

Extensive Geographical Mixing of 2009 Human H1N1 Influenza A Virus in a Single University Community

Edward C. Holmes, Elodie Ghedin, Rebecca A. Halpin,
Timothy B. Stockwell, Xing-Quan Zhang, Regina Fleming,
Richard Davey, Constance A. Benson, Sanjay Mehta, Randy
Taplitz, Yu-Tseung Liu, Kimberly C. Brouwer, David E.
Wentworth, Xudong Lin, the INSIGHT FLU002 Study Group
and Robert T. Schooley
J. Virol. 2011, 85(14):6923. DOI: 10.1128/JVI.00438-11.
Published Ahead of Print 18 May 2011.

Updated information and services can be found at:
<http://jvi.asm.org/content/85/14/6923>

SUPPLEMENTAL MATERIAL

These include:

<http://jvi.asm.org/content/suppl/2011/06/16/85.14.6923.DC1.htm>

REFERENCES

This article cites 24 articles, 9 of which can be accessed free at:
<http://jvi.asm.org/content/85/14/6923#ref-list-1>

CONTENT ALERTS

Receive: RSS Feeds, eTOCs, free email alerts (when new
articles cite this article), [more»](#)

Information about commercial reprint orders: <http://jvi.asm.org/site/misc/reprints.xhtml>
To subscribe to to another ASM Journal go to: <http://journals.asm.org/site/subscriptions/>

Extensive Geographical Mixing of 2009 Human H1N1 Influenza A Virus in a Single University Community^{∇†}

Edward C. Holmes,^{1,2*} Elodie Ghedin,^{3,4} Rebecca A. Halpin,³ Timothy B. Stockwell,³
Xing-Quan Zhang,⁶ Regina Fleming,⁷ Richard Davey,⁸ Constance A. Benson,⁶
Sanjay Mehta,⁶ Randy Taplitz,⁶ Yu-Tseung Liu,⁶ Kimberly C. Brouwer,⁹
David E. Wentworth,^{5‡} Xudong Lin,^{5‡} the INSIGHT FLU002
Study Group, and Robert T. Schooley⁶

Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802¹; Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892²; J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850³; Center for Vaccine Research, Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261⁴; Wadsworth Center, New York State Department of Health, Albany, New York 12201⁵; Division of Infectious Diseases, Department of Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0711⁶; Student Health Service, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093⁷; Clinical Center, NIAID, NIH, Bethesda, Maryland 21205⁸; and Division of Global Public Health, Department of Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093⁹

Received 4 March 2011/Accepted 4 May 2011

Despite growing interest in the molecular epidemiology of influenza virus, the pattern of viral spread within individual communities remains poorly understood. To determine the phylogeography of influenza virus in a single population, we examined the spatial diffusion of H1N1/09 influenza A virus within the student body of the University of California, San Diego (UCSD), sampling for a 1-month period between October and November 2009. Despite the highly focused nature of our study, an analysis of complete viral genome sequences revealed between 24 and 33 independent introductions of H1N1/09 into the UCSD community, comprising much of the global genetic diversity in this virus. These data were also characterized by a relatively low level of on-campus transmission as well as extensive spatial mixing, such that there was little geographical clustering by either student residence or city ZIP code. Most notably, students experiencing illness on the same day and residing in the same dorm possessed phylogenetically distinct lineages. H1N1/09 influenza A virus is therefore characterized by a remarkable spatial fluidity, which is likely to impede community-based methods for its control, including class cancellations, quarantine, and chemoprophylaxis.

Although there is growing interest in the rates, patterns, and determinants of the spread of viral infections such as influenza, little is known about the spatial dynamics of viruses within individual communities. However, such information is central to the design of effective intervention strategies. For example, the more frequently a virus enters a specific population, and the lower the frequency of autochthonous transmission, then it is possible that methods for infection control employed at the level of local populations will be less able to contain viral spread. Similarly, understanding the pathways by which viruses spread on a spatial scale will greatly assist in predicting, at least to some extent, their future patterns of movement. Indeed, simple spatial patterns such as unidirectional waves represent one of the cases in which epidemiological predictions can be made with statistical rigor (22).

One virus where spatial dynamics have received consid-

erable attention is pandemic H1N1/09 influenza A virus, which emerged globally in April 2009 and which remains a major cause of influenza infections in some localities. Unsurprisingly, there have been attempts to reconstruct the spread of H1N1/09 at both global (11, 13, 21) and national (8, 9, 14, 15, 19, 24, 26) scales. However, little is known about the patterns and dynamics of H1N1/09 spread at finer spatial scales. For example, while a recent study explored the origins and structure of genetic diversity in three populations from the United States (Houston, TX, Milwaukee, WI, and New York State), there was no attempt to examine the phylogeography of H1N1/09 within each of these populations, in large part because the relevant spatial data were unavailable (14). Similarly, a recent analysis of H1N1/09 virus in Buenos Aires, Argentina (2), did not consider patterns of virus diffusion within this city. As a consequence, how H1N1/09 evolves and spreads within a local population remains unclear.

