1) and, for some patients, lower on the central part of viral polymerase coding segments (segments 1–3; Figure 1). After excluding segment 2, a mean depth on all segments above 10,000X was obtained for 78 A(H3N2) genomes, and a mean depth of 2000X for 142 A(H3N2) genomes. These depths of coverage allow calling variants above 1% and 5% frequencies, respectively.

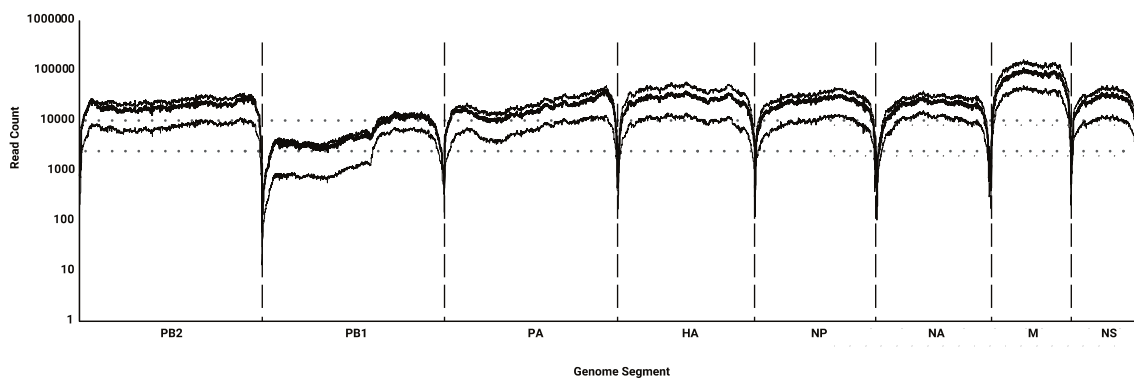


Figure 1. Segment coverage of A(H3N2) influenza virus genomes. Mean depth line for each segment is framed by thinner lines corresponding to the first and third quartiles of depth of coverage. Minimal depth thresholds for 5% (2000X) and 1% (10,000X) variant calling are represented by dot lines.

3.2. Distribution of Consensus Variants According to Clades, Severity, or Vaccination

Substitution associations with severity or vaccine failure were sought. All 155 A(H3N2) whole genome reads sufficiently covered for the generation of consensus sequences were used in phylogenetic and consensus variant analysis. Among them, 121 viruses belonged to genetic clade 3c.2a1, eight belonged to 3c.2a2, 24 belonged 3c.2a3, one belonged to 3c.2a4, and one belonged to genetic clade 3c.3a (the only virus not part of the 3c.2a clade; Figure 2). Within clade 3c.2a1, different subgroups could be further defined based on additional HA substitutions. One virus belonged to the recently-defined subclade 3c.2a1a (T135K, HA2 G150E); 32 viruses belonged to a clade 3c.2a1 subgroup, close to vaccine strain A/Singapore/Infimh-16-0019/2016, defined by N121K and, for 14 of them, by G150E (HA2 numbering); 88 viruses belonged to another clade 3c.2a1 subgroup, close to A/England/919/2016 (a Bolzano-like strain), whether defined by R142G or other various substitutions (A/England/919/2016 cluster; Figure 2).

Overall, severe, vaccinated, and Lyon University Hospitals cases were broadly distributed, period- and clade-wise. None of the well-described treatment resistance substitutions (neuraminidase (NA): E119V, D151E, I222V, R224K, E276D, R292K, and R371K) [32] were present in any consensus genome, except for the conserved amantadine resistance substitution S31N. However, several substitutions across the whole genome could significantly distinguish mild from severe cases and vaccinated from non-vaccinated cases. These substitutions were distributed on the whole influenza proteome, however most were on HA ($n = 8$), NA ($n = 7$), and non-structural protein 1 (NS1; $n = 4$). Substitutions associated with vaccination status exclusively ($n = 19$) were more varied but had weaker statistical significance than those exclusively associated with severity. For the latter, two substitutions, V263I on NA and K196E on NS1, were particularly significant ($p < 0.0001$; Table 2a). As represented in Figure 2, NA V263I and NS1 K196E were mostly found in a specific 3c.2a1 clade background (as defined by HA phylogeny). The NA substitution was only present alone in two strains associated with mild disease outside this background, whereas the NS1 substitution was not necessarily in NA-substituted strains. However, when co-occurrent, these mutations distinguished mild from severe cases more precisely (Table 2b). Based on 3D structures, it could be noted that some of the following substitutions were found: On HA positions 131 and 144 which are close to its receptor binding site (Figure 3a), on NS1 p85 β binding site positions 99 and 146 (Figure 3b), and on NA position 149 which is close to its sialidase active site (Figure 3c). Additionally, reported substitutions were verified as not being linked to depth-specific features of the respiratory tract sites. When comparing Lyon University Hospital cases, comprising both lower and higher respiratory tract samples (50 and 21 samples included in the analysis, respectively), none of the reported substitutions were associated with sample type.

