GENETIC CHARACTERIZATION OF INFLUENZA A(H3N2) VIRUSES FROM VACCINATED AND UNVACCINATED CHILDREN DURING THAILAND 2013 AND 2014 INFLUENZA SEASONS

Wudtichai Manasatienkij[1,2], Piyawan Chinnawirotpaisan[2], Weerayuth Kittichotirat[3], Louis R Macareo[2], Damon W Ellison[2], Supapon Cheevadhanarak[1], Tawee Chotpitayasunondh[4], Piyarat Suntarattiwong[4], Stefan Fernandez[2], In-Kyu Yoon[2], Sonja J Olsen[5], Wanitchaya Kittikraisak[6], Kathryn B Anderson[2], Wiriya Rutvisuttinunt[7] and Chonticha Klungthong[2]

[1]Biotechnology Division, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi; [2]Department of Virology, Armed Forces Research Institute of Medical Sciences; [3]Systems Biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi; [4]Queen Sirikit National Institute of Child Health, Ministry of Public Health, Bangkok, Thailand; [5]Influenza Division, Centers for Disease Control and Prevention, Atlanta, GA, USA; [6]Influenza Program, Thailand Ministry of Public Health - US Centers for Disease Control and Prevention Collaboration, Nonthaburi, Thailand; [7]Viral Diseases Branch, Walter Reed Army Institute of Research, Silver Spring, MD, USA

Abstract. Vaccination is the best strategy to reduce the burden of influenza infection. Nevertheless, mutations can cause antigenic mismatches between the vaccine strains and circulating strains. These differences can lead to reduced vaccine effectiveness (VE) and illness. Further investigations and molecular characterizations of influenza viruses associated with antigenic drift will increase our understanding of viral genetic factors contributing to reduction in VE. In this study, during the 2013 and 2014 (with low VE) influenza seasons, 15 influenza A(H3N2) viruses from Thai children enrolled in a VE study, who either received (n = 5) or did not receive (n = 10) vaccines of influenza A(H3N2) strain belonging to clade 3C.1, were sequenced. Phylogenetic analysis demonstrated the viruses belonged to clade 3C.2 [5 (1 from vaccinated and 4 from unvaccinated children) from 2013] and clade 3C.2a [10 (4 from vaccinated and 6 from unvaccinated children) from 2014]. The number of single nucleotide variants in viruses from individual children was not different in both seasons. The low VE observed in 2014 suggests a relationship with influenza A(H3N2) clade 3C.2a circulating in that year.

Keywords: influenza A(H3N2) virus, influenza virus genetics, vaccine effectiveness, Thailand

Correspondence: Dr Chonticha Klungthong, Molecular Virology Section, Department of Virology, USAMC-AFRIMS, 315/6 Ratchawithi Road, Ratchatewi, Bangkok 10400, Thailand.
Tel: +66 (0) 2696 2700 Ext 4341, mobile: +66 (0) 81 2685685
E-mail: chontichak.fsn@afrims.org

INTRODUCTION

Influenza is an acute respiratory disease of importance in humans, which spreads worldwide and causes yearly epidemics causing 290,000-650,000 deaths annually (Iuliano *et al*, 2018). Although annual vaccination is the most effective way to prevent influenza virus infection (Castilla *et al*, 2013; Shen *et al*, 2013), the vaccine effectiveness (VE) is dependent on several factors, especially the matching of vaccine strains with circulating viruses (Carrat and Flahault, 2007). In seasons where mismatches occur, this may lead to reduced VE and a higher than expected number of influenza cases.

Due to the high rate of error-prone replication (Webster *et al*, 1992), mutations can gradually accumulate at antigenic sites of the HA1 subunit, which plays a major role in binding to host receptors and is a major target of neutralizing antibodies (Bouvier and Palese, 2008). Major antigenic sites (defined as epitopes A, B, C, D, and E) are located on HA1 and mutations in these epitopes help drive antigenic drift, thereby reducing antibody recognition (Tewawong *et al*, 2015). Antigenic drift in globally circulating influenza viruses is monitored every year to help more appropriate vaccine development (Boni, 2008). Intra-host single nucleotide variants (iSNV) can be produced over the course of an infection and are able to transmit as part of the infecting population (Debbink *et al*, 2017). Many studies have attempted to uncover how host immunity and vaccine-induced immunity influence the level of iSNV (Archetti *et al*, 1950; Boni *et al*, 2006; Dinis *et al*, 2016; Debbink *et al*, 2017); however the significant differences in numbers of iSNVs and amino acid substitutions between those from unvaccinated and vaccinated subjects are not observed in these studies.

In Thailand, the Ministry of Public Health (MOPH) recommends influenza vaccination for persons ≥65 years of age, those with underlying medical conditions, children from 6 months through to 2 years (<36 months) of age, pregnant women, mentally ill individuals, persons weighing >100 kg, and healthcare personnel (Kittikraisak *et al*, 2015). A recent VE study using test negative design in Thai children during the 2013 and 2014 influenza endemic seasons estimated VE against all influenza viruses at 64% [95% confidence interval (CI): 21-84%] in 2013, and 26% (95% CI: -47-63%) in 2014 (Kittikraisak *et al*, 2016). VE against influenza A(H3N2) viruses is estimated at 73% (95% CI: -14-94%) and 6% (95% CI: -103-56%) for 2013 and 2014 seasons, respectively (Kittikraisak *et al*, 2016). Genetic characterization of influenza viruses obtained directly from VE studies may provide more precise information on the effect of viral genetic factors on VE and may assist in improving future vaccine development (Zhu *et al*, 2017).

In this study, 15 influenza A(H3N2) genomes obtained from Thai children enrolled in a referred VE study, who either received or did not receive influenza vaccines during the 2013 and 2014 influenza seasons, were examined for virus genetic factors that might have contributed to low VE of 6% in the 2014 season (Kittikraisak *et al*, 2016).

MATERIALS AND METHODS

Study cohort

This influenza VE study was conducted at the Queen Sirikit National Institute of Child Health (QSNICH), Bangkok, Thailand during the 2013 and

2014 influenza seasons (Kittikraisak *et al*, 2016). All specimens were collected from children of 10 to 56 months (average 25 months) of age.

The study was approved by the ethics committee of QSNICH (Document No.59-009) and the Institutional Review Board (IRB) of the Walter Reed Army Institute of Research, USA (WRAIR#2094). The IRB of the US Centers for Disease Control and Prevention relied on QSNICH ethical approval. Prior written informed consent was obtained from legal guardians of the children enrolled.