To explore the patterns and determinants of H1N1/09 virus evolution at a highly localized scale, we analyzed the patterns of spread of this virus as it diffused among the population of college students attending the University of California, San Diego (UCSD). UCSD has over 29,000 undergraduate, graduate, and medical/pharmacy students. On-campus residence

* Corresponding author. Mailing address: Department of Biology, The Pennsylvania State University, University Park, PA 16802. Phone: (814) 863-4689. Fax: (814) 865-9131. E-mail: ech15@psu.edu.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

‡ Present address: J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850.

[∇] Published ahead of print on 18 May 2011.

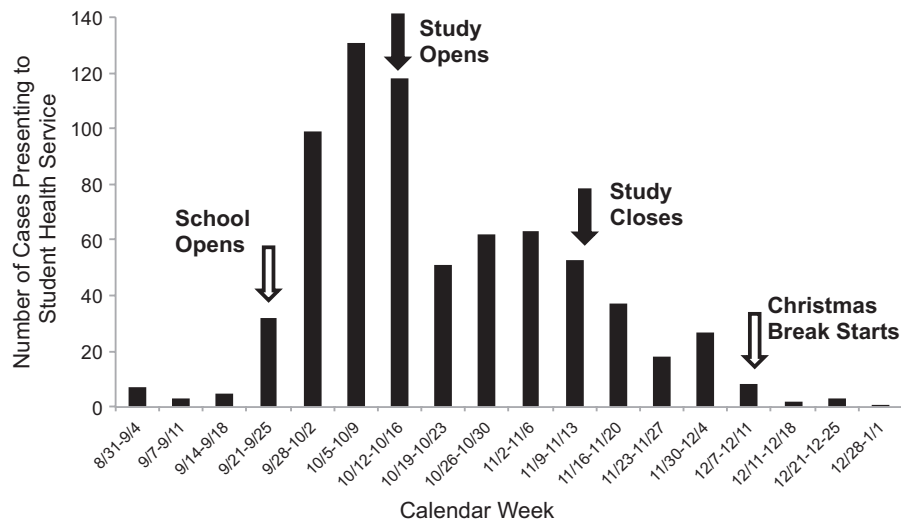


FIG. 1. Course of the second (Fall) wave of H1N1/09 influenza in UCSD, showing the duration of sampling undertaken in this study.

halls and graduate apartments accommodate approximately 6,500 students. The campus represents a particularly informative location for this study because some of the earliest cases of H1N1/09 in the United States were reported from California (16). Complete virus genome sequences were collected from students residing in dorms on campus and in off-campus accommodation situated within the greater San Diego area. Both groups reported to the Student Health Service center with influenza-like illness (ILI). Importantly, our sampling was also performed over a limited time span, representing 1 month from 12 October to 12 November 2009 during the second and largest wave of H1N1/09 influenza A virus in the United States (Fig. 1). Hence, our sampling regime enables us to collect viruses from students who developed symptoms on the same day and who live in very close proximity to each other, providing a uniquely detailed examination of the phylogeography of H1N1/09 influenza A virus. In addition, by comparing H1N1/09 viruses sampled in San Diego with those collected on a global scale, we were able to estimate the number of independent introductions of this virus into the UCSD community as well as their likely geographical origin. Because the H1N1/09 virus is relatively conserved at the molecular level (13), reflecting its recent emergence in humans, and because we found no evidence that reassortment was disrupting evolutionary patterns (see Results), we used complete (i.e., concatenated) genome sequences to maximize phylogenetic resolution.

MATERIALS AND METHODS

Sample collection. Students visiting the UCSD Student Health Service from 12 October to 12 November 2009 with influenza-like-illness (ILI) were asked if they would be interested in participating in an observational study of respiratory virus infection. Interested students were then referred to a mobile research unit that was parked on the grounds of the student health service. Additional students and staff with ILI approached the research unit independently of referral from the Student Health Service. Those aged at least 18 years with a respiratory illness that included cough and/or sore throat and who had a temperature of 37.8°C or greater at presentation or within the prior 12 h were eligible for enrollment in the study. After informed consent was obtained, three nasal and oropharyngeal swabs were obtained. One was used locally for culture and PCR for diagnosis.

Another was shipped overnight to a central specimen repository for subsequent processing for full-length sequencing. Information on the residence of each student, either on or off the UCSD campus, was collected at the time of the study. Because of the time course of the outbreak of ILI, our sampling regime covered only 1 month in late 2009 (12 October to 12 November) (Fig. 1). The number of cases of ILI by week depicted in the figure is derived from a daily log maintained by the Student Health Service to monitor the campus for clinical activity related to respiratory viruses in an ongoing fashion. Data were geocoded or hand digitized with ArcMap 9.3 (ESRI, Redlands, CA). In total, 57 complete genome sequences were collected from UCSD students, for which information on residence was available from 55 (Fig. 2; see also Table S1 in the supplemental material). These data were collected under the auspices of the INSIGHT FLU002 study program, an Institutional Review Board-approved protocol in which the students gave written informed consent to participate.

