

RESEARCH ARTICLE

Molecular Evidence of Transmission of Influenza A/H1N1 2009 on a University Campus

Ramandeep Kaur Virk^{1*}, Vithiagaran Gunalan², Hong Kai Lee³, Masafumi Inoue⁴, Catherine Chua⁵, Boon-Huan Tan⁶, Paul Anantharajah Tambyah¹

1 Department of Medicine, National University of Singapore, Singapore, **2** Bioinformatics Institute, Agency for Science, Technology and Research, Singapore, **3** Department of Laboratory Medicine, National University Hospital, Singapore, **4** Experimental Therapeutics Unit, Agency for Science, Technology and Research, Singapore, **5** University Health Centre, National University of Singapore, Singapore, **6** Detection & Diagnostics Laboratory, DSO National Laboratories, Singapore

* drramandeepvirk@yahoo.com.sg



OPEN ACCESS

Citation: Virk RK, Gunalan V, Lee HK, Inoue M, Chua C, Tan B-H, et al. (2017) Molecular Evidence of Transmission of Influenza A/H1N1 2009 on a University Campus. PLoS ONE 12(1): e0168596. doi:10.1371/journal.pone.0168596

Editor: Patrick Tang, Sidra Medical and Research Center, QATAR

Received: August 7, 2016

Accepted: December 3, 2016

Published: January 6, 2017

Copyright: © 2017 Virk et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All sequenced files are available from the GenBank database accession numbers: HA protein: AGY41900-AGY41903, AGY41905, AGY41906, AGY41910, AGY41911, AGY41915, AGY41918- AGY41921, AGY41923, AGY41924, AGY41927- AGY41932, AGY41935, AGY41936, AGY41907, AGY41912, AGY41917, AGY41913, AGY41914, AGY41926, AGY41904, AGY41916), (AGY41922, AGY41925, AGY41933, AGY41934, AGY41937-41, AGU69918, AGU69919, AGU69920, AGU69921, AGU69922, AGU69923, AHB72781, AHB72782, AGU69931, AGU69932, AGU69930. PB2: KP222548- KP222581; PB1:

Abstract

Background

In the recent years, the data on the molecular epidemiology of influenza viruses have expanded enormously because of the availability of cutting-edge sequencing technologies. However, much of the information is from the temperate regions with few studies from tropical regions such as South-east Asia. Despite the fact that influenza has been known to transmit rapidly within semi-closed communities, such as military camps and educational institutions, data are limited from these communities.

Objectives

To determine the phylogeography of influenza viruses on a university campus, we examined the spatial distribution of influenza virus on the National University of Singapore (NUS) campus.

Methods

Consenting students from the NUS who sought medical attention at the UHC provided two nasopharyngeal swabs and demographic data. PCR was used for detection of influenza viruses. 34 full-genomes of pH1N1/09 viruses were successfully sequenced by Sanger method and concatenated using Geneious R7. Phylogenetic analysis was conducted using these 34 sequences and 1518 global sequences. Phylogeographic analysis was done using BaTS software and Association index and Fitch parsimony scores were determined.

Results

Integrating whole genome sequencing data with epidemiological data, we found strong evidence of influenza transmission on campus as isolates from students residing on-campus

KP222582- KP222615; PA: KP222616- KP222649; HA: KF667880- KF667887, KF667890- KF667894, KF667896- KF667898, KF667900, KF667902- KF667910, KF667912, KF667914, KF667916- KF667921; NP: KP222650- KP222666, KF709407- KF709418, KF709420- KF709424; NA: KF695073- KF695080, KF695083- KF695101, KF695103- KF695109; MP: KP222667- KP222682, KP222684, KP222686- KP222694, KP222696, KP222698, KP222700- KP222705; NS: KP222706- KP222739.

Funding: The work at NUS was in part supported and funded by the Infectious Diseases Research Fund and NUSMINDEF JPP grant. The work at DSO National Laboratories was funded by the Ministry of Defence, Singapore. The work at the Institute of Molecular and Cell Biology (IMCB) was in part supported and funded by the Biomedical Research Council, which is part of Singapore's Agency for Science, Technology, and Research (A*STAR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

Competing Interests: Dr. Paul Tambyah has received research support from Baxter, GSK, ADAMAS, Merlion Pharmaceuticals, Sanofi Pasteur, Fabentech and Inviragen. He has also received honoraria from Novartis and Astra-Zeneca. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

were highly similar to each other (AI, P value = 0.009; PS, P value = 0.04). There was also evidence of multiple introductions from the community.

Conclusions

Such data are useful in formulating pandemic preparedness plans which can use these communities as sentinel sites for detection and monitoring of emerging respiratory viral infections.

Introduction

Molecular data have been instrumental in studying the origin as well as persistence and migration of influenza viruses [1,2]. In addition, molecular studies provide detailed virus characteristics, such as type, subtype, lineage, virulence and drug resistance markers. They also provide insights into the pathways of viral transmission and help in understanding the phylogeography of influenza viruses when integrated with epidemiological data [3]. A better integration of molecular and epidemiological data may (i) help unfold the dynamics of spatial transmission, (ii) provide insights into the global transmission patterns of influenza viruses and the existence of source populations, and (iii) help public health authorities to plan more targeted intervention strategies. On a global level, in an event of impending pandemic, a better understanding of transmission patterns of novel or emerging viruses may help public health authorities to develop effective pandemic preparedness plans.