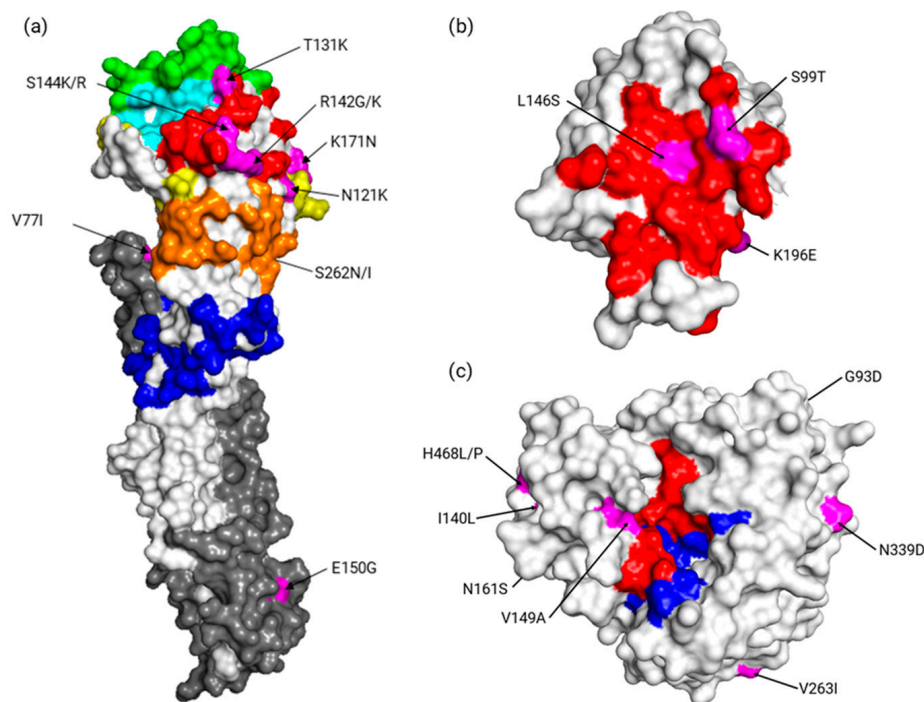


Figure 3. Contextualized locations of substitutions described in our study (in magenta). (a) On HA (PDB accession no. 4O5N) with antigenic sites A (red), B (green), C (blue), D (yellow), and E (orange) highlighted and the receptor binding site in cyan. (b) On NS1 (PDB accession no. 3EE8) with the p85 β -binding site colored red. (c) On NA (PDB accession no. 6BR6) with sialidase catalytic residues in red and framework residues in blue. The image was generated with PyMOL software (Delano Scientific, San Carlos, CA, USA).

Table 2. Substitutions significantly distinguishing mild from severe cases, and vaccinated from non-vaccinated cases. (a) The amino acid of each significant position is indicated for the seasonal consensus, observed variants and, when significant, variants associated with a group. Cases where, for a given position, no amino acid was significantly associated with a group are designated by dashes. (b) NA V263I and NS1 K196E occurrences in mild and severe groups.

(a)																
	PB2		PB1		PA	HA1						HA2		NP		
	255	480	52	614	565	121	131	142	144	171	262	77	150	450	472	
Significant <i>p</i> <	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.01	0.05	0.05	0.05	0.01	0.01	0.05	0.05	
Consensus	V	V	R	E	V	N	T	R	S	K	S	V	E	G	A	
Variants	I	I	W/K	D	M	K	K	G/K	K/R	N	N/I	I	G	S/N	T	
Severe cases	-	-	-	-	-	N	T	R	-	-	-	-	-	-	-	
Vaccinated cases	V	V	W/K	D	M	N	-	-	S	K	S	V	E	G	A	
	NA						M1	M2	NS1				NS2			
	93	140	149	161	263	339	468	147	23	99	146	196	224	67		
Significant <i>p</i> <	0.05	0.01	0.001	0.05	1×10^{-4}	0.05	0.05	0.05	0.05	0.05	0.05	1×10^{-4}	0.05	0.05		
Consensus	G	I	V	N	V	N	H	V	S	S	L	K	R	E		
Variants	D	L	A	S	V	D	L/P	M	N	T	S	E	S/K	A/K		
Severe cases	-	I	A	-	I	-	-	-	S	-	-	E	R	E		
Vaccinated cases	G	I	A	N	-	N	H	M	-	S	L	E	-	-		