Virus sequencing. Extraction and amplification of the entire viral RNA genome were performed at the Wadsworth Center, New York State Department of Health, Albany, NY, as described previously (12, 27). Whole-genome sequencing was performed at the J. Craig Venter Institute (JCVI) in Rockville, MD. Samples were processed on the Sanger capillary pipeline by sequencing tiled amplicons designed to have an optimal size of 550 bp with 100-bp overlap. DNA was amplified using Accuprime *Taq* at 35 cycles (denaturation, 0.5 min at 94°C; annealing, 0.5 min at 55°C; and extension, 2 min at 68°C), and amplicons were treated with shrimp alkaline phosphatase and exonuclease I. Sequencing, genome assembly, and closure reactions were performed as described previously (7, 13). All assemblies were submitted to GenBank and are publicly available at the National Center for Biotechnology Information (NCBI) Influenza Virus Resource (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>) (1).

Phylogenetic analysis. Three different data sets of H1N1/09 viruses were analyzed here. First, to place the evolution of H1N1/09 in San Diego in as broad a context as possible, a phylogeny was inferred using all H1N1/09 complete genome sequences available in GenBank, collected from 1 April 2009 to 2 March 2010, combined with the 57 genome sequences determined here. Because of the very large size of this data set, which severely limits computational tractability, identical viruses collected from the same place were excluded (other than those from UCSD). This resulted in a data set of 1,867 complete genome sequences, 13,113 nucleotides (nt) in length. Second, an equivalent phylogenetic analysis was undertaken using all sequences collected from the second wave of H1N1/09, defined here to run from August 2009 to March 2010. Again, this analysis combined sequences from UCSD and those available in GenBank, with identical sequences again excluded. This resulted in a data set of 1,036 complete genome sequences, 13,113 nt in length. Finally, a phylogenetic analysis was conducted on the 57 complete genome sequences sampled from UCSD in isolation (sequence alignment of 13,113 nt in length). All sequences were aligned manually using the SE-AL program (20), with the overlapping regions in the M and NS gene segments excluded.

Because of the very large numbers of sequences involved, a multilayered approach to phylogenetic inference was employed, combining both parsimony