In recent years, molecular epidemiology has been applied to emerging influenza viruses, including pandemic H1N1 2009 (pH1N1/09) and H7N9 [4]. Although there are available data on the spatial and temporal transmission dynamics of influenza at global and national levels [5–10], transmission dynamics have not been thoroughly explored at the level of individual communities, especially in semi-closed settings such as universities or military camps. A better understanding of transmission dynamics in localized communities helps to identify foci of influenza transmission for more targeted interventions, such as school closure or household prophylaxis [11]. An examination of the spatial diffusion of pH1N1/09 virus in a localized community of the University of California found that although there was intra-campus transmission, there was no significant association of on-campus residence with clustering of similar viral strains [12].

We undertook this study to identify distinctive epidemiological factors associated with transmission of influenza on a university campus and to explore the molecular epidemiology of influenza in a university setting. We found that 79% (11/14) of the strains that formed clusters on phylogenetic analysis were from students staying on-campus ($p < 0.05$), which suggests that a significantly higher proportion of the influenza was transmitted within campus than was introduced from sources outside of the university. This has tremendous implications on response to future pandemics. For instance, if the majority of influenza is found to be transmitted on campus, i.e. there is substantial virus clustering on campus and the influenza strains isolated are proven by molecular epidemiology to be closely linked, then, possible closure of large institutions *or* on-campus intervention methods to halt influenza transmission, such as, quarantine of hostels, cancellation of lectures, shut-down of canteen areas, closure of facility areas or ring prophylaxis may be important and useful strategies. If on the other hand, the majority of influenza strains are distinct and there is no evidence of clustering, suggesting

importation from contacts off campus, then the major intervention strategies would be very different.

Materials and Methods

Samples

The study was approved by National University of Singapore (NUS) Institutional Review Board (reference number, 06–156; approval number, NUS-282) and written informed consent was obtained from the participants prior to sample and data collection. Two nasopharyngeal swabs and demographic data (age, gender, citizenship status, smoking status, residence address, and contact history) were collected from students and staff presenting with influenza-like illness (ILI) to the university health centre at the NUS [13]. Influenza infection was confirmed by PCR. Details of this cohort have been described previously [14].

Virus sequencing

RNA was extracted using QIAamp Viral RNA mini kit (Qiagen, Inc., Valencia, CA, USA) according to manufacturer's instruction. Reverse-transcription of vRNA was carried out using SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen Corporation, CA, USA) and Uni 12 primer (5' AGCRAAAGCAGG3') [15]. Full-length genes were obtained by performing PCR using primers described previously [15–17], and with primers obtained from the WHO Collaborating Centre for Reference and Research on Influenza in Melbourne (Dr. TBoon-Huan Tan, personal communication). Sanger sequencing was performed with ABI Prism Big Dye Terminator (ThermoFisher Scientific Inc.). The assembly and editing of raw sequence data were done using SeqMan (DNASTAR, Lasergene Version 7, Madison, USA). The sequences were deposited in GenBank and can be retrieved from the National Center for Biotechnology Information's (NCBI) Influenza Virus Resource (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>).

Preliminary analysis for related transmissions based on HA sequence and epidemiology

Eleven hemagglutinin (HA) genes of seasonal (sH1N1 and H3N2) and 40 HA genes of pH1N1/09 influenza A viruses (IAVs) were included in this analysis. The fifty-one sequences of HA gene were translated into protein sequences and were then divided into two groups—“shared” strains versus “non-shared” strains which represented clustered transmission groups versus isolated introductions respectively. The criteria for a shared strain was 100% amino acid (aa) identity in two or more strains with date of collection within 5 days of each other, or from the preceding one in the cluster respectively. This would probably suggest that some kind of active transmission was going on. The two groups were then compared across various demographic characteristics such as age, gender, on-campus residence, course of study/faculty, and citizenship status.

In-depth genomic analysis of pH1N1/09 viruses using comparative phylogenetic methods

Analysis was conducted on full-genomes of 34 pH1N1/09 viruses successfully sequenced in our study.

Phylogenetic analysis. For the whole genome concatenation Geneious R7 <http://www.geneious.com> [18] was used. The coding regions of gene segments PB-2, PB-1, PA, HA, NP, NA, M1, M2, NS1 and NS2 were concatenated using the software. The sequences were tested

for any reassortment between strains using the RDP4 software [19]. Multiple alignment was performed with ClustalW in BioEdit [20]. The whole genomes were then used for the construction of phylogenetic trees using the best-fit nucleotide substitution model (TVM +Gamma+Invariant sites) determined by Modeltest 3.7 [21]. Bootstrapping with 1000 replicates was undertaken to assess the robustness of the phylogenetic trees. All the analyses were performed in the PAUP package [22].

