Open Forum Infectious Diseases

MAJOR ARTICLE

SARS-CoV-2 Genomic Surveillance Reveals Little Spread From a Large University Campus to the Surrounding Community

Andrew L. Valesano,[1] William J. Fitzsimmons,[2] Christopher N. Blair,[2] Robert J. Woods,[2] Julie Gilbert,[3] Dawn Rudnik,[4] Lindsey Mortenson,[4] Thomas C. Friedrich,[5] David H. O'Connor,[6] Duncan R. MacCannell,[7] Joshua G. Petrie,[3] Emily T. Martin,[3] and Adam S. Lauring[1,2,©]

[1]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA, [2]Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA, [3]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA, [4]University Health Service, University of Michigan, Ann Arbor, Michigan, USA, [5]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA, [6]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, Wisconsin, USA, and [7]Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Background. Coronavirus disease 2019 (COVID-19) has had high incidence rates at institutions of higher education (IHE) in the United States, but the transmission dynamics in these settings are poorly understood. It remains unclear to what extent IHE-associated outbreaks have contributed to transmission in nearby communities.

Methods. We implemented high-density prospective genomic surveillance to investigate these dynamics at the University of Michigan and the surrounding community during the Fall 2020 semester (August 16–November 24). We sequenced complete severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomes from 1659 individuals, including 468 students, representing 20% of cases in students and 25% of total cases in Washtenaw County over the study interval.

Results. Phylogenetic analysis identified >200 introductions into the student population, most of which were not related to other student cases. There were 2 prolonged student transmission clusters, of 115 and 73 individuals, that spanned multiple on-campus residences. Remarkably, <5% of nonstudent genomes were descended from student clusters, and viral descendants of student cases were rare during a subsequent wave of infections in the community.

Conclusions. The largest outbreaks among students at the University of Michigan did not significantly contribute to the rise in community cases in Fall 2020. These results provide valuable insights into SARS-CoV-2 transmission dynamics at the regional level.

Keywords. genomic epidemiology; infection prevention; SARS-CoV-2; transmission; university.

Institutions of higher education (IHE) have been associated with high incidence of coronavirus disease 2019 (COVID-19) in the United States [1–3]. Congregate settings, such as on-campus housing and off-campus social gatherings, have led to large outbreaks despite prevention efforts [4–8]. It is essential to gain a better understanding of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission dynamics surrounding IHE to inform prevention strategies [9–11], especially as more contagious variants circulate. An important question is to what extent IHE-related outbreaks have contributed to transmission in the communities where IHE are geographically located. One study showed that counties with large IHE that opened for in-person learning had higher incidence of COVID-19 compared with matched counties with remote-only learning [2]. However, it is possible that this reflects transmission into and among student populations without spread into surrounding communities. Some studies have analyzed case counts and anonymized behavioral and movement data to assess mixing between populations, but these approaches are limited in their ability to track specific transmission chains [12, 13]. Complementary approaches, such as contact tracing and genomic epidemiology, may be useful to more directly assess whether IHE-related outbreaks have spread into nearby communities.

Virus genome sequencing has been an important epidemiologic tool during the COVID-19 pandemic, enabling the characterization of transmission lineages and their connections across different populations [14–16]. Phylogenetic analysis from well-sampled populations can reveal the number of unique introductions of SARS-CoV-2, the growth and persistence of lineages, and the frequency of transmission crossover between groups. An important advantage of genomic surveillance is that it can rule out or establish transmission relatedness between groups that appear to be associated through

Received 27 September 2021; editorial decision 5 October 2021; accepted 7 October 2021; published online XX XX XXXX.

Correspondence: Adam S. Lauring, MD, PhD, MS2 4742C, SPC 1621, 1137 Catherine Street, Ann Arbor, MI 48109 (alauring@med.umich.edu).

Open Forum Infectious Diseases®2021

https://doi.org/10.1093/ofid/ofab518

analyses of case counts and onset dates. However, the power of genomic surveillance to answer these questions relies on dense, comprehensive sampling in the populations under investigation.