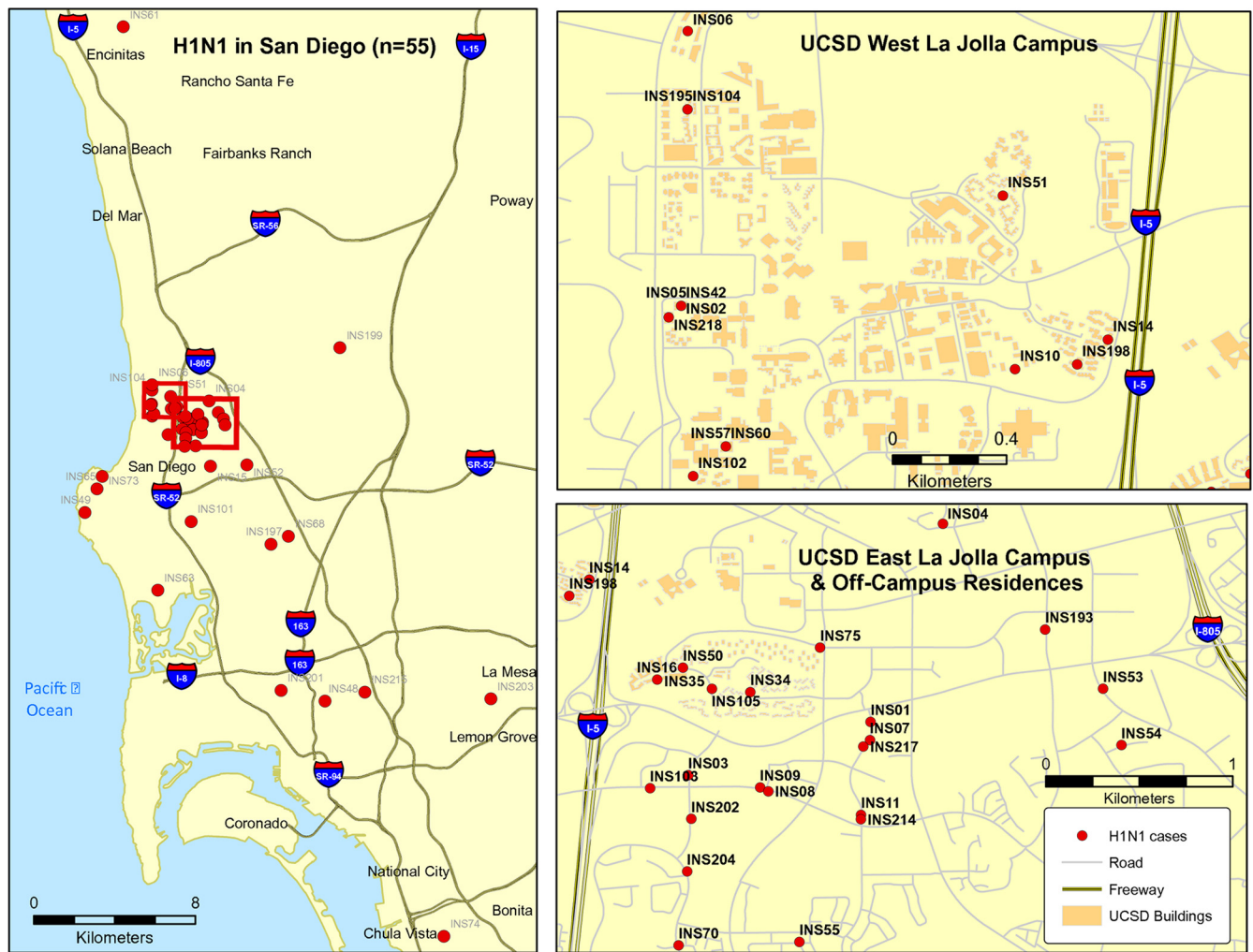


FIG. 2. Map of the UCSD campus (right panels) and the greater San Diego region (left panel) showing the residences (red circles) of individuals sampled in this study. For clarity, the name of each sequence has been abbreviated to “INS.” The east and west campus residences of UCSD are divided by the I-5 freeway.

and maximum likelihood (ML) methods. An initial parsimony analysis was employed to search a broad area of tree space. Given the similarity of the sequences in question, parsimony represents an appropriate phylogenetic method. Accordingly, we were able to search 5.08×10^9 and 4.18×10^{10} trees for the complete and second-wave data sets, respectively. The phylogenetic trees obtained under the parsimony analysis were then used as starting trees in a ML analysis. From these starting trees, we estimated the optimal value of the transition/transversion ratio under the HKY85 model of nucleotide substitution and then attempted to improve the likelihood by employing nearest-neighbor interchange (NNI) branch swapping. Again, the similarity of the sequences in question meant that all inferences could be based on a simple model of nucleotide substitution. Importantly, the phylogenetic positions of the UCSD sequences did not change substantially under the parsimony and ML analysis and were always indicative of multiple introductions (see Results). For the third data set, representing the 57 genome sequences from UCSD in isolation, we employed only an ML analysis with the best-fit nucleotide model (GTR+I) determined by MODELTEST (18) (all parameter values are available from the authors on request). To assess the robustness of the phylogenetic groupings observed, a neighbor-joining bootstrap analysis (1,000 replications) was undertaken using the substitution models described above. All these analyses were conducted using the PAUP* package (25).

Finally, we undertook two further analyses to ensure that our phylogenetic results were not adversely affected by segment reassortment. First, we used the BOOTSCAN, GENECOV, and RDP methods available within the RDP3 program (10) to determine whether there was any evidence for reassortment among

the 57 concatenated H1N1/09 genome sequences sampled from UCSD. Second, using the integrative parsimony and ML approach described above, we inferred separate phylogenetic trees for the HA and NA gene segments from the second H1N1/09 wave data set, comprising sequence alignments of 1,698 nt and 1,407 nt, respectively.

Phylogeographic analysis. To determine the strength and pattern of the geographical structure of H1N1/09 influenza A virus within the UCSD community, we employed three phylogeny-trait association test statistics: the parsimony score (PS), the association index (AI), and the maximum monophyletic clade size (MC) statistics. The first two indicate the strength of phylogenetic clustering by place of isolation, while the MC statistic assesses the association between specific locations and phylogeny by estimating the size of the largest cluster of sequences sampled from a specific geographical location. For this analysis, we utilized the 57-sequence UCSD complete genome data set and coded it in three different ways (although precise spatial information was not available from two sequences): (i) by subdividing the sequences as to whether they were collected from students residing “on” or “off” campus (2 spatial categories); (ii) by ZIP code, such that all on-campus sequences are categorized together (16 spatial categories); and (iii) by ZIP code, where the on-campus sequences were subdivided into those sampled from east and west campus residences, located on either side of the I-5 highway (17 spatial categories). Similarly, to determine whether there was phylogenetic clustering by time, we grouped each of the viruses into five discrete time classes, representing each week of sampling; to determine whether there was phylogenetic clustering by the sex of the patient, we classified viruses ac-

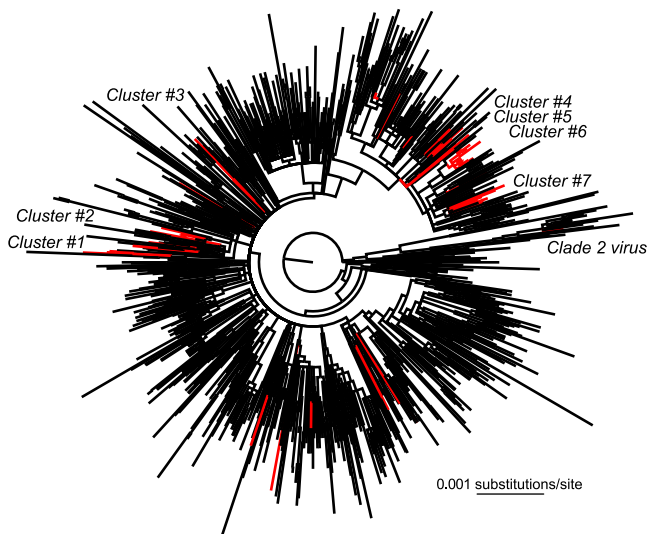


FIG. 3. Maximum likelihood phylogenetic tree of 1,036 complete genome sequences from the second wave of H1N1/09 influenza A virus. Those viruses sampled from the UCSD community are highlighted in red. The approximate positions of the seven clusters of sequences from UCSD are also indicated, as is that of the single clade 2 virus (all other UCSD sequences were classified as clade 7 viruses). All branches are drawn to a scale of nucleotide substitutions per site.