(b)									
	Mild	NS1		Total		Severe	NS1		Total
		K196	196E				K196	196E	
NA	V263	64	13	77	NA	V263	32	9	41
	263I	2	8	10		263I	0	27	27
Total		66	21	87	Total		32	36	68

HA clustering of the strains did not seem to originate from within-subtype reassortments, possibly implying the substitutions could have hitchhiked alongside another event rather than being implicated directly in observed correlations. Phylogenetic trees generated from proteins other than HA found that the same strains clustered consistently, however without being similar to such an extent as to indicating a possible common index case. When verified by GiRaF, only one strain was detected as possibly resulting from a reassortment event. This strain, A/Poitiers/917/2017, presented genomic segments closely related to A/Switzerland/9715293/2013, the 3c.3a reference strain, except for its NS1 coding segment corresponding to a 3c.2a1 genotype.

3.3. Diversity and Quasispecies Analysis

When analyzed individually, no significant role of minor variants was observed in distinguishing the clinical groups. These minor variants were rarely at the same positions as the consensus substitutions. An exception to this was found on the HA. The HA 160K substitution, resulting in the loss of a glycosylation site for these quasispecies, was found in 14 cases (with frequencies ranging from 1.2% to 98%, median 7.1%) and was significantly associated with severity ($n = 11$; $p < 0.05$). Otherwise, the same minor variant was rarely found in several patients and at most in three patients ($n < 4$) and a major part of polymerase-coding segment diversity corresponded to deletions. In order to discern potential associations between viral particle diversity and clinical features, we had to perform the analysis of global diversity by segment in each case. Considering that sequencing depth was variable, groups were defined according to segment and targeted depth threshold. The mean diversities of segments 3, 4, and 6 (coding for polymerase acidic (PA), HA, and NA proteins, respectively) were significantly higher for severe cases than in mild cases (Figure 4).

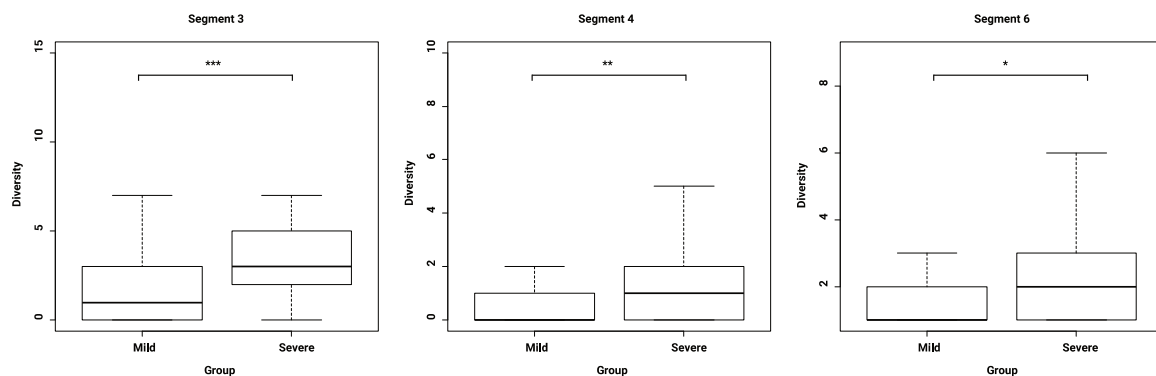


Figure 4. Box plots comparing viral diversity, when considering variants above 1%, for segments with a significant p -value. The whiskers of the boxplots show the range of the data comprised in the first and last quartiles. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$.

Note that a considerable number of mutants had frequencies between 1% and 5%, thus lowering the depth threshold to 1% was necessary to notice the diversity difference for segment 6. In comparison, diversity did not vary significantly with vaccination status (Table 3). Once more, as a control, diversity was verified as not being significantly associated with respiratory tract sampling depth in the cases from the Lyon University Hospitals.

Table 3. Estimated mean quasispecies diversity comparison according to segment. Number of samples (n), mean diversity, and Mann Whitney Wilcoxon test *p*-value resulting from their comparison, when considering whether variants above 1% or 5% are reported in each case.