Clinical specimens selected for influenza virus genomic sequencing

Samples and definitions used as inclusion criteria for influenza virus genomic sequencing and analysis were: (i) having specimen volume ≥1 ml, (ii) having a cycle threshold (Ct) value from an RT-quantitative PCR assay of influenza virus <22 cycles (Rutvisuttinunt *et al*, 2015), and (iii) were collected from children whose vaccination status was verified as fully vaccinated or unvaccinated (Kittikraisak *et al*, 2016). The influenza seasons were defined as follows: the 2013 season, June 2013 - May 2014 and the 2014 season, June 2014 - May 2015. Based on the selection process (Fig 1), 15 influenza A(H3N2) virus-positive specimens were selected, 5 specimens from 2013 and 10 specimens from 2014 (Table 1). These represented 5/66 (8%) and 10/173 (6%) specimens from the VE study in the 2013 and 2014 seasons, respectively. Among these specimens, 5 (1 from 2013 and 4 from 2014 season) were collected from children who received full vaccination and 10 (4 from 2013 and 6 from 2014 season) from unvaccinated children. The one vaccinated child in 2013 received the 2013 Southern Hemisphere vaccine composed of A/California/7/2009 (2009 pandemic

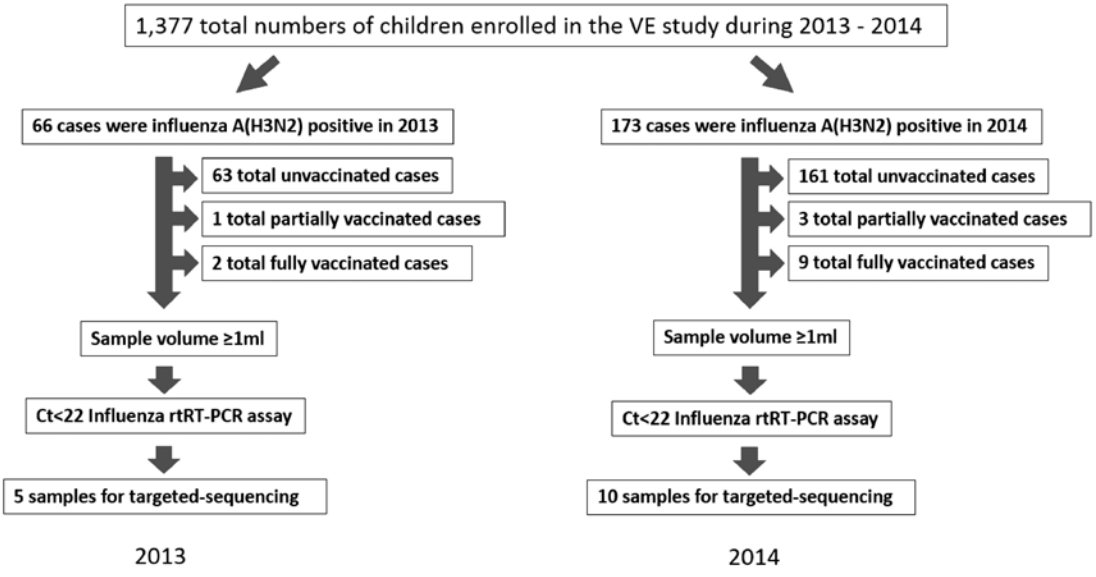

Fig 1-Flow chart depicting selection process in selecting influenza A samples for the study (Kittikraisak *et al*, 2016).

Table 1
Influenza A virus samples obtained from children participating in an influenza vaccine effectiveness study, Bangkok, Thailand, 2013-2015.

No.	Sample ID	Age (month)	Sample collection date	Vaccination date		Vaccine Influenza A(H3N2) strain	rtRT-PCR	
				First	Second		Ct Influenza A virus (M gene)[a]	Ct Influenza A(H3N2) (HA gene)[b]
1	G2-069	55	5 November 2013	28 June 2013	16 August 2013	A/Victoria/361/2011	21	22
2	G1-444	12	18 February 2015	22 November 2014	27 December 2014	A/Texas/50/2012	20	20
3	G1-449	16	22 February 2015	7 July 2014	15 August 2014	A/Texas/50/2012	20	22
4	G1-480	17	13 March 2015	19 July 2014	23 August 2014	A/Texas/50/2012	20	19
5	G1-511	16	2 April 2015	10 June 2014	2 December 2014	A/Texas/50/2012	19	19
6	G2-030	56	8 October 2013	-	-	-	15	15
7	G1-016	22	7 October 2013	-	-	-	15	15
8	G1-027	10	14 October 2013	-	-	-	17	18
9	G1-043	11	8 November 2013	-	-	-	17	18
10	G1-335	12	8 November 2014	-	-	-	15	15
11	G2-564	39	2 December 2014	-	-	-	16	16
12	G1-334	12	8 November 2014	-	-	-	16	17
13	G2-628	34	20 January 2015	-	-	-	17	17
14	G1-491	17	18 March 2015	-	-	-	17	16
15	G2-693	47	3 March 2015	-	-	-	17	17

[a]Cycle threshold (Ct) value obtained from detection of M gene. [b]Ct value obtained from detection of HA gene.

H1N1-like), A/Victoria/361/2011 (H3N2) and B/Wisconsin/1/2010 (Yamagata lineage) (available in Thailand during May-December 2013), and the four in 2014 received the 2014 Southern Hemisphere vaccine composed of A/California/7/2009, A/Texas/50/2012 (H3N2) and B/Massachusetts/2/2012 (Yamagata lineage) (available in Thailand during May-December 2014) (Table 1).

DNA library preparation and whole genome sequencing

RNA of specimens was extracted using QIAamp Viral RNA Mini Kit (QIAGEN, Valencia, CA) and employed in RT-PCR to amplify the influenza virus whole genome. RT-PCR was performed using influenza specific primers, SuperScript III One-Step RT-PCR System and Platinum *Taq* High Fidelity DNA Polymerase (ThermoFisher, Carlsbad, CA) as previously described (Zhou *et al*, 2009). In brief, approximately 10 μg total RNA were added to the reaction mixture containing 0.2 μM MBTuni-12 and MBTuni-13 primers. Thermocycling was performed in Mastereycler Nexus gradient thermal cycler (Eppendorf, Hamburg, Germany) as follows: 45°C for 60 minutes; 94°C for 2 minutes; 5 cycles of 94°C for 30 second, 48°C for 30 seconds and 68°C for 3 minutes; 35 cycles of 94°C for 30 seconds, 57°C for 30 seconds and 68°C for 3 minutes; and a final step of 68°C for 7 minutes. Amplicons were purified using QIAquick® PCR Purification Kit (QIAGEN) and DNA concentration was measured using Qubit® dsDNA HS Assay Kit (ThermoFisher). DNA library preparation employed QIAseq FX DNA library kit (QIAGEN) and 600 μl aliquot of a library pool (15 samples per pool) was placed in a flow cell of a 500 cycle MiSeq reagent Kit v2 (Illumina, San Diego, CA). Pair-end sequencing (2 x 250 bp) was performed on MiSeq instrument (Illumina).

Next generation sequencing (NGS) data analysis

Data of sequence reads obtained from the MiSeq were trimmed to nucleotides (nts) below quality 30 using Trimmomatic tool (Bolger *et al*, 2014) before the paired-end data were analyzed by Trinity V.2.2.0 (Grabherr *et al*, 2011). Contigs longer than 500 nts were compared to a non-redundant nt NCBI database using BLASTN (McGinnis and Madden, 2004) and the best hits were considered reference strains for mapping. Reference sequences were updated with their consensus sequences before being re-mapped using BWA-MEM aligner (Li and Durbin, 2009; Fonseca *et al*, 2012). The consensus sequence was extracted using samtools (Li *et al*, 2009) and bcftools (Narasimhan *et al*, 2016). In order to obtain the whole sequences and iSNV for the influenza virus eight segments, the filtered short reads were aligned to their updated reference sequences.

Read mapping analysis

The average depth of coverage (DOC) was observed at above ~10,000 reads per base with even coverage across the coding region of all segments indicating sufficient redundancy to identify virus consensus sequences and iSNVs (Table 2).

Phylogenetic tree construction, sequence alignment and glycosylation site prediction

A total of 587 influenza A(H3N2) HA1 gene sequences were used to construct the phylogenetic tree. These sequences included 15 sequences obtained from the study, 357 sequences from GenBank and 195 sequences from GISAID databases; all sequences were from influenza A(H3N2) viruses circulating in Thailand from 2011 to 2017. Other sequences were from GenBank and GISAID: 9 sequences of vaccine strains and 11 reference strain sequences with known virus clades. Maximum

Table 2
Influenza A virus genome deep sequencing of samples in Bangkok, Thailand 2013 - 2015.