cording to whether they were sampled from males or females, and to determine whether there was phylogenetic clustering by the age of the patient, we grouped each virus into one of nine age classes: years 18, 19, 20, 21, 22, 23, 24, 25, and 26 and over. All these analyses were performed using the BaTS program (17), which calculates empirical distributions of these statistics from the posterior distribution of trees provided by Bayesian phylogenetic inference. The Bayesian trees in this instance were estimated using the Bayesian Markov Chain Monte Carlo approach available in the BEAST package, incorporating the day of sampling of each sequence in question (4). For these trees, the GTR+ Γ model of nucleotide substitution was utilized, along with a relaxed (uncorrelated lognormal) molecular clock and Bayesian skyline coalescent prior. Analyses were run until all parameter values had converged, with statistical uncertainty represented in values of the 95% highest probability density (HPD).

Nucleotide sequence accession numbers. The GenBank accession numbers of the viruses determined here are given in Table S1 in the supplemental material.

RESULTS

The most striking observation from our study was the number of independent entries of H1N1/09 into the UCSD community, despite the relatively small number of sequences collected and the short duration of sampling. For ease of presentation, only the ML tree of the second wave of H1N1/09 virus is shown here (Fig. 3), with that for all available H1N1/09 sequences shown in Fig. S1 in the supplemental material (and these trees did not differ in any substantive way with respect to the UCSD sequences). Because of the lack of resolution in some parts of the H1N1/09 phylogenies, such that most viruses differ by very few nucleotide differences even at the genomic scale, estimating the number of independent introductions of this virus into the UCSD community is complex. An upper-bound estimate of the number of independent introductions is provided by the number of phylogenetically distinct and strongly supported UCSD lineages; that is, well-supported clusters of UCSD viruses ($>70\%$ bootstrap support in the tree of sequences from the second wave of H1N1/09) that are

separated from each other by non-UCSD viruses as well as singleton lineages. Under these criteria, a total of 33 separate introductions into the UCSD community are apparent. A lower-bound estimate of the number of introductions can be obtained by simply counting the total number of UCSD lineages of H1N1/09 that may be considered phylogenetically distinct and hence indicative of independent entry, irrespective of their bootstrap support, plus singleton lineages. With the use of this second approach, a total of 24 separate introductions can be inferred. Importantly, we observed no clear signal for reassortment either with the use of the RDP3 analysis of concatenated genomes or in the individual HA and NA phylogenies. Although there were topological differences among the HA and NA phylogenies and to the whole genome tree, these likely reflect a lack of phylogenetic signal, as bootstrap values in the HA and NA trees were consistently very low, and both trees clearly presented a picture of multiple entries of H1N1/09 into UCSD (Fig. S2). As such, concatenating genomic segments as we have done here represents a valid approach for studying the microevolution of H1N1/09.

Visual inspection of the phylogenies revealed that in all cases the sequences most closely related to the UCSD viruses, and hence their probable source populations, were also sampled from the United States, and often from California (although more-precise location information within California was unavailable). This suggests that there has been at least some regional evolution of H1N1/09 within the United States, although the clear bias toward U.S. sequences in the available data from the second wave means that all conclusions here should be drawn with caution. All but one of the UCSD sequences fell into clade 7 of H1N1/09 influenza A virus, which appears to dominate infections globally during the second wave of H1N1/09 (14). The exception was a single sequence (A/San Diego/INS35/2009) associated with clade 2 that commonly circulated during the first (Spring) wave of H1N1/09 virus (Fig. 3 and 4). Hence, not all early circulating viral lineages suffered extinction with the arrival of the second wave.

Of equal note was the extensive spatial mixing of H1N1/09 within the UCSD community. In total, we observed seven well-supported clusters of sequences from UCSD students that were compatible with intracommunity transmission (Fig. 3 and 4), although the composition of some of the clusters is hard to define and it is striking that none of these clusters contained sequences sampled exclusively from on-campus residences. The remaining sequences fell as singletons, such that each may represent an independent entry into the UCSD community.