			Segment								
			1	2	3	4	5	6	7	8	
Mild/Severe	Threshold	1	n	62/47	6/9	58/46	74/56	61/45	64/49	82/68	60/46
		Mean	1.88/2.38	2.33/2.55	2.03/3.46	0.99/1.54	2.75/2.73	1.64/2.24	0.41/0.65	0.42/0.59	
		p	0.17	0.81	<0.001	<0.005	0.88	0.03	0.53	0.17	
	Threshold	5	n	80/69	53/38	80/67	85/69	81/65	83/68	91/73	83/68
		Mean	1.68/1.95	0.83/1.05	1.71/2.84	0.93/1.43	2.40/2.23	1.64/2.07	0.41/0.60	0.42/0.54	
		p	0.40	0.26	<0.001	<0.005	0.73	0.27	0.75	0.25	
Not Vaccinated/Vaccinated	Threshold	1	n	45/23	5/3	44/22	56/26	46/22	51/23	65/26	45/19
		Mean	2.22/1.78	3.4/1.3	2.84/2.13	1.20/0.77	3.22/2.27	1.84/1.47	0.51/0.35	0.44/0.47	
		p	0.57	0.36	0.49	0.33	0.05	0.49	0.76	0.58	
	Threshold	5	n	62/26	37/17	62/26	67/27	63/26	65/27	69/27	65/25
		Mean	1.85/1.69	1.14/0.53	2.23/2.00	1.18/0.74	2.68/2.15	1.75/1.63	0.51/0.33	0.45/0.52	
		p	0.76	0.26	0.89	0.28	0.30	0.77	0.62	0.47	

4. Discussion

In the present study we conducted a comprehensive genomic characterization of H3N2 viruses from either the French national influenza surveillance network or the Lyon University Hospitals' intensive care wards using a combination of several approaches. This allowed the impact of antigenic proximity, genomic substitutions, quasispecies diversity, and reassortment events to be studied concomitantly, and the results therefore provide a comprehensive view of mechanisms involved in the severity of the 2016–2017 seasonal influenza in France.

These data were generated from a secondary analysis of a cohort designed to investigate molecular evolution of the influenza virus. The size of the study did not permit the performance of multivariate analysis correcting for sample nature, comorbidities, geographic, or temporal confounding factors that could influence influenza severity, a limitation that is shared with previous studies [8]. Furthermore, most severe cases were, by design, from the same region albeit with several hospital centers. However, sequencing was performed directly on clinical samples, avoiding any cell culture or experimental infection bias [33,34]. Clinical samples from this study were collected exclusively from treatment-naïve patients. They were fairly equally distributed between mild and severe cases, and can be considered to be representative of the spectrum of influenza diseases in the 2016–2017 season. It should be noted that the specimens of surveillance network patients were mostly collected in the upper respiratory tract and earlier after the onset of symptoms than those from hospitalized cases. However, there was no significant difference according to site of respiratory tract sampling within the Lyon University Hospital cases, as this group comprised both upper and lower respiratory tract sample types which could be a source of diversity by itself or by being correlated to viral copy number. In addition, we did not find any substitutions in the HA receptor binding site contrary to what could be expected if influenza viruses' mutations were driven by location on the respiratory tract. Nonetheless, these results should be interpreted as hypothesis generating and future studies validating the impact of mutations described here are warranted.

Several substitutions were associated with severity or vaccination status in the study population. Among these, most had not been extensively studied *in vitro* or *in vivo*, hence determining their actual involvement would need further investigations. The two most significant ones, NA V263I and NS1 K196E ($p < 0.001$), have previously been described in other contexts and were strongly associated with the severe group. Note that these substitutions were mostly found in strains belonging to a subgroup of the 3c.2a1 clade. This subgroup differed from the others in the 3c.2a1 clade by not harboring commonly found substitutions HA1 N121K and HA2 G150E. Interestingly, both substitutions were reported co-varying alongside substitutions altering p85 β binding by NS1 in H1N1 strains [35]. Additionally, the substitution NS1 K196E was reported to increase virulence in mice when induced in an H5N1 backbone [36]. To the best of our knowledge, NA V263I has not been studied in an H3N2 background yet. However, a comprehensive study of NA V263I involvement at the structural level suggested the isoleucine, which is two methyl groups higher than a valine, could allow important changes by altering hindrance effects [37]. It was also suggested that this site was part of an epitope that was suitable for therapeutic monoclonal antibodies, and that an amino acid change at this position modified the immune response [38]. We also reported that a substitution on V149A on NA, which is close to the active site, may have some impact on the sialidase activity. Substitutions significantly associated with vaccination status comprised many clade-defining HA substitutions (HA1 N121K, S144K, N171K, HA2 I77V, G150E), reflecting strains antigenically less related to A/Hong Kong/4801/2014, the 2016–2017 H3N2 vaccine strain. Notably, substitutions were found on position 131 and 144 close to positions 155 and 145 which have been shown to be crucial for the evolution of H3N2 viruses [5]. As vaccination efficacy relies on the antigenic proximity with vaccinal strains, these associations with putative antigenic changes, albeit statistically weaker, make sense and tend to support the validity of the results presented herein.