No.	Sample ID	Total number of reads	Alignment		
			Number of mapped reads	Mapped read (%)	Average DOC
1	G2-069	1,964,970	1,908,118	97.11	26,011.75
2	G1-444	2,697,677	2,659,465	98.58	36,627.45
3	G1-449	2,495,812	2,441,404	97.82	33,099.12
4	G1-480	2,040,210	2,023,045	99.16	27,941.59
5	G1-511	2,963,937	2,943,452	99.31	41,869.09
6	G2-030	2,187,102	2,111,782	96.56	27,777.87
7	G1-016	2,214,503	2,172,512	98.10	29,067.43
8	G1-027	2,178,026	2,155,147	98.95	29,954.13
9	G1-043	2,358,913	2,329,273	98.74	31,859.59
10	G1-335	2,340,645	2,309,014	98.65	31,270.54
11	G2-564	2,007,967	1,989,210	99.07	28,065.65
12	G1-334	2,171,203	2,095,983	96.54	29,548.11
13	G2-628	1,946,210	1,932,683	99.30	26,823.80
14	G1-491	1,492,841	1,483,947	99.40	21,649.45
15	G2-693	1,907,269	1,867,785	97.93	25,462.17

DOC, depth of coverage.

likelihood (ML) tree was constructed with Molecular Evolutionary Genetic Analysis (MEGA) version 6.0 using HKY+G model with 1,000 bootstrap replicates support (Tamura *et al*, 2013). The tree was visualized and annotated using FigTree version 1.4.2 (Rambaut, 2012). Nt and amino acid sequence alignments were performed using MEGA version 6.0. Potential N-linked glycosylation site was predicted using NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) (Blom *et al*, 2004). Amino acid consensus sequences of N-X-S/T, where X represents any amino acid except P, with a threshold value >0.5 is considered to be a potential glycosylation site (Shakin-Eshleman *et al*, 1996; Gupta *et al*, 2004). GenBank accession numbers, including those of the 15 sequences from the study, are listed in Table 3.

ISNV analysis

ISNV analysis was conducted using LoFreq, a per site quality score-based statistical test (Wilm *et al*, 2012; McElroy *et al*, 2013). Positive SNVs were identified using the following criteria: (i) frequency ≥0.02, (ii) minimum depth of coverage ≥10,000, and (iii) variant-supporting reads in both orientations.

Statistical analysis

Shapiro-Wilk test was used to analyze normal distribution of the number of Ct values and the number of iSNVs. Overall

Table 3
GenBank accession numbers of influenza A reference and vaccine strains used in phylogenetic analysis.

Sample	Accession number
Influenza A (H3N2) vaccine strains 2006 - 2017 (n = 9)	KM821334, KM821341, KM978061, GQ293081, KM821347, KC892952, EPI554566, EPI1312170, EPI1313136
Influenza A (H3N2) reference strains (n = 11)	EPI461953, EPI750018, EPI346607, EPI426061, EPI326342, EPI467994, EPI426077, KC892582, EPI460558, EPI460540, EPI1058553
Influenza A (H3N2) circulating in Thailand, 1979 -2010 (n = 99)	CY114429, EF568929, JN617980, JN617982, EF568928, EF568925, EF568926, JN617979, JN617981, EF568924, EF568927, EU021284, EU021274, EU021266, EU021282, EU021278, EU021268, EPI158219, EPI162297, EPI158222, EPI162300, EPI155715, EPI158231, EPI162303, EPI158243, EPI162315, EU021270, EU021272, EU021276, KP637918, KP637926, KP637934, KP637942, KP638046, KP638054, KP638094, KP638142, KP638222, EPI163135, EPI163137, EPI160589, EPI160590, EPI160591, EPI185820, EPI175300, EPI175302, EPI185823, EPI175305, EU625363, EU625364, EU625365, FJ912976, FJ912984, FJ912992, KP637950, KP637958, KP638158, KP638166, KP638102, KP638150, KP638230, EPI606093, EPI232665, KT888617, EPI232585, EPI232602, EPI211734, EPI232608, EPI232614, EPI232617, EPI232620, EPI232623, GQ983548, GQ902809, GU271974, GU271982, GQ902793, GQ902817, GQ902825, CY074942, GU271990, KP637966, KP637974, KP638062, KT888618, EPI278796, EPI294228, EPI919303, EPI606256, KT889260, EPI295271, CY074950, CY074958, CY074966, KP637982, KP637990, KP638174, KP638070, KP638110
Influenza A (H3N2) circulated in Thailand, 2011 -2017 (n = 453)	KT888619, KT888620, KT888621, KT888622, KT888623, KT888624, KT888626, KT888627, KT888630, KT888644, EPI346317, EPI331594, EPI346455, EPI348478, EPI346452, EPI346320, EPI346458, EPI670323, EPI272025, KP335865, KP335869, KP335867, KP335870, KP335871, KP335872, KP335874, KP335875, KP335876, KP335878, KP335877, KP335879, KP335880, KP335881, KP335882, KP335883, KP335884, KP335885, KP335886, KP335887, KP335888, KP335866, KP335868, KP335873, KT888625, KT888628, KT888631, KP637998, KP638006, KP638182, KP638190, KX721054, KP638078, KP638118, KP638214, KT888643, EPI526661, EPI406048, EPI379391, EPI432652, EPI450268, EPI450271, EPI406057, EPI379385, KP335890, KP335891, KP335892, KP335893, KP335894, KP335896, KP335900, KP335902, KP335903, KP335904, KP335889, KP335895, KP335897, KP335898, KP335899, KP335901, KT888629, KT888639, KP638014, KP638022, KP638198, KP638126, KP638206, KJ577149, KJ577157, KJ577189, EPI584728, EPI553213, EPI585034, EPI584730, EPI585037, EPI529568, EPI467248, EPI585036, EPI585038, EPI467244, EPI432623, EPI529571, EPI467242, EPI584731, EPI553204, EPI584733, EPI529587, EPI584732, EPI585035, KP335918, KP335920, KP335921, KP335923, KP335922, KP335925, KP335924, KP335917, KP335928, KP335937, KP335939, KP335916, KP335905, KP335906, KP335907, KP335908, KP335909, KP335910, KP335911, KP335912, KP335913, KP335914, KP335915, KP335919, KP335926, KP335927, KP335929, KP335931, KP335932, KP335933, KP335934, KP335935, KP335940, KP335941, KP335943, KP335930, KP335945, KP335936, KP335938, KP335942, KP335944, KP638030, KP638038, KP638086, KP638134, KP877345, KP877346, KP877347, KP877348, KP877349, KP877350, KP877351, KP877352, KP877353, KP877354, KP877355, KP877356, KP877357, KP877358, KP877359, KP877360, EPI566017, EPI543708, EPI516705, EPI541343,

Table 3
(Continued)