Phylogeographic/ Bayesian Tip-Association (BaTS) analysis. BaTS (<http://evolve.zoo.ox.ac.uk/Evolve/BaTS.html>) was used to assess the degree to which the phylogeny of the viruses was related with the phenotypic traits [23]. The analysis was undertaken to determine the spatial dynamics of influenza infection in the localized university community of NUS. For the spatial component there were two categories: the location of the residence ('On-campus' or 'Off-campus') and faculty ('Life Sciences': Science, Nursing and Medicine faculties or 'Non-Life Sciences': Engineering, Business, Design and Environment, Computing, Arts and Social Science). The basis of separation into two faculties was the location on the campus. Life Sciences faculties are physically clustered in one region of the university whereas the Non-Life Sciences faculties are clustered together in altogether separate region of the university. The other phenotypic factors that were analysed for phylogenetic clustering were age (<25 or >25); gender (Male or Female); Singaporean (Yes or No). The BaTS program has a pre-requisite of Bayesian analysis as it uses the tree file with posterior probabilities that is generated in the BEAST v1.8 software [24]. The Bayesian analysis was performed using Bayesian Markov Chain Monte Carlo (MCMC) method with GTR+I+G substitution model along with estimated relaxed (uncorrelated lognormal) molecular clock. The analyses were performed with a time-aware Gaussian Markov Random Field Bayesian Skyride coalescent tree, with an Unweighted Pair Group Method with Arithmetic Mean-derived starting tree. All model parameters were given with relatively uninformative priori (default setting), except a uniform prior distribution for the mean substitution rate with initial value of 0.005 substitution per site per year and lower/upper limits of 0.0/1.0 [25, 26]. Analysis was performed with length of MCMC chain of 50 million to reach convergence and sampled at every 5000th generation. The analysis generated a total of 10000 samples for parameter estimates. Ten percent burn-in was applied for the analysis. This provided with a posterior set of trees which were then used to investigate each of the phenotypic traits listed earlier. Each trait was analysed as a binary character annotated into the posterior set of trees from the Beast analysis using the BaTS program. Two statistics (a) Association index (AI) and (b) Parsimony score (PS) were calculated to determine the phylogeny and trait association [27,28]. These two statistics determine the strength of phylogenetic clustering by place of isolation. A P value of ≤ 0.05 was considered statistically significant which means that the clustering was in fact associated with the trait under analysis and was not merely by chance.

Panoramic phylogenetic analysis. Further, these 34 genomes were analyzed in the context of 1518 global pH1N1/09 genomes retrieved from GenBank for the period from May 2009 to Sep 2009 using RAXML [29]. Trees were generated using the GTR substitution matrix with a GAMMA model of rate heterogeneity (all parameters were allowed to be estimated) and 1000 rapid bootstrap replicates were performed and examined with an ML search to find the best-scoring tree, which was subsequently visualized in MEGA 6 [30].

Accession numbers

"Shared" strains (GenBank accession numbers for HA protein)- cluster S1: AGY41900-AGY41903, AGY41905, AGY41906, AGY41910, AGY41911, AGY41915, AGY41918- AGY41921, AGY41923, AGY41924, AGY41927- AGY41932, AGY41935, AGY41936, cluster S2: AGY41907, AGY41912, AGY41917, cluster S3: AGY41913, AGY41914, AGY41926, cluster S4: AGY

41904, AGY41916) and “non-shared” strains (AGY41922, AGY41925, AGY41933, AGY41934, AGY41937-41, AGU69918, AGU69919, AGU69920, AGU69921, AGU69922, AGU69923, AHB72781, AHB72782, AGU69931, AGU69932, AGU69930. GenBank nucleotide accession numbers of 34 pH1N1/09 strains sequenced in this study: **PB2**: KP222548- KP222581; **PB1**: KP222582- KP222615; **PA**: KP222616- KP222649; **HA**: KF667880- KF667887, KF667890- KF667894, KF667896- KF667898, KF667900, KF667902- KF667910, KF667912, KF667914, KF667916- KF667921; **NP**: KP222650- KP222666, KF709407- KF709418, KF709420- KF709424; **NA**: KF695073- KF695080, KF695083- KF695101, KF695103- KF695109; **MP**: KP222667- KP222682, KP222684, KP222686- KP222694, KP222696, KP222698, KP222700- KP222705; **NS**: KP222706- KP222739.

Results

Preliminary analysis for related transmissions based on HA sequence and epidemiology

In the first analysis, using HA protein sequences and our own criteria for categorization of IAVs into shared and non-shared strains, we identified 31 ‘shared strains’ and 20 ‘non-shared strains’. This suggests at least some degree of transmission on-campus together with a limited amount of introductions. The 31 strains in the shared group were all pH1N1/09 strains which could be further sub-divided into 4 clusters with 100% identical aa patterns. One of the clusters had 23 strains, two clusters had three strains each, and another cluster had two strains. The non-shared group had 9 distinct pH1N1/09 strains and 11 distinct seasonal influenza strains (sH1N1 and H3N2). Residence at a student hostel was identified as a risk factor for having a shared strain (Table 1).