Next generation sequencing (NGS) has brought to light new areas of research in microbiology, in particular on genetic diversity and on the implication of quasispecies in physiopathology [39–44].

For instance, certain influenza virus quasispecies have been reported fostering cell-to-cell mechanisms under treatment pressure [13], but also changing viral tropism [8]. When analyzed on a global genomic diversity scale, we found an association between severity and high diversity whereas vaccination status had no impact, as previously described [45,46]. Herein, we also found that this association concerned segments 3, 4, and 6 (coding for PA, HA, and NA respectively) only. This is likely to be in relation to specific selective pressure experienced by each segment that is mainly dependent on the “swarm” needs, viral particles constitution, and coded proteins [11]. For now, hypothesizing why these segments show a significant difference would not be supported by any experimental evidence. Thus, we believe genetic diversity should be considered for each segment independently. As explained in the sequencing efficacy description of the samples herein, an inconsistent read depth was observed on polymerase-coding segments, which could suggest the existence of defective particles in certain cases [15]. We could not conclude the reliability of these observations as the study was not designed to explore this complex problem. An approach focused on anticipating the difficulties related to molecular biology and bioinformatic analysis in this case is required.

In conclusion, the global approach of influenza sequencing results reported herein indicate that multiple factors need to be considered to understand the causes of the 2016–2017 season epidemic severity and support the potential added value of whole genome sequencing for influenza surveillance.

Author Contributions: Conceptualization, M.P. and L.J.; methodology, M.P., B.S.; software, B.S.; validation, B.S., L.J.; formal analysis, B.S., M.R.; investigation, M.P., G.B.; resources, M.V., L.J.; data curation, B.S., M.V.; writing—original draft preparation, B.S., M.P., L.J.; writing—review and editing, B.S., L.J., M.P., B.L., M.V., M.R., G.B.; visualization, B.S.; supervision, B.L., L.J.; project administration, L.J.; funding acquisition, L.J., B.L.

Funding: This study was performed in the field of the global surveillance of influenza viruses performed by the WHO and by the “Centre National de Référence des Virus des Infections Respiratoires” supported by “Santé Publique France” (SPF) in France. The DRIVE Project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777363. This joint undertaking receives support from the European Union’s Horizon 2020 research and innovation and EFPIA.

Acknowledgments: The authors are very grateful to all clinicians, virologists, and laboratory technicians for their active participation and help in this study. We want to thank Philip Robinson (DRCI, Hospices Civils de Lyon) for his valuable help in the preparation of this manuscript.

Conflicts of Interest: M.P. received a research grant from Theradiag and Hologic and is a member of the young microbiologist section of the French Society for Microbiology. B.L. is the co-chair of the Global Influenza Initiative, a member of the scientific committee of the Global Hospital Influenza Surveillance Network, and a member of the Foundation for Influenza. He received a travel grant from Alere to attend the RICAI 2017 Conference, and a travel grant from Seegene to attend the ECCMID conference in 2017. All personal remuneration stopped in September 2010. L.J. has served on the French Advisory Board for SANOFI on the ‘Quadrivalent Influenza Vaccine’ in 2017 and 2018. The other authors (B.S., M.V., G.B.) declare that there are no conflicts of interest associated with this study.

References

1. WHO Influenza (Seasonal) Fact Sheet. Prevention and Control of Influenza Pandemics and Annual Epidemics. Available online: <http://www.who.int/mediacentre/factsheets/fs211/en/> (accessed on 24 January 2018).
2. Uhart, M.; Bricout, H.; Clay, E.; Largeron, N. Public health and economic impact of seasonal influenza vaccination with quadrivalent influenza vaccines compared to trivalent influenza vaccines in Europe. *Hum. Vaccin. Immunother.* **2016**, *12*, 2259–2268. [CrossRef] [PubMed]
3. Chowell, G.; Bertozzi, S.M.; Colchero, M.A.; Celia, A.A.; Mauricio, H. Severe Respiratory Disease Concurrent with the Circulation of H1N1 Influenza. *N. Engl. J. Med.* **2009**, *361*, 674–679. [CrossRef] [PubMed]
4. Jamieson, D.J.; Honein, M.A.; Rasmussen, S.A.; Williams, J.L.; Swerdlow, D.L.; Biggerstaff, M.S.; Lindstrom, S.; Louie, J.K.; Christ, C.M.; Bohm, S.R. H1N1 2009 influenza virus infection during pregnancy in the USA. *Lancet* **2009**, *374*, 451–458. [CrossRef]
5. Koel, B.F.; Burke, D.F.; Bestebroer, T.M.; van der Vliet, S.; Zondag, G.C.; Vervaet, G.; Skepner, E.; Lewis, N.S.; Spronken, M.I.; Russell, C.A.; et al. Substitutions Near the Receptor Binding Site Determine Major Antigenic Change During Influenza Virus Evolution. *Science* **2013**, *342*, 976–979. [CrossRef]