Sample	Accession number
Influenza A (H3N2) circulated in Thailand, 2011 -2017 (n = 453) (*Continued*)	EPI567267, EPI565993, EPI543717, EPI520397, EPI520359, EPI516708, EPI545581, KP335946, KP335959, KP335958, KP335961, KP335962, KP335963, KP335964, KP335966, KP335965, KP335967, KP335969, KP335970, KP335971, KP335972, KP335973, KP335974, KP335975, KP335976, KP335977, KP335979, KP335980, KP335981, KU558934, KU558935, KP335947, KP335948, KP335949, KP335951, KP335952, KP335957, KP335956, KP335950, KP335953, KP335954, KP335955, KP335960, KP335968, KP335978, EPI649620, EPI649661, KP877341, KP877342, KP877343, KP877344, KP877361, KP877362, KP877363, KP877364, KP877365, KP877366, KP877367, KP877368, KP877369, KP877370, KP877371, KP877372, EPI710153, EPI710161, EPI652263, EPI715261, EPI649557, EPI649643, EPI649549, EPI649683, EPI649731, EPI746900, EPI649506, EPI565397, EPI710321, EPI710329, EPI710337, EPI710345, EPI704151, EPI647860, EPI647884, EPI647892, EPI652257, EPI652260, EPI715572, EPI710385, EPI715596, EPI715604, EPI775472, EPI715612, EPI704154, EPI715620, EPI649490, EPI647876, EPI652272, KU558939, KU558943, KU558936, KU558937, KU558938, KU558940, KU558941, KU558946, KU558947, KU558948, KU558949, KU558950, KU558951, KU558952, KU558953, KU558954, KU558955, KU558957, KU558958, KU558959, KU558960, KU558963, KU558964, KU558965, KU558966, KU558968, KU558969, KU558970, KU558971, KU558972, KU558974, KU558976, KU558977, KU558978, KU558942, KU558945, KU558962, KU558967, KU558973, KU558975, KU558944, KU558956, KU558961, EPI925039, EPI925071, EPI925095, EPI825071, EPI925031, EPI919402, EPI813833, EPI813841, EPI813849, EPI813865, EPI919249, EPI813825, EPI919243, EPI925063, EPI925103, EPI813873, EPI919240, EPI925023, EPI836954, EPI919405, EPI925079, EPI960533, EPI925111, EPI836924, EPI919246, EPI919408, EPI925047, EPI925055, EPI813857, EPI919155, EPI925087, EPI960541, EPI919411, EPI903904, EPI331543, EPI903901, MF673262, MF673231, MF673232, MF673233, MF673241, MF673234, MF673235, MF673236, MF673237, MF673238, MF673239, MF673242, MF673243, MF673244, MF673245, MF673246, MF673247, MF673249, MF673250, MF673251, MF673252, MF673253, MF673254, MF673255, MF673256, MF673259, MF673260, MF673261, MF673263, MF673264, MF673265, MF673266, MF673267, MF673268, MF673269, MF673240, MF673248, MF673257, MF673258, EPI1058157, EPI1058165, EPI1044762, EPI1058173, EPI1182959, EPI1044765, EPI1198421, EPI1179551, EPI1058213, EPI1058221, EPI1181106, EPI1182940, EPI1045321, EPI1179437, EPI1182948, EPI1179578, EPI1058397, EPI1197694, EPI1197686, EPI1205488, EPI1179434, EPI1198631, EPI1058493, EPI1197678, EPI1198684, EPI1179548, EPI1044759, EPI1179347, EPI1182929, EPI1185478, EPI1179345, EPI1058523, EPI1058531, EPI1198599, EPI1198655, EPI1179575, EPI1058561, EPI272019, EPI1058569, EPI1058577, MF673270, MF673271, MF673272, MF673273, MF673274, MF673277, MF673278, MF673279, MF673280, MF673281, MF673282, MF673283, MF673285, MF673290, MF673275, MF673284, MF673286, MF673287, MF673288, MF673289, MF673291, MF673292, MF673276, EPI1181072, EPI1058585, EPI1058593
Influenza A (H3N2) from the study (n = 15)	MH410500, MH410501, MH410502, MH410503, MH410504, MH410505, MH410506, MH410507, MH410508, MH410509, MH410510, MH410511, MH410512, MH410513, MH410514

correlation between Ct values and number of iSNVs was determined by Pearson's correlation method. The total number of iSNVs between vaccinated and unvaccinated groups was compared as well as the total number of iSNVs between 2013 and 2014 seasons. As prior studies have demonstrated that viral load influences sensitivity and specificity of iSNV detection (McCrone and Lauring, 2016), Ct values of virus specimens from vaccinated and unvaccinated individuals (set 1) and of the 2013 and 2014 seasons (set 2) were compared using an independent-samples *t*-test. Then subsequent within-set comparisons of the total numbers of iSNVs for the entire influenza genome and for each of the eight genome segments were conducted on set 2. Independent-samples *t*-test was used when the numbers of iSNVs (total genome or an individual segment) for the two groups were normally distributed, otherwise Mann-Whitney test was applied. A p-value <0.05 is considered statistically significant.

RESULTS

Phylogenetic analysis

A phylogenetic tree constructed from 587 influenza A(H3N2) HA1 gene sequences revealed that the 15 [A/H3N2] viruses from the 2013 and 2014 seasons belonged to clade 3C.2 (5/15) and 3C.2a (10/15), respectively, whereas vaccine strains for 2013 (A/Victoria/361/2011) and 2014 (A/Texas/50/2012) belonged to clade 3C.1 (Fig 2A). In the 2014 season, within clade 3C.2a, the four sequences from fully vaccinated children were not clustered together but scattered among other sequences in the same clade, as was also observed for sequences from unvaccinated children. As there was only one sequence from a fully vaccinated child in 2013, such analysis could not be undertaken. Percentage of each influenza A(H3N2) clade identified in Thailand during 2011 to 2017 was shown in Fig 2B. It was observed that the influenza A(H3N2) clade 3C.1 viruses predominated in 2011, representing 80% (41/51) of all influenza A(H3N2) viruses sequenced in this year. Influenza A(H3N2) clade 3C.2 viruses emerged in 2012 (3/22, 14%) and gained dominance in 2013 (52/58, 90%). Influenza A(H3N2) clade 3C.2a viruses emerged in 2013 (1/58, 2%) and gained dominance in 2015 (93/96, 97%). Influenza A(H3N2) clade 3C.2a1 viruses emerged in 2015 (1/96, 1%) and gained dominance in 2017 (33/40, 83%).

Amino acid substitutions in HA subunits

When the amino acid sequences in the HA subunits of the 15 viruses were compared with those from vaccine strains in the corresponding season, viruses from the 2013 season (clade 3C.2) contained 6 amino acid substitutions in the antigenic site (epitopes A to D) and an amino acid substitution in the non-antigenic site, while viruses from the 2014 season (clade 3C.2a) contained 10 amino acid substitutions in the antigenic site (epitope A to D) and an amino acid substitution at the non-antigenic site (Table 3). In the HA2 subunit, viruses from both seasons compared to the vaccine strains contained D160N substitution (Table 4). There was no unique amino acid differences in amino acid sequences of the HA subunits from vaccinated compared to unvaccinated children in both seasons.