Of the seven UCSD clusters, three comprised only two sequences, another two contained three sequences, and one cluster comprised four sequences. Of most significance was the observation of a major cluster (no. 6) of 14 sequences that received strong bootstrap support (82% in the tree of sequences from the second wave of H1N1/09) and was composed of many short branches, including some identical sequences (see below), as expected if some of the individuals were linked by direct, or near-direct, transmission. Although the presence of this cluster is strongly suggestive of intra-UCSD transmission, closer inspection of the phylogenetic trees again reveals a complex picture of geographical movement. In particular, there is no clear clustering according to whether individuals were residing in the same dorm or on the east or west part of

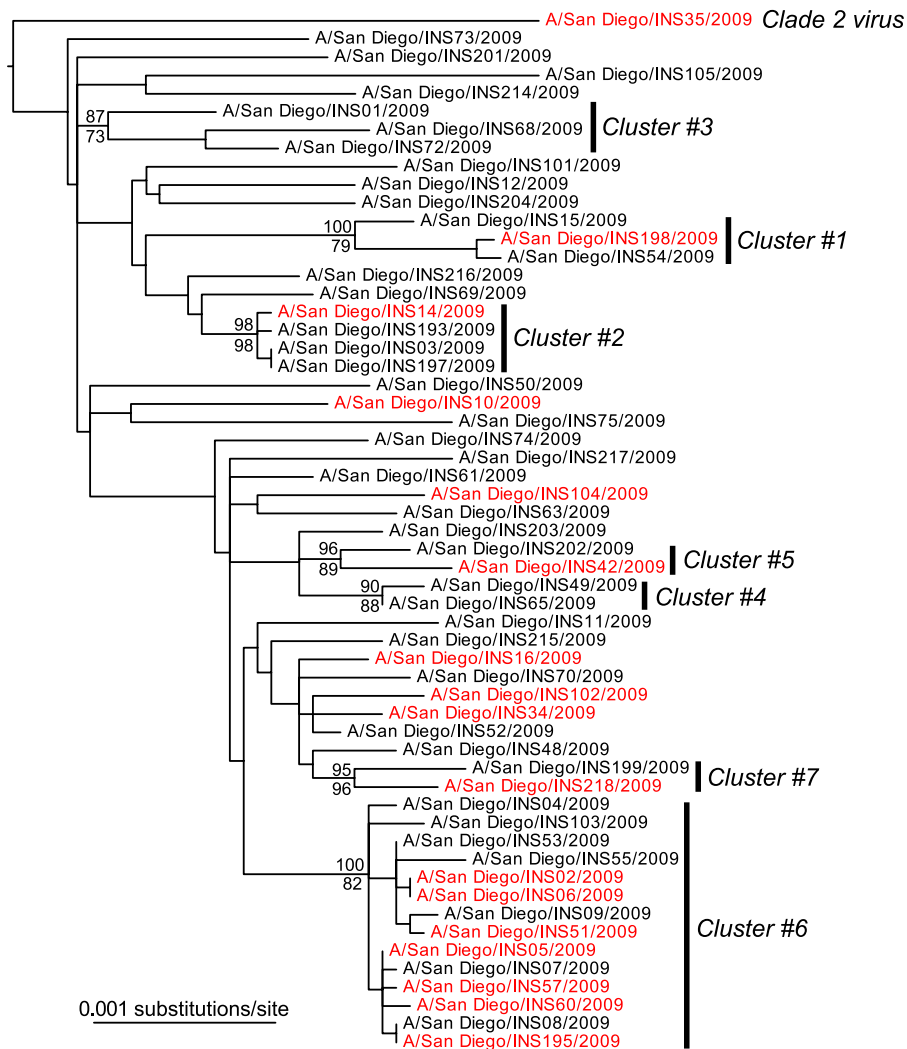


FIG. 4. Maximum likelihood phylogenetic tree of 57 complete genome sequences of H1N1/09 sampled from UCSD. Each of the seven putative UCSD transmission clusters is indicated, with viruses collected from students residing on campus shown in red. Bootstrap support values are shown for the seven clusters of UCSD sequences; values above the branch are from the 57-sequence UCSD phylogeny, while values below the branch are from the phylogeny of the 1,036-sequence second-H1N1/09-wave data set (Fig. 2). Horizontal branches are drawn to a scale of nucleotide substitutions per site.

the campus (see below). For example, sequences A/San Diego/INS02/2009 and A/San Diego/INS05/2009 occupy divergent phylogenetic positions within cluster no. 6, although they were taken from individuals residing in the same dorm and who reported sick on the same day. Such extensive spatial mixing is also apparent from the three identical pairs of genome sequences that were sampled from UCSD. One such pair comprised two sequences (A/San Diego/INS03/2009 and A/San Diego/INS197/2009) sampled off-campus from different ZIP codes in the San Diego region and 9 days apart. A second pair (A/San Diego/INS08/2009 and A/San Diego/INS195/2009) comprised one sequence from an on-campus location and one from an off-campus location, sampled 2 days apart. Most interestingly, a third pair of sequences (A/San Diego/INS02/2009 and A/San Diego/INS06/2009) was sampled on the same day, both on-campus, but from different residences.