In-depth genomic analysis of pH1N1/09 viruses using comparative phylogenetic methods

To ensure that our phylogenetic analysis of concatenated genomes is valid, we performed analysis for reassortment using RDP4 software and found no evidence of reassortment. We required the reassortment event to be consistently identified by at least three of the seven algorithms used in the RDP4 software. The second analysis which is phylogenetic analysis of full genomes of 34 pH1N1/09 viruses isolated in our study showed that both clade 6 and 7 were co-circulating on campus during the first pandemic wave in Singapore from July to August at NUS campus (Fig 1). It is possible that all the earlier clades might have been displaced by clade 6 and clade 7 viruses. From the phylogenetic tree (Fig 1), five well-supported clusters (bootstrap value >70) of pH1N1/09 viruses from the university students and staff can be discerned. This clustering suggests that there was some degree of intra-university/on-campus transmission going on. The clusters were assigned names from A to E as shown in Fig 1. All the clusters

Table 1. Association of epidemiological factors with clustered strains.

Epidemiological factor	Odds Ratio (95% CI)	P value
On-campus vs Off-campus residence	4.2 (1.2–14.9)	0.02
Age <25 vs Age >25 years	2.3 (0.6–8.9)	0.22
Gender	0.8 (0.2–2.7)	0.74
Singaporean vs Foreigner	1.7 (0.5–5.2)	0.38
Life Sciences vs Non-life sciences faculty	1.7 (0.5–5.9)	0.41

P value of ≤0.05 was statistically significant (shown in bold)

doi:10.1371/journal.pone.0168596.t001

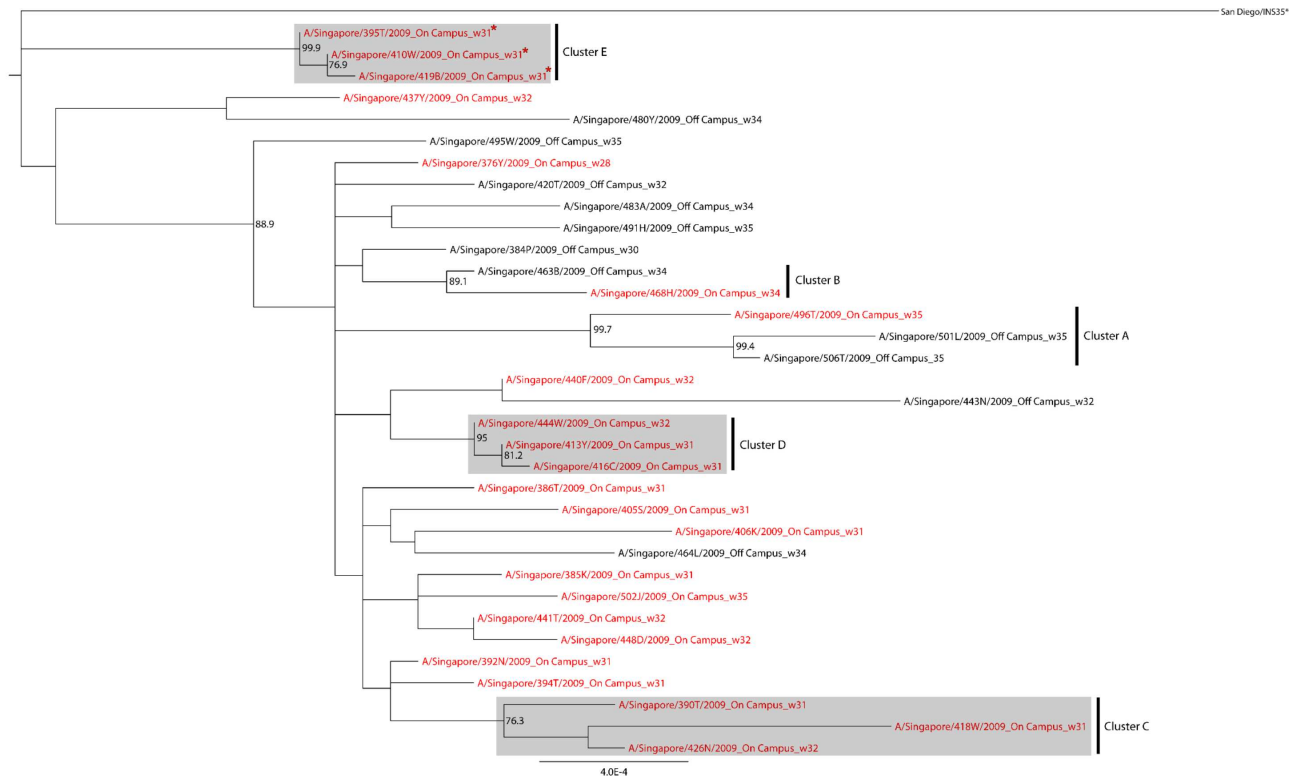


Fig 1. Maximum Likelihood phylogenetic tree of 34 concatenated genomes of pH1N1/09 viruses from NUS campus. Clade 6 viruses are marked with asterisk. Strain name is followed by residence status and week of isolation. On-campus sequences are in red font and Off-campus sequences are in black font. Clusters were identified with strong bootstrap support (>70%). Clusters with exclusively On-campus sequences are highlighted in grey color.