6. Abed, Y.; Baz, M.; Boivin, G. Impact of neuraminidase mutations conferring influenza resistance to neuraminidase inhibitors in the N1 and N2 genetic backgrounds. *Antivir. Ther.* **2006**, *11*, 971–976. [[PubMed](#)]
7. Boivin, G. Detection and management of antiviral resistance for influenza viruses. *Influenza Other Respir. Viruses* **2013**, *7*, 18–23. [[CrossRef](#)] [[PubMed](#)]
8. Wedde, M.; Wählisch, S.; Wolff, T.; Schweiger, B. Predominance of HA-222D/G Polymorphism in Influenza A(H1N1) pdm09 Viruses Associated with Fatal and Severe Outcomes Recently Circulating in Germany. *PLoS ONE* **2013**, *8*, e57059. [[CrossRef](#)]
9. Hung, I.F.; To, K.K.; Lee, C.K.; Lin, C.K.; Chan, J.F.; Tse, H.; Cheng, V.C.; Chen, H.; Ho, P.L.; Tse, C.W.; et al. Effect of Clinical and Virological Parameters on the Level of Neutralizing Antibody against Pandemic Influenza A Virus H1N1 2009. *Clin. Infect. Dis.* **2010**, *51*, 274–279. [[CrossRef](#)] [[PubMed](#)]
10. Lai, S.; Qin, Y.; Cowling, B.J.; Ren, X.; Wardrop, N.A.; Gilbert, M.; Tsang, T.K.; Wu, P.; Feng, L.; Jiang, H.; et al. Global epidemiology of avian influenza A(H5N1) virus infection in humans, 1997–2015: A systematic review. *Lancet Infect. Dis.* **2016**, *16*, e108–e118. [[CrossRef](#)]
11. Domingo, E.; Sheldon, J.; Perales, C. Viral Quasispecies Evolution. *Microbiol. Mol. Biol. Rev.* **2012**, *76*, 159–216. [[CrossRef](#)]
12. Pappas, C.; Yang, H.; Carney, P.J.; Pearce, M.B.; Katz, J.M.; Stevens, J.; Tumpey, T.M. Assessment of transmission, pathogenesis and adaptation of H2 subtype influenza viruses in ferrets. *Virology* **2015**, *477*, 61–71. [[CrossRef](#)] [[PubMed](#)]
13. Mori, K.; Murano, K.; Ohniwa, R.L.; Kawaguchi, A.; Nagata, K. Oseltamivir Expands Quasispecies of Influenza Virus through Cell-to-cell Transmission. *Sci. Rep.* **2015**, *5*, 9163. [[CrossRef](#)] [[PubMed](#)]
14. Dinis, J.M.; Florek, N.W.; Fatola, O.O.; Moncla, L.H.; Mutschler, J.P.; Charlier, O.K.; Meece, J.K.; Belongia, E.A.; Friedrich, T.C. Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. *J. Virol.* **2016**, *90*, 3355–3365. [[CrossRef](#)] [[PubMed](#)]
15. Vasilijevic, J.; Zamarreño, N.; Oliveros, J.C.; Rodriguez-Frandsen, A.; Gómez, G.; Rodriguez, G.; Pérez-Ruiz, M.; Rey, S.; Barba, I.; Pozo, F. Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. *PLoS Pathogens*. **2017**, *13*, e1006650. [[CrossRef](#)]
16. Ghedin, E.; Holmes, E.C.; DePasse, J.V.; Pinalla, L.T.; Fitch, A.; Hamelin, M.E.; Papenburg, J.; Boivin, G. Presence of Oseltamivir-Resistant Pandemic A/H1N1 Minor Variants Before Drug Therapy With Subsequent Selection and Transmission. *J. Infect. Dis.* **2012**, *206*, 1504–1511. [[CrossRef](#)] [[PubMed](#)]
17. Xue, K.S.; Hooper, K.A.; Ollodart, A.R.; Dings, A.S.; Bloom, J.D. Cooperation between distinct viral variants promotes growth of H3N2 influenza in cell culture. *eLife* **2016**, *5*, e13974. [[CrossRef](#)] [[PubMed](#)]
18. Ferguson, N.M.; Galvani, A.P.; Bush, R.M. Ecological and immunological determinants of influenza evolution. *Nature* **2003**, *422*, 428–433. [[CrossRef](#)]
19. Rambaut, A.; Pybus, O.G.; Nelson, M.I.; Viboud, C.; Taubenberger, J.K.; Holmes, E.C. The genomic and epidemiological dynamics of human influenza A virus. *Nature* **2008**, *453*, 615–619. [[CrossRef](#)]
20. Muscatello, D.J.; Newall, A.T.; Dwyer, D.E.; MacIntyre, C.R. Mortality Attributable to Seasonal and Pandemic Influenza, Australia, 2003 to 2009, Using a Novel Time Series Smoothing Approach. *PLoS ONE* **2013**, *8*, e64734. [[CrossRef](#)]
21. Simonsen, L. The global impact of influenza on morbidity and mortality. *Vaccine* **1999**, *17*, S3–S10. [[CrossRef](#)]
22. Équipes de surveillance de la grippe. Surveillance de la grippe en France, saison 2016-2017. *Bull. Épidémiol. Hebd.* **2017**, 466–475. Available online: http://invs.santepubliquefrance.fr/beh/2017/22/2017_22_1.html (accessed on 28 January 2019).
23. Zost, S.J.; Parkhouse, K.; Gumina, M.E.; Kim, K.; Diaz Perez, S.; Wilson, P.C.; Treanor, J.J.; Sant, A.J.; Cobey, S.; Hensley, S.E. Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12578–12583. [[CrossRef](#)] [[PubMed](#)]
24. Zhou, B.; Donnelly, M.E.; Scholes, D.T.; St George, K.; Hatta, M.; Kawaoka, Y.; Wentworth, D.E. Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for Classical and Swine Origin Human Influenza A Viruses. *J. Virol.* **2009**, *83*, 10309–10313. [[CrossRef](#)] [[PubMed](#)]
25. Pichon, M.; Simon, B.; Valette, M.; Bal, A.; Picard, C.; Escuret, V.; Ottmann, M.; Gillet, Y.; Ader, F.; Lina, B.; et al. Evolution of influenza genome diversity during infection in immunocompetent patients. *bioRxiv* **2018**, 435263. [[CrossRef](#)]