To obtain a more quantitative measure of the pattern of the

geographical structure of H1N1/09 in the UCSD community, we employed a number of phylogeny-trait association statistics, in which the trait in question represents a discrete measure of the home residence of each patient. Accordingly, the strength of clustering was explored according to (i) whether individuals were resident “on” or “off” campus, (ii) the ZIP code of each individual residence, and (iii) the ZIP code of each individual, with a division into those individuals resident in either the east or the west part of the UCSD campus. The results here were striking: in no case did we find a significant association between phylogeny and geography ($P > 0.4$ in all cases) (Table 1). Hence, there is no more clustering by residence in these data than might be expected from chance alone. In addition, we observed no significant viral clustering by time (week of sampling) under the AI statistic and only a relatively weakly significant result ($P = 0.044$) under the PS statistic, which was in large part due to a clustering of sequences from week 1 of

TABLE 1. Results of phylogeny trait association tests for H1N1/09 genome sequences sampled from UCSD

Comparison	Statistic ^a	Value (95% CI ^b)		P
		Observed	Null	
On vs off campus	AI	3.42 (2.99–3.86)	2.87 (2.05–3.62)	0.874
	PS	16.51 (15.0–17.0)	16.69 (14.59–18.54)	0.563
ZIP code	AI	5.27 (4.74–5.80)	5.22 (4.54–5.85)	0.520
	PS	35.28 (34.0–36.0)	34.77 (32.22–37.17)	0.541
E/W campus plus ZIP code ^c	AI	5.40 (4.90–5.91)	5.45 (4.74–6.02)	0.409
	PS	36.48 (35.0–37.0)	36.62 (34.13–38.88)	0.578
Wk of sampling	AI	4.21 (3.69–4.74)	4.77 (3.95–5.50)	0.123
	PS	28.02 (27.0–29.0)	30.93 (28.16–33.29)	0.044 ^d
Age of patient (yr)	AI	5.44 (4.93–5.98)	5.68 (5.0–6.25)	0.248
	PS	38.73 (37.0–40.0)	40.01 (37.16–42.66)	0.258
Sex of patient	AI	2.70 (2.22–3.19)	2.65 (1.90–3.44)	0.544
	PS	15.69 (14.0–17.0)	15.52 (13.48–17.40)	0.629

^a AI, association Index; PS, parsimony score.

^b CI, confidence interval.

^c In this analysis, UCSD is divided into the east (E) and west (W) campus residences.

^d The single statistically significant result.

sampling (12 October to 18 October; $P = 0.005$), representing the majority of those collected on campus. Finally, we observed no significant phylogenetic clustering by either the age or the sex of the patients in question (Table 1).

DISCUSSION

Our microevolutionary study of the spatial spread of H1N1/09 influenza A virus within a single university community has revealed a remarkably fluid picture of virus dynamics. Although all the viruses sampled from UCSD were collected over a 1-month period during October to November 2009, and from a restricted geographical area, we revealed at least 24 independent entries of H1N1/09 into the university community, and only small clusters of viruses that seemed to be suggestive of intra-UCSD transmission. Hence, although our study considers only a single community, the large number of independent virus entries and their diverse positions on phylogenetic trees indicates that UCSD contains an effectively global sample of H1N1/09 genetic diversity. Similarly, there was a marked lack of clustering of viruses by residence, either on or off campus, with no significant clustering by spatial location. Therefore, in contrast to what might have been anticipated, we find no evidence for a point source outbreak in which a single virus lineage entered the UCSD community and then diffused through an interconnected population. As such, our study illustrates how H1N1/09 influenza A virus is able to move rapidly through geographic space, exploiting the complex movement patterns of human populations and inhibiting attempts at predicting its future pattern of spread. Although we consider only a single university population in the United States, we believe that these results are likely to be generally applicable to the spread of influenza virus within modern and highly mobile populations.

Another unexpected observation was the marked lack of clustering by university residence. Although a larger sample of viruses will doubtless reveal more evidence for dorm-based transmission, its absence from the current data suggests that the virus may be more routinely transmitted in classrooms, social areas, etc., where students regularly interact. Indeed,

such a hypothesis is compatible with the strong phylogenetic connections between those viruses sampled on and off campus. Similarly, despite the highly localized sampling undertaken here, as well as the comparison of complete genome sequences, it is clearly going to be difficult to use genomic data in isolation to determine exactly who infected whom (although this was not part of our study). Hence, contact tracing for influenza A virus may require the analysis of intrahost gene sequence data, such as those produced by next-generation sequencing methods, which allow the analysis of the multiple virus lineages that may pass between hosts. Indeed, recent studies of both seasonal (5) and H1N1/09 (6) influenza virus, as well as experimental studies with equine influenza virus (12), have shown that multiple virus strains may be passed on at interhost virus transmission.

Finally, the data presented here have important implications for how health authorities might attempt to control the spread of influenza viruses. That H1N1/09 influenza A virus has moved onto the campus so many times independently, with relatively little on-campus transmission, suggests that on-campus methods for controlling the spread of influenza, such as campus- or dormitory-based quarantine, class cancellation, distribution of respiratory isolation equipment, or population-based chemoprophylaxis of university students and staff, will be insufficient to fully prevent virus transmission. Although recommendations for nonpharmaceutical interventions to prevent the spread of pandemic influenza on University campuses have been drafted, in the example of UCSD the continual daily influx of students and staff who live in the greater San Diego area, and hence the continual in-flow of viruses, would greatly limit the effectiveness of on-campus control measures (3, 23). A number of factors must be considered with each outbreak as to the potential impact of isolation and quarantine measures at the community level. These include the pathogenicity and infectivity of the agent, the level of underlying immunity in the affected communities, and the social topography of those at risk within specific subpopulations. In the face of such remarkably fluid dynamics, widespread vaccination would seem to be the only appropriate way to control the spread of influenza A

virus, including those students who reside outside the main campus.