doi:10.1371/journal.pone.0168596.g001

(A, C, D and E) comprised three virus strains each, except one cluster (B), which had two strains. Of the five clusters, only cluster E had the three clade 6 virus strains while the rest had clade 7 viruses (Fig 1). Apart from clustered viruses, there were singleton viruses which may suggest independent viral introduction events. The phylogeny of the pH1N1/09 viruses was subsequently correlated with the epidemiological characteristics. The location of residence was compared among the various strains in each cluster. In three of the clusters: C, D and E, all the member strains were isolated from students residing on campus. This suggests possible intra-university transmission of influenza. However, in the other two clusters: A and B, both on-campus and off-campus strains clustered together. Universities represent a semi-closed setting and are themselves a small community, where the student interaction/mixing may happen at various places like hostels, classrooms, canteens, or while using the various university facilities. A closer look into the clusters with the aid of integrated phylogenetic and epidemiological data provided more insights into the transmission dynamics on campus. Cluster D contained three strains from students who reported sick on the same day, were from the same hostel, and even the same faculty which provide very strong evidence of intra-campus transmission. Furthermore, A/Singapore/413Y/2009 strain was isolated from a student who reported contact history with a sick neighbour in Malaysia. It is quite possible that he acquired infection in Malaysia and then spread it to other fellow students on campus. However, in cluster E, the first strain (according to the date of collection), A/Singapore/395T/2009, was isolated from a student from non-life sciences faculty and from a different hostel than the other two strains, A/Singapore/410W/2009 and A/Singapore/419B/2009. These 2 strains were isolated from students

Table 2. Results of phylogeny trait association for pH1N1/09 viruses.

Demographic Characteristic	Statistic	Observed value (95% CI)	Null value (95% CI)	P value
On vs Off campus residence	AI	0.76 (0.47–1.08)	1.69 (1.14–2.23)	0.009
	PS	7.08 (6.0–8.0)	8.90 (7.10–9.93)	0.04
Age <25 vs >25	AI	0.68 (0.42–0.96)	0.80 (0.48–1.2)	0.34
	PS	3.88 (3.0–4.0)	3.89 (3.06–4.0)	1.0
Male vs Female	AI	1.61 (1.06–1.9)	1.47 (1.06–1.99)	0.71
	PS	6.68 (6.0–7.0)	7.38 (6.0–8.0)	0.25
Singaporean vs Foreigner	AI	1.78 (1.40–2.16)	1.88 (1.19–2.6)	0.45
	PS	11.04 (10.0–12.0)	11.18 (8.91–13.04)	0.43
Life Sciences vs Non-life Sciences faculty	AI	0.50 (0.27–0.82)	1.38 (0.91–1.96)	0.02
	PS	5.87 (5.0–7.0)	7.33 (6.0–8.0)	0.05

AI- Association index; PS-Parsimony score; P value ≤ 0.05 was statistically significant, Life Sciences- Medicine, Nursing, Sciences; Non-life Sciences- Engineering, Computing, Business, Arts and Social science, Design and Environment.

doi:10.1371/journal.pone.0168596.t002

from life sciences faculty, who shared the same hostel and reported sick on the same day. This suggests the student with A/Singapore/395T/2009 strain could have possibly infected the students with strains A/Singapore/410W/2009 or A/Singapore/419B/2009 on campus but at a location other than hostel and classroom, for instance, at library, canteen or while using hostel facilities, suggesting some kind of spatial mixing on campus. Interestingly, there were some students who reported sick on the same day or one day apart and were from the same hostel and even same faculty but did not share highly similar strains.

To further strengthen the analysis and to clear the ambiguity associated with phylogenetic analysis, another test of phylogeny-trait associations was conducted using the BaTS software. Two statistics, AI and PS scores, were calculated for various demographic characteristics to explore the strength of clustering (Table 2).

The phylogeography of the viruses was analysed employing two characteristics: residence location and faculty. The residence location was annotated as ‘On-campus’ or ‘Off-campus’ and faculty was annotated as Life Sciences or Non-Life Sciences. The results showed that there was in fact some degree of clustering according to residence and this was not merely due to chance (AI P = 0.009; PS P = 0.04). There was also evidence of clustering with relation to faculty (AI P = 0.02; PS P = 0.05).

Panoramic phylogenetic analysis

The third phylogenetic analysis showed (Fig 2) the estimated upper-bound of introductions on campus was 8 (NUS sequences separated from each other by non-NUS sequences as well as singleton lineages), and lower-bound of introductions was 4 (distinct lineages of pH1N1/09 viruses). This further implies that there was indeed influenza transmission going on on-campus.

Discussion

This study was undertaken on a university campus to understand the spatial dynamics of spread and transmission of IAVs. There are some outbreak studies in university populations [31–33], but few phylogeographic studies in the literature.