Glycosylation sites

Eleven potential glycosylation sites in influenza A(H3N2) clade 3C.2 HA1 were identified at amino acid positions 8, 22, 38, 45, 63, 126, 133, 144, 165, 246, and 285 (data not shown). Loss of N122 glycosylation site makes the glycosylation pattern dif-

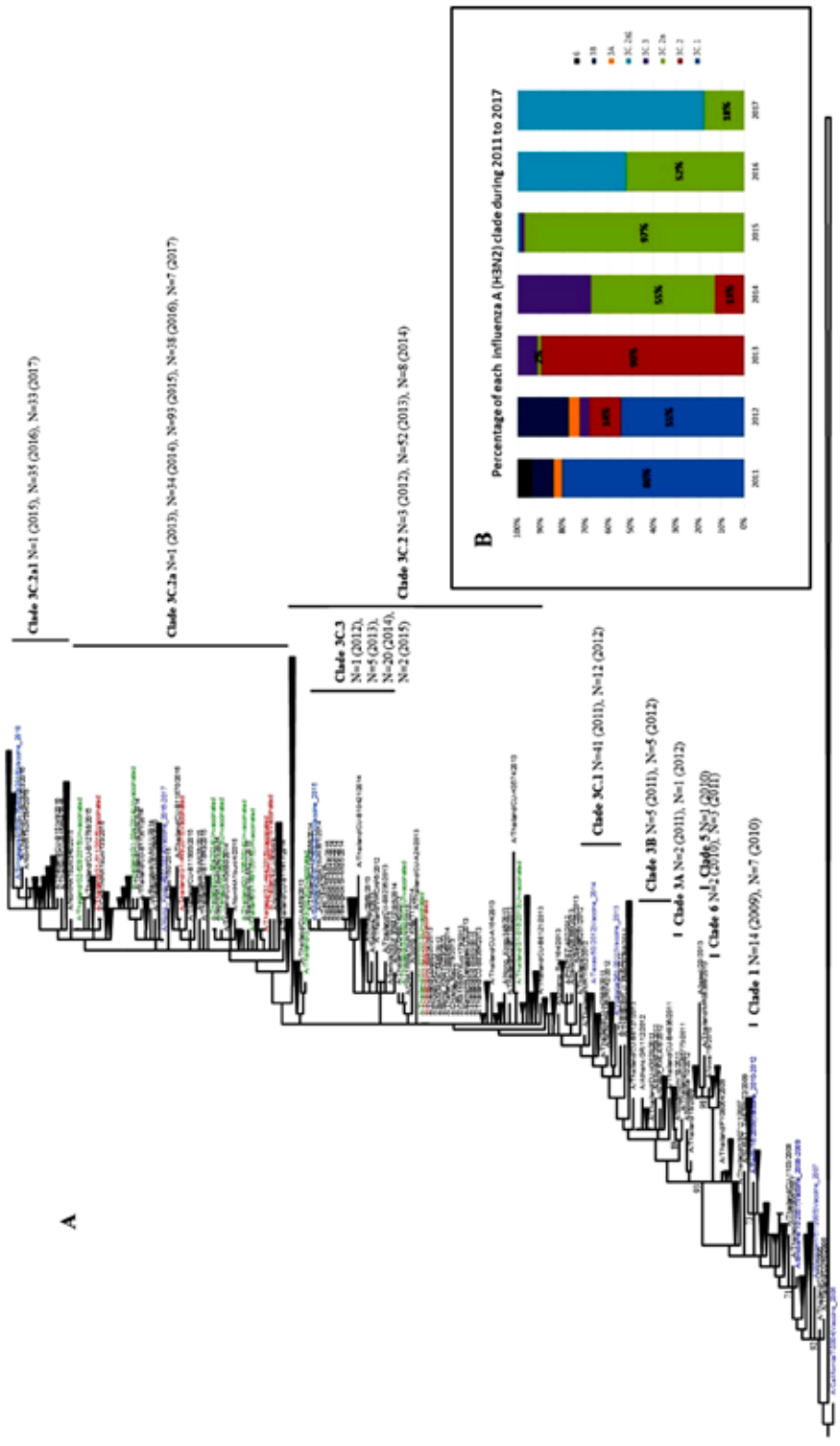

Fig 2- Phylogenetic tree of 587 hemagglutinin domain 1 (HA1) nucleotide sequences of influenza A(H3N2) viruses including 15 strains from the study (A), and percent each influenza A(H3N2) clade identified in Thailand during 2011 - 2017. A). Bootstrap values >60% are labeled at branch nodes. B). Calculated from numbers of HA1 sequences. Green, vaccinated; red, unvaccinated; blue, vaccine strains recommended by the World Health Organization; black, strains from GenBank and GISAID. GenBank accession numbers are listed in Table 3. Scale bar represents 2% nucleotide difference between close relations.

ferent from that of A/Victoria/361/2011, the 2013 vaccine strain (Stucker *et al*, 2015); whereas, clade 3C.2a virus also contained 11 potential glycosylation sites at the same positions as in clade 3C.2 except for S144 (in epitope A) and T160 (in epitope B) that caused loss in potential glycosylation sites, and a new potential glycosylation site at position158 (in epitope B), resulting in a glycosylation pattern different from that of the 2014 vaccine strain A/Texas/50/2012 (Stucker *et al*, 2015) (data not shown).

Characterization of iSNVs

A total of 132 iSNVs were identified across the entire genome of the 15 influenza A (H3N2) viruses, 36 synonymous and 96 non-synonymous mutations (Table 5). Each virus sample contained 4-15 iSNVs, with 121/132 (92%) variants present at a frequency of <0.1 and only 11 (8%) present at frequencies >0.1, latter found on PB1 (2 variants at frequency of 0.106 and 0.108), PA (4 variants at frequency of 0.167-0.259), HA (1 variant at frequency of 0.408), and NP (1 variant at frequency of 0.499) (Fig 3). The 3 variants with frequency >0.2 were found only in strains from 2014 season, namely, variants on PA (C708T), HA (T703C), and NP (T943C). Among these, only 1 variant on the HA protein of sample G1-511 (a sample with clade 3C.2a virus from a vaccinated child) contained a non-synonymous mutation (HA-S219P) located in epitope D.

In addition, there were 10 low frequency variants, 7 at HA1 antigenic sites and 3 at HA2 subunit, the former (5 non-synonymous and 2 synonymous mutations) consisting of 5 variants in epitope D (A563T, G563T, A572T, A581T, and A761G), 1 in epitope B (T606A), and 1 in epitope C (C960A). The 5 non-synonymous mutations were found in samples HA1-E172V, HA1-G172V, HA1-D175V, HA1-Y178F, and HA1-K238R. For the 3 low frequency variants in HA2 subunit, all non-synonymous mutations were T1088G (HA2-V18G), T1099G (HA2-Y21D) and G1558A (HA2-G174R). From both seasons, 4 variants (frequency of 0.020-0.046) were found in almost all viruses, in HA (A572T), NA (T1013A and T1020A) and PA (A1878T), with non-synonymous mutation in HA-D191V (HA1-175 on antigenic site D), NA-L338stop and PA-K626N.

Overall, Ct values showed normal distribution and inversely correlated with the numbers of iSNVs (correlation coefficient of -0.67, $p = 0.008$) (data not shown), indicating low Ct value reflected a high number of iSNVs. As the Ct values from vaccinated and unvaccinated groups were different (independent-samples *t*-test, $p =$ 1.2 E-05), comparison of the numbers of iSNVs between these two groups could not be analyzed. Comparison of the total number of iSNVs across the entire genome showed total number of iSNVs per sample from the same season is not significantly different (Mann-Whitney *U* test, $p = 0.54$). Total number of iSNVs in each genome segment is also not statistically significantly different (Mann-Whitney *U* test, $p = 0.07$, 0.12, 0.24, 0.81, 0.81, and 0.95 for PB2, PB1, NA, M, NS, and NP, respectively and independent-samples *t*-test, $p = 0.05$ and 0.63 for HA and PA, respectively) (data not shown).