ACKNOWLEDGMENTS

We thank all the UCSD students who took part in this study. We also thank Tari Gilbert, Jill Kunkel, Linda Meixner, and Edward Seefried of the UCSD Antiviral Research Center and Rubina Ghazaraian for sample and data collection. E.C.H. is supported in part by NIH grant R01 GM080533. Support was also provided in part by NIAID/NIH contract HHSN272200900007 to E.C.H., E.G., R.A.H., and T.B.S., by cooperative agreement U01 AI074521 to R.T.S., by NIH grant K01DA020364 to K.C.B., and by support from the INSIGHT Network. INSIGHT is funded by NIH grant U01-AI068641, and the FLU002 study is funded by SAIC—Frederick, Inc., prime contract HHSN261200800001E, NCI—Frederick, Frederick, MD.

REFERENCES

1. Bao, Y., et al. 2008. The influenza virus resource at the National Center for Biotechnology Information. *J. Virol.* **82**:596–601.
2. Barrero, P. R., M. Viegas, L. E. Valinotto, and A. S. Mistchenko. 2011. Genetic and phylogenetic analyses of influenza A H1N1pdm virus in Buenos Aires, Argentina. *J. Virol.* **85**:1058–1066.
3. Centers for Disease Control and Prevention. 23 January 2011, accession date. Interim pre-pandemic planning guidance: community strategy for pandemic influenza mitigation in the United States—early, targeted, layered use of non-pharmaceutical interventions, Atlanta, 2007. <http://www.flu.gov/professional/community/commitigation.html#app7>.
4. Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**:214.
5. Ghedin, E., et al. 2009. Mixed infection and the genesis of influenza diversity. *J. Virol.* **83**:8832–8841.
6. Ghedin, E., et al. 2011. Deep sequencing reveals mixed infection with pandemic influenza A (H1N1) virus strains and the emergence of oseltamivir resistance. *J. Infect. Dis.* **203**:168–174.
7. Ghedin, E., et al. 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* **437**:1162–1166.
8. Graham, M., et al. 2011. Nationwide molecular surveillance of pandemic H1N1 influenza A virus genomes: Canada, 2009. *PLoS One* **6**:e16087.
9. Ikonen, N., et al. 2010. Genetic diversity of the 2009 pandemic influenza A(H1N1) viruses in Finland. *PLoS One* **5**:e13329.
10. Martin, D. P., et al. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**:2462–2463.
11. Martinez-Hernandez, F., et al. 2010. What happened after the initial global spread of pandemic human influenza virus A (H1N1)? A population genetics approach. *Virol. J.* **7**:196.
12. Murcia, P. R., et al. 2010. The intra- and interhost evolutionary dynamics of equine influenza virus. *J. Virol.* **84**:6943–6954.
13. Nelson, M. I., et al. 2009. The early diversification of influenza A/H1N1pdm. *PLoS Curr.* **1**:RRN1126.
14. Nelson, M. I., et al. 2011. Phylogeography of the spring and fall waves of the H1N1/09 pandemic influenza virus in the United States. *J. Virol.* **85**:828–834.
15. Nougairède, A., et al. 2010. Novel virus influenza A (H1N1sw) in South-Eastern France, April–August 2009. *PLoS One* **5**:e9214.
16. Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team. 2009. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N. Engl. J. Med.* **360**:2605–2615.
17. Parker, J., A. Rambaut, and O. G. Pybus. 2008. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect. Genet. Evol.* **8**:239–246.
18. Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
19. Potdar, V. A., et al. 2010. Genetic characterization of the influenza A pandemic (H1N1) 2009 virus isolates from India. *PLoS One* **5**:e9693.
20. Rambaut, A. 2002. Sequence alignment editor. University of Edinburgh, Edinburgh, United Kingdom. <http://tree.bio.ed.ac.uk/software/seal/>.
21. Rambaut, A., and E. C. Holmes. 2009. The early molecular epidemiology of the swine-origin A/H1N1 human influenza pandemic. *PLoS Curr.* **1**:RRN1003.
22. Russell, C. A., D. L. Smith, J. E. Childs, and L. A. Real. 2005. Predictive spatial dynamics and strategic planning for raccoon rabies emergence in Ohio. *PLoS Biol.* **3**:e88.
23. Santibañez, S., A. E. Fiore, T. L. Merlin, and S. Redd. 2009. A primer on strategies for prevention and control of seasonal and pandemic influenza. *Am. J. Public Health* **99**:S216–S224.
24. Shiino, T., et al. 2010. Molecular evolutionary analysis of the influenza A(H1N1)pdm, May–September, 2009: temporal and spatial spreading profile of the viruses in Japan. *PLoS One* **5**:e11057.
25. Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.
26. Valli, M. B., et al. 2010. Evolutionary pattern of pandemic influenza (H1N1) 2009 virus in the late phases of the 2009 pandemic. *PLoS Curr.* **2**:RRN1149.
27. Zhou, B., et al. 2009. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza viruses. *J. Virol.* **83**:10309–10313.