We conducted three separate analyses to study the transmission dynamics of IAVs. The univariate epidemiologic analysis was based on the percentage of aa identity for HA protein sequences: ‘Shared’ & ‘Non-shared’ strains. Shared strains outnumbered non-shared strains,

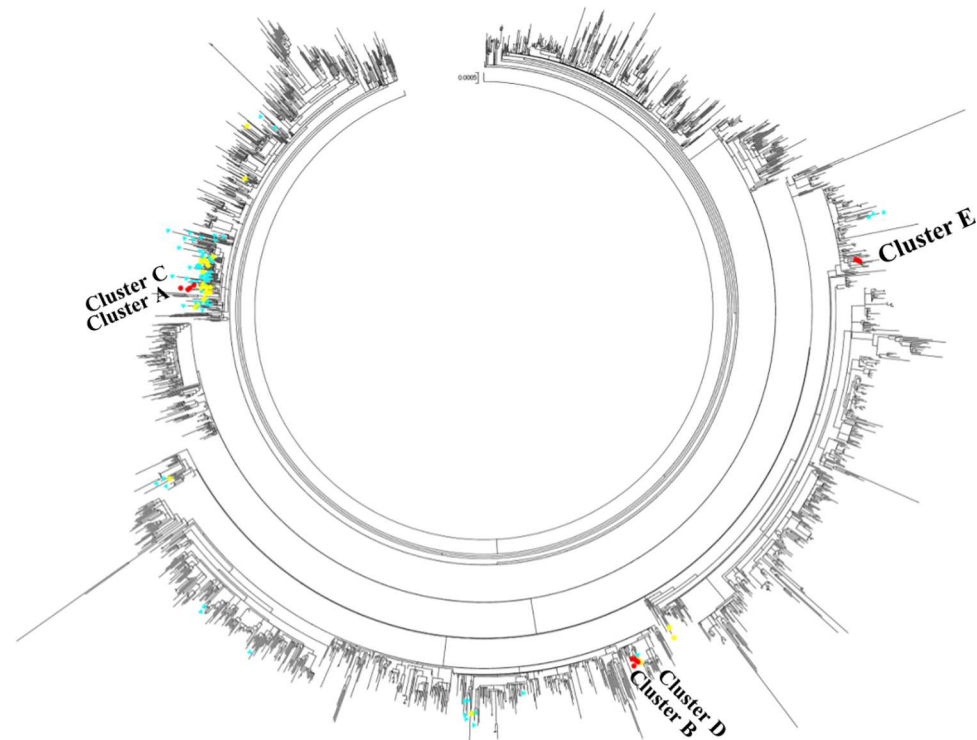


Fig 2. Maximum Likelihood phylogenetic tree of 1518 global genomes and 34 concatenated genomes of pH1N1/09 viruses. NUS sequences are in circles (red represent clusters and yellow others) and non-NUS Singaporean sequences are in blue triangles. Clusters were identified with strong bootstrap support (>70%). Trees were generated in RAxML using the GTR substitution matrix and GAMMA model of rate heterogeneity with 1000 bootstrap replicates. The best scoring tree was visualized in MEGA 6.

doi:10.1371/journal.pone.0168596.g002

and all the shared strains were pH1N1/09 viruses. This suggests that the pH1N1/09 viruses were more transmissible than seasonal IAVs and this could be due to the viruses captured in the early phase of pandemic in this study when the virus had not undergone significant genetic drift. Residence at hostel was identified as an independent risk factor for having a shared strain. Clustering of identical strains in subjects staying on campus in hostels suggests that influenza transmission occurred on campus. While planning control strategies for influenza epidemics or mitigation strategies for pandemics, this demographic risk factor needs to be taken into consideration. Our sample size was too small for a meaningful multivariable analysis with multiple comparisons so in order to identify transmission clusters we also performed the genomic and phylogeographic analysis to confirm the epidemiological analysis. For genomic analysis, concatenated whole genome approach (34 genomes of pH1N1/09 viruses) was undertaken to determine the molecular epidemiology of influenza. The concatenation provides more concrete phylogenetic evidence than the single gene approach in understanding the complex microevolution of IAVs since genetic drift occurs in internal genes as well due to the error prone nature of low fidelity RNA polymerase. In our study, the majority of the circulating pH1N1/09 viruses belonged to either clade 6 or 7. The early pH1N1/09 viruses diversified into 7 clades as revealed by Nelson *et al.*, over a short period since their appearance in 2009 [6]. This phylogenetic analysis provided some evidence of clustering of highly similar sequences suggesting intra-campus transmission. A strong evidence of clustering based on the residence and faculty of the students (P value for AI and PS <0.05) was supported by the phylogeographic analysis. Although a larger dataset is required to strongly substantiate the

conclusion, this study still gives a good evidence of association of phylogeny with geography in a localized community of NUS. On the contrary, another study by Holmes *et al.*, found no association between phylogeny and residence analysing 57 complete genomes of viruses on campus of the University of California, San Diego while there was intra-campus transmission [12].

The panoramic phylogenetic analysis provides a broader view of the process of transmission (Fig 2). The estimated upper-bound of introductions on campus was 8 and lower-bound of introductions was 4, which contributed to lower genetic diversification of virus and hence more clustering was seen. Clustering of the on-campus strains further supports the hypothesis that influenza transmission indeed occurred on-campus.