DISCUSSION

Phylogenetic analysis demonstrates influenza A(H3N2) clade 3C.1 viruses predominated in 2011, representing 80% (41/51) of all influenza A(H3N2) viruses sequenced in that year. Influenza A(H3N2) clade 3C.2 viruses emerged in 2012 (3/22, 14%) and gained dominance

Table 4
Amino acid substitutions found in hemagglutinin domain 1 (HA1) subunits and domain 2 (HA2) of influenza A(H3N2) strains compared to vaccine strain, Bangkok, Thailand.

Protein subunits	HA1																				HA1				HA2					
Antigenic site	E	E	B	A	A	A	B	B	B	B	B	B	B	B	D	D	E	C	C	C	Non-antigenic site				Non-antigenic site					
Amino acid position	88	94	128	142	144	145	156	159	160	172	186	190	197	198	219	225	261	278	304	311	3	33	114	49	82	160	184	186	199	202
2013 Season																														
A/Victoria/361/2011	V	Y	T	R	N	N	R	F	K	E	V	E	Q	S	Y	N	R	N	A	Q	L	Q	S	N	K	D	D	I	C	L
G2-069 (Vaccinated)	.	.	.	.	.	S	H	.	.	.	G	D	.	.	S	.	.	K	.	.	.	R	.	.	.	N	.	.	.	.
G2-030 (Unvaccinated)	.	.	.	.	.	S	H	.	.	.	G	D	.	.	S	.	.	K	.	.	.	R	.	.	R	N	.	.	.	.
G1-016 (Unvaccinated)	.	.	.	.	.	S	H	.	.	.	G	D	.	.	S	.	.	K	.	.	.	R	.	.	.	N	.	.	.	.
G1-027 (Unvaccinated)	.	.	.	.	.	S	H	.	.	.	G	D	.	.	S	.	.	K	.	.	.	R	.	.	.	N	.	V	.	.
G1-043 (Unvaccinated)	.	H	.	.	.	S	H	.	.	.	G	D	.	.	S	.	Q	K	.	.	.	R	.	S	.	N	.	.	.	.
2014 Season																														
A/Texas/50/2012	.	.	N	.	.	.	H	.	.	.	.	D	.	P	F	.	.	K	.	.	.	R	.	.	.	.	.	.	.	.
G1-444 (Vaccinated)	.	.	.	.	S	S	H	Y	T	.	G	D	.	.	S	D	.	K	.	H	I	R	.	.	.	N	.	.	.	S
G1-449 (Vaccinated)	.	.	.	K	S	S	H	Y	T	G	G	D	R	.	S	D	.	K	.	H	I	R	.	.	.	N	.	.	.	.
G1-480 (Vaccinated)	.	.	.	.	S	S	H	Y	T	.	G	D	.	.	S	D	.	K	.	H	I	R	.	.	.	N	.	.	.	.
G1-511 (Vaccinated)	.	.	.	.	S	S	H	Y	T	.	G	D	.	.	S	D	.	K	.	H	I	R	T	.	.	N	.	.	.	.
G1-335 (Unvaccinated)	.	.	.	.	S	S	H	Y	T	.	G	D	.	.	S	D	.	K	.	H	I	R	.	.	.	N	.	.	.	.
G2-564 (Unvaccinated)	.	.	.	.	S	S	H	Y	T	.	G	D	.	.	S	D	.	K	.	H	I	R	.	.	.	N	.	.	.	.
G1-334 (Unvaccinated)	.	.	.	.	S	S	H	Y	T	.	G	D	.	P	S	D	.	K	P	H	I	R	.	.	.	N	.	.	.	.
G2-628 (Unvaccinated)	.	.	.	.	S	S	H	Y	T	.	G	D	.	.	S	D	.	K	.	H	I	R	T	.	.	N	.	.	.	.
G1-491 (Unvaccinated)	.	.	.	.	S	S	H	Y	T	.	G	D	K	.	S	D	.	K	.	H	I	R	.	.	.	N	.	.	S	.
G2-693 (Unvaccinated)	I	.	.	.	S	S	H	Y	T	.	G	D	.	.	S	D	.	K	.	H	I	R	.	.	.	N	N	.	.	.

Table 5
Numbers of synonymous and non-synonymous mutations in genome of influenza A(H3N2) strains, Bangkok, Thailand.

Sample	Sample ID	Vaccinated	Season	Number of iSNVs	Number of non-synonymous mutations	Number of synonymous mutations
1	G2-069	YES	2013	4	3	1
2	G2-030	NO	2013	12	10	2
3	G1-016	NO	2013	9	8	1
4	G1-027	NO	2013	11	9	2
5	G1-043	NO	2013	13	10	3
6	G1-444	YES	2014	6	3	3
7	G1-449	YES	2014	6	4	2
8	G1-480	YES	2014	6	5	1
9	G1-511	YES	2014	10	5	5
10	G1-335	NO	2014	15	12	3
11	G2-564	NO	2014	12	9	3
12	G1-334	NO	2014	6	5	1
13	G2-628	NO	2014	12	7	5
14	G1-491	NO	2014	6	3	3
15	G2-693	NO	2014	4	3	1

iSNVs, intra-host single nucleotide variants.

in 2013 (52/58, 90%). Influenza A(H3N2) clade 3C.2a viruses emerged in 2013 (1/58, 2%) and gained predominance in 2015 (93/96, 97%), while influenza A(H3N2) clade 3C.2a1 viruses emerged in 2015 (1/96, 1%) and gained predominance in 2017 (33/40, 83%). The circulation of clades 3C.2 and 3C.2a from the study was similar to that reported in other studies (D'Mello *et al*, 2015; Tewawong *et al*, 2015; Wedde *et al*, 2015; Monamele *et al*, 2017; Yokoyama *et al*, 2017).

This report describes for the first time in Thailand genetic characteristics of influenza A(H3N2) viruses identified from vaccinated and unvaccinated children (albeit a small number) from a previous VE study of the 2013 and 2014 flu seasons (Kittikraisak *et al*, 2016). The results reveal HA1 sequences at antigenic sites of clade 3C.2 and 3C.2a (the predominant clade in the 2013 and 2014 season, respectively) were different from the clade 3C.1 in the vaccine strains for 2013 (A/Victoria/361/2011) and 2014 (A/Texas/50/2012). Both fully vaccinated and unvaccinated children were infected with viruses circulated in the corresponding season. Although not all viruses from fully vaccinated children in the VE study were investigated, our findings imply that differences in circulating influenza A viruses from vaccine strain

Fig 3-Number and frequency of intra-host single nucleotide variants (iSNVs) among influenza A viruses from 2013 (blue) and 2014 (red) flu seasons, Bangkok, Thailand.

might account for the breakthrough infections; however, further studies need to be conducted to verify these speculations.

Based on sequence data alone, low VE might have been expected in both seasons as the vaccine strains employed were from a different clade from the circulating viruses. However, in this population VE in 2013 is 73% (95% CI: -14-94%) when clade 3C.2 is the predominant clade and A/Victoria/361/2011 clade 3C.1 is the vaccine strain (Kittikraisak *et al*, 2016). Similar findings were observed in Japan (1992 - 1993 flu season), USA (2004 - 2005 season) and Thailand (2011 - 2012 season), with VE of 55-68% during the circulation of antigenically dissimilar strains (Sugaya *et al*, 1994; Ohmit *et al*, 2006; Kittikraisak *et al*, 2015). Apparently the 6 mutated antigenic sites in HA1 epitopes A to D and 11 potential glycosylation sites found in the circulating clade 3C.2 viruses did not cause a significant antigenic difference from vaccine strain. A recent study in the USA demonstrated adults vaccinated with egg-adapted A/Victoria/361/2011-IVR-165 vaccine strain produce antibodies recognizing both the vaccine strain and viruses from the dominant wild type clades, 3C.2 and 3C.3 circulating during 2012-2013 flu seasons (Cobey *et al*, 2018), which lends support to the notion that vaccine-induced antibodies can cross-react with other strains not expressing the vaccine unique epitopes.