26. Rotmistrovsky, K.; Agarwala, R. BMTagger: Best Match Tagger for removing human reads from metagenomics datasets. Available online: <ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/> (accessed on 26 April 2018).
27. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **2011**, *17*, 10–12. [[CrossRef](#)]
28. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
29. Blankenberg, D.; Von Kuster, G.; Bouvier, E.; Baker, D.; Afgan, E.; Stoler, N.; Taylor, J.; Nekrutenko, A. Dissemination of scientific software with Galaxy ToolShed. *Genome Biol.* **2014**, *15*, 403. [[CrossRef](#)] [[PubMed](#)]
30. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)] [[PubMed](#)]
31. Nagarajan, N.; Kingsford, C. GiRaF: Robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Res.* **2011**, *39*, e34. [[CrossRef](#)] [[PubMed](#)]
32. Summary Table of Neuraminidase Amino Acid Substitutions Associated with Reduced Inhibition by Neuraminidase Inhibitors (NAI). Available online: http://www.who.int/influenza/gisrs_laboratory/antiviral_susceptibility/nai_overview/en/ (accessed on 24 January 2018).
33. Murcia, P.R.; Hughes, J.; Battista, P.; Lloyd, L.; Baillie, G.J.; Ramirez-Gonzalez, R.H.; Ormond, D.; Olivier, K.; Elton, D.; Mumford, J.A.; et al. Evolution of an Eurasian Avian-like Influenza Virus in Naïve and Vaccinated Pigs. *PLoS Pathog.* **2012**, *8*, e1002730. [[CrossRef](#)] [[PubMed](#)]
34. Lee, H.K.; Tang, J.W.-T.; Kong, D.H.-L.; Loh, T.P.; Chiang, D.K.; Lam, T.T.; Koay, E.S. Comparison of Mutation Patterns in Full-Genome A/H3N2 Influenza Sequences Obtained Directly from Clinical Samples and the Same Samples after a Single MDCK Passage. *PLoS ONE* **2013**, *8*, e79252. [[CrossRef](#)] [[PubMed](#)]
35. Lopes, A.M.; Domingues, P.; Zell, R.; Hale, B.G. Structure-Guided Functional Annotation of the Influenza A Virus NS1 Protein Reveals Dynamic Evolution of the p85 β -Binding Site During Circulation in Humans. *J. Virol.* **2017**, *91*, e01081-17. [[CrossRef](#)] [[PubMed](#)]
36. Min, J.-Y.; Santos, C.; Krug, R.; Subbarao, K. Glutamic Acid at Position 196 of the NS1 Protein is Associated with Both a Block in IRF3 Activation and the Virulence of a HPAI H5N1 Virus in Mice. Fourth ESWI Influenza Conference Poster. Available online: <https://core.ac.uk/download/pdf/37965740.pdf> (accessed on 24 April 2018).
37. Ramírez-Salinas, G.L.; García-Machorro, J.; Quiliano, M.; Zimic, M.; Briz, V.; Rojas-Hernández, S.; Correa-Basurto, J. Molecular modeling studies demonstrate key mutations that could affect the ligand recognition by influenza AH1N1 neuraminidase. *J. Mol. Model.* **2015**, *21*, 292.