Conclusions

Phylogenetic data showed evidence of well-supported geographical clustering of highly similar pandemic 2009 H1N1 influenza virus sequences with the majority from on-campus students suggesting some degree of intra-campus transmission. Furthermore, phylogeographic analysis strengthened the evidence of geographical clustering by providing statistically significant association of residence and faculty with clustering. Integration of molecular, epidemiological and statistical methods can help public health authorities to identify foci of transmission in localized communities such as universities. Identification of transmission hot spots can lead to more targeted intervention strategies, such as, closures of universities or campus-based quarantine and even potentially targeted chemoprophylaxis if there is sufficient evidence of intra-campus transmission.

Acknowledgments

We thank all the university students who participated in this study. We also thank Senthamarai Chelvi for help with sample collection, and Elizabeth Ai-Sim Lim, Ka-Wei Chan, Pei Jun and Kim Pern Toh for providing technical assistance.

Author Contributions

Conceptualization: RKV PAT BHT VG.

Data curation: RKV PAT BHT VG HKL MI CC.

Formal analysis: RKV PAT BHT VG HKL.

Funding acquisition: RKV PAT BHT MI.

Investigation: RKV PAT BHT VG HKL.

Methodology: RKV PAT BHT VG HKL MI.

Project administration: RKV PAT BHT.

Resources: RKV PAT BHT VG HKL MI CC.

Software: RKV VG HKL.

Supervision: RKV PAT BHT VG.

Validation: RKV PAT BHT VG HKL MI CC.

Visualization: RKV PAT BHT VG HKL MI CC.

Writing – original draft: RKV.

Writing – review & editing: RKV PAT BHT VG HKL MI CC.

References

1. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*. 2009; 459(7250): 1122–1125. doi: [10.1038/nature08182](https://doi.org/10.1038/nature08182) PMID: [19516283](https://pubmed.ncbi.nlm.nih.gov/19516283/)
2. Le MQ. Migration and persistence of human influenza A viruses, Vietnam, 2001–2008. *Emerg Infect Dis*. 2013; 19(11): 1756–1765. doi: [10.3201/eid1911.130349](https://doi.org/10.3201/eid1911.130349) PMID: [24188643](https://pubmed.ncbi.nlm.nih.gov/24188643/)
3. Cheng X, Tan Y, He M, Lam TT, Lu X, Viboud C, et al. Epidemiological dynamics and phylogeography of influenza virus in southern China. *J Infect Dis*. 2013; 207(1): 106–114 doi: [10.1093/infdis/jjs526](https://doi.org/10.1093/infdis/jjs526) PMID: [22930808](https://pubmed.ncbi.nlm.nih.gov/22930808/)
4. Kidd M. Influenza viruses: update on epidemiology, clinical features, treatment and vaccination. *Curr Opin Pulm Med*. 2014; 20(3): 242–246. doi: [10.1097/MCP.000000000000049](https://doi.org/10.1097/MCP.000000000000049) PMID: [24637227](https://pubmed.ncbi.nlm.nih.gov/24637227/)
5. Rambaut A, Holmes E. The early molecular epidemiology of the swine-origin A/H1N1 human influenza pandemic. *PLoS Currents*. 2009; 1.
6. Nelson M, Spiro D, Wentworth D, Fan J, Beck E, George KS, et al. The early diversification of influenza A/H1N1pdm. *PLoS Currents Influenza*. 2009.
7. Shiino T, Okabe N, Yasui Y, Sunagawa T, Ujike M, Obuchi M, et al. Molecular evolutionary analysis of the influenza A (H1N1) pdm, May–September, 2009: temporal and spatial spreading profile of the viruses in Japan. *PLoS ONE*. 2010; 5(6): e11057. doi: [10.1371/journal.pone.0011057](https://doi.org/10.1371/journal.pone.0011057) PMID: [20548780](https://pubmed.ncbi.nlm.nih.gov/20548780/)
8. Mullick J, Cherian SS, Potdar VA, Chadha MS, Mishra AC. Evolutionary dynamics of the influenza A pandemic (H1N1) 2009 virus with emphasis on Indian isolates: Evidence for adaptive evolution in the HA gene. *Infect Genet Evol*. 2011; 11(5): 997–1005. doi: [10.1016/j.meegid.2011.03.015](https://doi.org/10.1016/j.meegid.2011.03.015) PMID: [21457796](https://pubmed.ncbi.nlm.nih.gov/21457796/)
9. Nelson MI, Lemey P, Tan Y, Vincent A, Lam TT, Detmer S, et al. Spatial dynamics of human-origin H1 influenza A virus in North American swine. *PLoS Pathog*. 2011; 7(6): e1002077. doi: [10.1371/journal.ppat.1002077](https://doi.org/10.1371/journal.ppat.1002077) PMID: [21695237](https://pubmed.ncbi.nlm.nih.gov/21695237/)
10. Graham M, Liang B, Van Domselaar G, Bastien N, Beaudoin C, Tyler S, et al. National Influenza A/H1N1pdm Genomics Study Team. Nationwide molecular surveillance of pandemic H1N1 influenza A virus genomes: Canada. *PLoS ONE*. 2009; 6(1): e16087.
11. Viboud C, Nelson MI, Tan Y, Holmes EC. Contrasting the epidemiological and evolutionary dynamics of influenza spatial transmission. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 2013; 368(1614): 20120199. doi: [10.1098/rstb.2012.0199](https://doi.org/10.1098/rstb.2012.0199) PMID: [23382422](https://pubmed.ncbi.nlm.nih.gov/23382422/)
12. Holmes EC, Ghedin E, Halpin RA, Stockwell TB, Zhang XQ, Fleming R, et al. Extensive geographical mixing of 2009 human H1N1 influenza A virus in a single university community. *J Virol*. 2011; 85(14): 6923–6929. doi: [10.1128/JVI.00438-11](https://doi.org/10.1128/JVI.00438-11) PMID: [21593168](https://pubmed.ncbi.nlm.nih.gov/21593168/)
13. Campbell DS, Rumley MH. Cost-effectiveness of the influenza vaccine in a healthy, working-age population. *J Occup Env Med*. 1997; 39(5): 408–414.
14. Tan AL, Virk RK, Tambyah PA, Inoue M, Lim EA, Chan KW, et al. Surveillance and clinical characterization of influenza in a university cohort in Singapore. *PloS ONE*. 2015; 10(3): e0119485. doi: [10.1371/journal.pone.0119485](https://doi.org/10.1371/journal.pone.0119485) PMID: [25790305](https://pubmed.ncbi.nlm.nih.gov/25790305/)
15. Hoffmann E, Stech J, Guan Y, Webster RG, Perez DR. Universal primer set for the full-length amplification of all influenza A viruses. *Arch Virol*. 2001; 146(12): 2275–2289. PMID: [11811679](https://pubmed.ncbi.nlm.nih.gov/11811679/)
16. Poddar SK. Influenza virus types and subtypes detection by single step single tube multiplex reverse transcription-polymerase chain reaction (RT-PCR) and agarose gel electrophoresis. *J Virol Methods*. 2002; 99(1): 63–70.
17. Ghedin E, Sengamalay NA, Shumway M, Zaborsky J, Feldblyum T, Subbu V, et al. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature*. 2005; 437(7062): 1162–1166. doi: [10.1038/nature04239](https://doi.org/10.1038/nature04239) PMID: [16208317](https://pubmed.ncbi.nlm.nih.gov/16208317/)
18. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28(12): 1647–1649. doi: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199) PMID: [22543367](https://pubmed.ncbi.nlm.nih.gov/22543367/)
19. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* 1. 2015;vev003. doi: [10.1093/ve/vev003](https://doi.org/10.1093/ve/vev003) PMID: [27774277](https://pubmed.ncbi.nlm.nih.gov/27774277/)
20. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 1999; 41: 95–98.

21. Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics*. 1998; 14(9): 817–818. PMID: [9918953](#)
22. Swofford DL. PAUP. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4. Sinauer Associates, Sunderland, MA. 2002.
23. Parker J, Rambaut A, Pybus OG. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol*. 2008; 8(3): 239–246. doi: [10.1016/j.meegid.2007.08.001](#) PMID: [17921073](#)
24. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007; 7: 214. doi: [10.1186/1471-2148-7-214](#) PMID: [17996036](#)
25. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. *Nature*. 2008;29; 453(7195): 615–619. doi: [10.1038/nature06945](#) PMID: [18418375](#)
26. Westgeest KB, Russell CA, Lin X, Spronken MI, Bestebroer TM, Bahl J, et al. Genomewide analysis of reassortment and evolution of human influenza A (H3N2) viruses circulating between 1968 and 2011. *J Virol*. 2014; 88(5): 2844–2257. doi: [10.1128/JVI.02163-13](#) PMID: [24371052](#)
27. Wang TH, Donaldson YK, Brettell RP, Bell JE, Simmonds P. Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *J Virol*. 2001; 75(23): 11686–11699. doi: [10.1128/JVI.75.23.11686-11699.2001](#) PMID: [11689650](#)
28. Fitch WM. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst Biol*. 1971; 20(4): 406–416.
29. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014; 30(9): 1312–1313. doi: [10.1093/bioinformatics/btu033](#) PMID: [24451623](#)
30. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013; 30(12): 2725–2729. doi: [10.1093/molbev/mst197](#) PMID: [24132122](#)
31. Pons VG, Canter J, Dolin R. Influenza A/USSR/77 (H1N1) on a university campus. *Am J Epidemiol*. 1980; 111(1): 23–30. PMID: [7352457](#)
32. Okoror LE, Okoror OI, Okoror OI. Epidemiology of influenza virus in a suspected outbreak of influenza in a university. *Int J Trop Med Public Health*. 2011; 1(1): 1–10.
33. Mitchell T, Massoudi M, Swerdlow DL, Dee DL, Gould LH, Kutty PK, et al. Swine flu in college: early campus response to outbreak control measures. *Am J Health Behav*. 2014; 38(3): 448–464. doi: [org/10.5993/AJHB.38.3.14](#) PMID: [25181765](#)