Influenza A(H3N2)-specific VE estimate of the 2014 flu season is 6% (95% CI: -103-56%) when predominately clade 3C.2a viruses were circulate and A/Texas/50/2012 (clade 3C.1) was the vaccine strain (Kittikraisak *et al*, 2016), consistent with VE estimates from Canada, UK and USA, which showed low or negligible protection against the circulating clade 3C.2a viruses during the 2014-2015 influenza season (Lednicky *et al*, 2016; Skowronski *et al*, 2016). The mutations at 9 antigenic site in HA1 epitopes A to D of clade 3C.2a viruses discovered in current study are similar to those reported in other studies (Lednicky *et al*, 2016; Skowronski *et al*, 2016; Huang *et al*, 2017; Monamele *et al*, 2017; Yokoyama *et al*, 2017; Zost *et al*, 2017). The unique amino acids (T128, S144, Y159, and T160) of HA protein in clade 3C.2a viruses were suggested to contribute to making HA structure more advantageous to evade pre-existing antibodies as well as increase ligand binding specificity (Yokoyama *et al*, 2017).

Although, the number of potential glycosylation sites on clade 3C.2a viruses was equal to that of clade 3C.2, a glycosylation site migration (change in location) from epitope A to epitope B in clade 3C.2a viruses might be critical in affecting VE. A study of pre-2009 influenza A(H1N1) viruses suggested glycosylation site migrations on HA and NA proteins have at least five possible functions, namely, more effectively mask antigenic sites, more effectively protect enzymatic cleavage sites of NA, help stabilize the polymeric structures, regulate receptor binding and catalytic activities, and balance HA binding activity with release activity of NA (Sun *et al*, 2012).

A comparison between clade 3C.2 and 3C.2a viruses reveals mutations to 3 charged amino acids in epitope B (K160T), epitope C (Q311H) and epitope D (N225D) are found only in the former viruses, which might drastically alter HA antigenicity and receptor binding affinity (Kobayashi and Suzuki, 2012; Li *et al*, 2013; Huang *et al*, 2017). Further investigations on the effects of charged amino acid mutations and antigenicity in clade 3C.2a viruses on VE are needed. However, it is also possible that low VE in 2014 season might have been

due to low vaccine immunogenicity in the study population (Lee *et al*, 2016; Petrie *et al*, 2016; Cobey *et al*, 2018).

As regards intra-host diversity, Debbink *et al* (2017) noted seasonal vaccination from 2004 to 2008 in healthy adults 18-49 years of age in the United States has minimal impact on the intra-host diversity of influenza A(H3N2) virus population. Although fewer samples were examined in the current study, similar results were observed in viruses from young Thai children during the 2013 and 2014 influenza seasons. However, our analysis was based on data from circulating viruses that were dissimilar to the corresponding vaccine, showing a greater reduction in HA genetic diversity of viruses from children who received 3C.1 clade vaccine. A variant with high frequency (0.408) found in HA gene of only 1 sample in clade 3C.2a virus from a vaccinated person might be construed as arising from vaccine-induced immunity driving a strong selective pressure towards novel variants in that individual. Thus, vaccination paradoxically may help promote antigenic drift-related variants and high mutation rates of HA gene (Cattoli *et al*, 2011); however, this notion needs more study. Comparison of intra-host diversity between viruses from the two seasons suggest host background immunity (3C.1 vaccine-induced immunity or pre-exiting host immunity) in each season has a low impact on intra-host diversity of clade 3C.2 and 3C.2a viruses. This also implies intra-host diversity was not associated with the difference in VE between 2013 to 2014 influenza seasons as there was no change in the number of intra-host diversity in the virus population from both seasons and no appearance in both seasons of a specific variant related to critical mutations in the antigenic sites.

The major limitation of the study was the small number of samples (only 15 viruses). In addition, no data were available on viral load, measurement of the existing antibody titer against influenza viruses or a plasmid control for monitoring false positive iSNVs. Viral load can influence sensitivity and specificity of iSNV detection (McCrone and Lauring, 2016). In the current study, there was a negative correlation between Ct values (assumed to reflect in part the viral load in samples) and number of variants obtained from all examined samples, but the comparison was performed on only the groups with no Ct value differences. Existing antibody against influenza virus, including vaccine-induced and natural infection-induced antibodies, would have enhanced accuracy of the analysis, but such information regarding the test samples was not available. Although there was no plasmid control for monitoring false positive iSNV results, sequencing was performed directly with swab specimens without amplification by passage in cell culture and high fidelity enzyme was used to eliminate false positive iSNVs.

In summary, genetic characterization of influenza A(H3N2) viruses from a previous VE study in Thai children during 2013 and 2014 influenza seasons shows a predominance of clade 3C.2a viruses in 2014 that may have been a cause of the low vaccine effectiveness in the 2014 season owing to mutations related with glycosylation site migration of the HA protein. No association was observed between lower vaccine effectiveness and intra-host diversity among viruses from the two seasons. Continued investigations of genetic characterization of influenza viruses from vaccine effectiveness studies and in particular vaccine failures will provide more precise information on the effect of viral genetic factors on vaccine

effectiveness and support improvements in future vaccine development.

ACKNOWLEDGEMENTS

The authors thank the study coordinators and clinical laboratory administrative personnel at Thailand Ministry of Public Health - US Centers for Disease Control and Prevention Collaboration, the US Army Medical Directorate of the Armed Forces Research Institute of Medical Sciences and QSNICH, and for their cooperation the Thai children for their participation in the previous vaccine effectiveness study for their generous cooperation. The research was funded by a grant from the Armed Forces Health Surveillance Center-Global Emerging Infections Surveillance and Response System. The funders had no role in study design, data collection and analysis, preparation of the manuscript, or decision to publish. Material has been reviewed by the Walter Reed Army Institute of Research and there is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the authors, and are not to be construed as official, or as reflecting true views of the Department of the Army, the Department of Defense, or the Centers for Disease Control and Prevention. The investigators have adhered to the policies for protection of human subjects as prescribed in AR 70–25.

REFERENCES

Archetti I, Horsfall FL Jr. Persistent antigenic variation of influenza A visuses after incomplete neutralization in vovo with heterologous immune serum. *J Exp Med* 1950; 92: 441-62.

Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 2004; 4: 1633-49.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; 30: 2114-20.

Boni MF. Vaccination and antigenic drift in influenza. *Vaccine* 2008; 26: 8-14.

Boni MF, Gog JR, Andreasen V, Feldman MW. Epidemic dynamics and antigenic evolution in a single season of influenza A. *Proc R Soc Lond B Biol Sci* 2006; 273(1592): 1307-16.

Bouvier NM, Palese P. The biology of influenza viruses. *Vaccine* 2008; 26: D49-53.

Carrat F, Flahault A. Influenza vaccine: the challenge of antigenic drift. *Vaccine* 2007; 25: 6852-62.

Castilla J, Godoy P, Domínguez A, *et al*; CIBERESP Cases and Controls in Influenza Working Group Spain. Influenza vaccine effectiveness in preventing outpatient, inpatient, and severe cases of laboratory-confirmed influenza. *Clin Infect Dis* 2013; 57: 167-75.