38. Wan, H.; Yang, H.; Shore, D.A.; Garten, R.J.; Couzens, L.; Gao, J.; Jiang, L.; Carney, P.J.; Villanueva, J.; Stevens, J.; et al. Structural characterization of a protective epitope spanning A(H1N1) pdm09 influenza virus neuraminidase monomers. *Nat. Commun.* **2015**, *6*, 6114. [[CrossRef](#)] [[PubMed](#)]
39. Solmone, M.; Vincenti, D.; Prosperi, M.C.F.; Bruselles, A.; Ippolito, G.; Capobianchi, M.R. Use of Massively Parallel Ultradeep Pyrosequencing to Characterize the Genetic Diversity of Hepatitis B Virus in Drug-Resistant and Drug-Naive Patients and To Detect Minor Variants in Reverse Transcriptase and Hepatitis B S Antigen. *J. Virol.* **2009**, *83*, 1718–1726. [[CrossRef](#)] [[PubMed](#)]
40. Kuroda, M.; Katano, H.; Nakajima, N.; Tobiume, M.; Aina, A.; Sekizuka, T.; Hasegawa, H.; Tashiro, M.; Sasaki, Y.; Arakawa, Y.; et al. Characterization of Quasispecies of Pandemic 2009 Influenza A Virus (A/H1N1/2009) by De Novo Sequencing Using a Next-Generation DNA Sequencer. *PLoS ONE* **2010**, *5*, e10256. [[CrossRef](#)] [[PubMed](#)]
41. Bartolini, B.; Chillemi, G.; Abbate, I.; Bruselles, A.; Rozera, G.; Castrignanò, T.; Paoletti, D.; Picardi, E.; Desideri, A.; Pesole, G.; et al. Assembly and characterization of pandemic influenza A H1N1 genome in nasopharyngeal swabs using high-throughput pyrosequencing. *New Microbiol.* **2011**, *34*, 391–397. [[PubMed](#)]
42. Nishijima, N.; Marusawa, H.; Ueda, Y.; Takahashi, K.; Nasu, A.; Osaki, Y.; Kou, T.; Yazumi, S.; Fujiwara, T.; Tsuchiya, S.; et al. Dynamics of Hepatitis B Virus Quasispecies in Association with Nucleos(t)ide Analogue Treatment Determined by Ultra-Deep Sequencing. *PLoS ONE* **2012**, *7*, e35052. [[CrossRef](#)] [[PubMed](#)]
43. Kukimoto, I.; Maehama, T.; Sekizuka, T.; Ogasawara, Y.; Kondo, K.; Kusumoto-Matsuo, R.; Mori, S.; Ishii, Y.; Takeuchi, T.; Yamaji, T.; et al. Genetic Variation of Human Papillomavirus Type 16 in Individual Clinical Specimens Revealed by Deep Sequencing. *PLoS ONE* **2013**, *8*, e80583. [[CrossRef](#)]

44. Ruiz-Carrascoso, G.; Romero-Gómez, M.P.; Plaza, D.; Mingorance, J. Rapid detection and quantitation of ganciclovir resistance in cytomegalovirus quasispecies. *J. Med. Virol.* **2013**, *85*, 1250–1257. [[CrossRef](#)] [[PubMed](#)]
45. Vignuzzi, M.; Stone, J.K.; Arnold, J.J.; Cameron, C.E.; Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions within a viral population. *Nature* **2006**, *439*, 344–348. [[CrossRef](#)] [[PubMed](#)]
46. Debbink, K.; McCrone, J.T.; Petrie, J.G.; Truscon, R.; Johnson, E.; Mantlo, E.K.; Monto, A.S.; Luring, A.S. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. *PLoS Pathog.* **2017**, *13*, e1006194. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).