Cattoli G, Fusaro A, Monne I, *et al*. Evidence for differing evolution dynamics of A/H5N1 viruses among countries applying or not applying avian influenza vaccination in poultry. *Vaccine* 2011; 29: 9368-75.

Cobey S, Gouma S, Parkhouse K, *et al*. Poor immunogenicity, not vaccine strain egg adaptation, may explain the low H3N2 influenza vaccine effectiveness in 2012-13. *Clin Infect Dis* 2018; 67: 327-33.

Debbink K, McCrone JT, Petrie JG, *et al*. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. *PLOS Pathog* 2017; 13: e1006194.

Dinis JM, Florek NW, Fatola OO, *et al*. Deep sequencing reveals potential antigenic variants at low frequencies in influenza A virus-infected humans. *J Virol* 2016; 90: 3355-65.

D'Mello T, Brammer L, Blanton L, *et al*. Update: influenza activity-United States, Septem-

ber 28, 2014-February 21, 2015. *MMWR Morb Mortal Wkly Rep* 2015; 64: 206-12.

Fonseca NA, Rung J, Brazma A, Marioni JC. Tools for mapping high-throughput sequencing data. *Bioinformatics* 2012; 28: 3169-77.

Grabherr MG, Haas BJ, Yassour M, *et al*. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* 2011; 29: 644-52.

Gupta R, Jung E, Brunak S. Prediction of N-glycosylation sites in human proteins. Database: NetNGlyc 1.0, 2004. [Cited 2018 Oct 10]. Available from: http://www.cbs.dtu.dk/services/NetNGlyc/

Huang ZZ, Yu L, Huang P, Liang LJ, Guo Q. Charged amino acid variability related to N-glyco -sylation and epitopes in A/H3N2 influenza: Hem-agglutinin and neuraminidase. *PLOS One* 2017; 12: e0178231.

Iuliano AD, Roguski KM, Chang HH, *et al*. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet* 2018; 391(10127): 1285-300.

Kittikraisak W, Suntarattiwong P, Ditsungnoen D, *et al*. Effectiveness of the 2013 and 2014 southern hemisphere influenza vaccines against laboratory-confirmed influenza in young children using a test-negative design, Bangkok, Thailand. *Pediatr Infect Dis J* 2016; 35: e318.

Kittikraisak W, Suntarattiwong P, Levy J, *et al*. Influenza vaccination coverage and effectiveness in young children in Thailand, 2011-2013. *Influenza Other Respir Viruses* 2015; 9: 85-93.

Kobayashi Y, Suzuki Y. Compensatory evolution of net-charge in influenza A virus hemagglutinin. *PLOS One* 2012; 7: e40422.

Lednicky JA, Iovine NM, Brew J, *et al*. Hemagglutinin gene clade 3C.2a influenza A (H3N2) viruses, Alachua County, Florida, USA, 2014-15. *Emerg Infect Dis* 2016; 22: 121-3.

Lee J, Boutz DR, Chromikova V, *et al*. Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat Med* 2016; 22: 1456-64.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754-60.

Li H, Handsaker B, Wysoker A, *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; 25: 2078-9.

Li Y, Bostick DL, Sullivan CB, *et al*. Single hemagglutinin mutations that alter both antigenicity and receptor binding avidity influence influenza virus antigenic clustering. *J Virol* 2013; 87: 9904-10.

McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 2004; 32: W20-5.

McCrone JT, Lauring AS. Measurements of intrahost viral diversity are extremely sensitive to systematic errors in variant calling. *J Virol* 2016; 90: 6884-95.

McElroy K, Zagordi O, Bull R, Luciani F, Beerenwinkel N. Accurate single nucleotide variant detection in viral populations by combining probabilistic clustering with a statistical test of strand bias. *BMC Genomics* 2013; 14: 501-12.

Monamele GC, Vernet MA, Njankouo MR, *et al*. Genetic and antigenic characterization of influenza A (H3N2) in Cameroon during the 2014-2016 influenza seasons. *PLOS One* 2017; 12: e0184411.

Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 2016; 32: 1749-51.

Ohmit SE, Victor JC, Rotthoff JR, *et al*. Prevention of antigenically drifted influenza by inactivated and live attenuated vaccines. *N Engl J Med* 2006; 355: 2513-22.

Petrie JG, Parkhouse K, Ohmit SE, Malosh RE, Monto AS, Hensley SE. Antibodies against the current influenza a (H1N1) vaccine strain do not protect some individuals from infection with contemporary circu-

lating influenza A (H1N1) virus strains. *J Infect Dis* 2016; 214:1947-51.

Rambaut A. FigTree v1. 4, 2012. [Cited 2019 Jan 14]. Available from: http://tree.bio.ed.ac.uk/software/figtree/

Rutvisuttinunt W, Chinnawirotpisan P, Thaisomboonsuk B, *et al*. Viral subpopulation diversity in influenza virus isolates compared to clinical specimens. *J Clin Virol* 2015;68: 16-23.

Shakin-Eshleman SH, Spitalnik SL, Kasturi L. The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J Biol Chem* 1996; 271: 6363-6.

Shen S, Campitelli MA, Calzavara A, Guttmann A, Kwong JC. Seasonal influenza vaccine effectiveness in pre- and full-term children aged 6-23 months over multiple seasons. *Vaccine* 2013; 31: 2974-8.

Skowronski DM, Sabaiduc S, Chambers C, *et al*. Mutations acquired during cell culture isolation may affect antigenic characterisation of influenza A (H3N2) clade 3C.2a viruses. *Euro Surveill* 2016; 21: 30112.

Stucker KM, Schobel SA, Olsen RJ, *et al*. Haemagglutinin mutations and glycosylation changes shaped the 2012/13 influenza A (H3N2) epidemic, Houston, Texas. *Euro surveill* 2015; 20: 21122.

Sugaya N, Nerome K, Ishida M, Matsumoto M, Mitamura K, Nirasawa M. Efficacy of inactivated vaccine in preventing antigenically drifted influenza type A and well-matched type B. *JAMA* 1994; 272: 1122-6.

Sun S, Wang Q, Zhao F, Chen W, Li Z. Prediction of biological functions on glycosylation site migrations in human influenza H1N1 viruses. *PLOS One* 2012; 7: e32119.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013; 30: 2725-9.

Tewawong N, Prachayangprecha S, Vichiwattana P, *et al*. Assessing antigenic rift of seasonal influenza A(H3N2) and A(H1N1)pdm09 viruses. *PLOS One* 2015; 10: e0139958.

Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of influenza A viruses. *Microbiol Rev* 1992; 56: 152-79.

Wedde M, Biere B, Wolff T, Schweiger B. Evolution of the hemagglutinin expressed by human influenza A (H1N1) pdm09 and A (H3N2) viruses circulating between 2008-2009 and 2013-2014 in Germany. *Int J Med Microbiol* 2015; 305: 762-75.

Wilm A, Aw PPK, Bertrand D, *et al*. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* 2012; 40: 11189-201.

Yokoyama M, Fujisaki S, Shirakura M, *et al*. Molecular dynamics simulation of the influenza A (H3N2) hemagglutinin trimer reveals the structural basis for adaptive evolution of the recent epidemic clade 3C.2a. *Front Microbiol* 2017; 8: 584.

Zhou B, Donnelly ME, Scholes DT, *et al*. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza A viruses. *J Virol* 2009; 83: 10309-13.

Zhu W, Wang C, Wang BZ. From variation of influenza viral proteins to vaccine development. *Int J Mol Sci* 2017; 18: E1554.

Zost SJ, Parkhouse K, Gumina ME, *et al*. 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; 114:12578-83.