Recent studies suggest that spillover from outbreaks at IHE occurs infrequently, but there are relatively few that have used high-density genomic surveillance. A genome sequencing study described the transmission of SARS-CoV-2 from an outbreak among students in La Crosse County, Wisconsin, into skilled nursing facilities, resulting in 2 deaths [17], but other studies have noted limited transmission from students. A seroprevalence study in Pennsylvania indicated that the community incidence of COVID-19 remained low throughout a period of high incidence in students [18]. Few or no viral descendants of student clusters in Dane County, Wisconsin, and Washington State have been detected in genomic surveillance of their respective communities with sampling density of 1%–3% [4, 6]. One notable study from a university in the United Kingdom used genomic surveillance (8% of the community) and contact tracing to demonstrate separate viral lineages in university vs community cases [19]. Because these studies have varied greatly in the depth and breadth of community sampling, the extent of transmission between students and the community in other settings is not clear.

To address this question, we conducted prospective, high-density genomic SARS-CoV-2 surveillance during the Fall 2020 semester at the University of Michigan–Ann Arbor and the surrounding area. We sequenced complete genomes from 1659 individuals, including 468 students and 1191 nonstudents. We captured 20% of confirmed cases in University of Michigan–Ann Arbor students and roughly one-quarter of confirmed cases in Washtenaw County, where the university is located. We detected >200 independent transmission introductions into the student population, including 2 large lineages that persisted for several weeks. However, there was very little crossover from student lineages into the community. The largest student-associated lineages waned by mid-November 2020, when community incidence drastically increased. We conclude that large outbreaks among students at the University of Michigan–Ann Arbor did not significantly drive the increase in community COVID-19 incidence in Southeastern Michigan in November 2020.

METHODS

Research Ethics and Sample Sources

Use of residual SARS-CoV-2-positive specimens from Michigan Medicine laboratories and collection of student status and on-campus residence were approved by the Institutional Review Board at the University of Michigan (HUM185966).

Genome Amplification, Sequencing, and Consensus Generation

We sequenced SARS-CoV-2 genomes as described in Valesano et al. [20]. Briefly, we extracted RNA from nasopharyngeal specimens with the PureLink RNA kit and reverse-transcribed RNA with SuperScript IV. We amplified SARS-CoV-2 cDNA in 2 pools using the ARTIC Network v3 primers and protocol. We combined polymerase chain reaction products of each pool in equal volumes for a given sample and purified with AMPure beads. We prepared libraries for sequencing with the NEBNext Ultra II Library Prep Kit. We quantified pooled libraries with a Qubit and sequenced on an Illumina MiSeq (v2 chemistry, 2×250 cycles). We mapped reads to the Wuhan/Hu-1/2019 reference genome (GenBank MN908947.3) with BWA-MEM [21]. We trimmed the amplification primer sequences and determined consensus sequences with iVar 1.2.1 [22], using bases with >50% frequency and placing an N at positions covered by <10 reads. Genomes with ≥29 000 unambiguous bases (>97%) were used in downstream analysis.

Case Definitions and Metadata

We used unique identifiers associated with each sample to obtain a single genome per individual and determine which individuals were students in the fall semester (undergraduate, graduate, or professional). We queried 2 databases hosted by the University Health Service to identify students and housing status. Students with a campus residence hall listed in either database were considered "on campus," and the rest were considered "unknown." We obtained case metrics in students from the university COVID-19 Dashboard (https://campusblueprint.umich.edu/dashboard/). We obtained data on new confirmed cases in Washtenaw County and Region 2S from the Michigan Department of Health and Human Services website (accessed March 4, 2021).

Phylogenetic Analysis and Discrete Trait Reconstruction

We subsampled genomes on GISAID using augur [23], excluding those with <27 000 bases, and aligned with MAFFT. We prioritized genomes with genetic similarity to Michigan sequences as follows: 100 genomes per month in Michigan, 5 genomes per month per division in the United States outside of Michigan, 1 genome per country per month in North America outside of the United States, and 2 genomes per country per month outside of North America. We included all complete genomes sequenced here (n = 1659) and all other genomes from Michigan collected in August–December 2020. This resulted in an alignment with 7174 sequences, including 3318 genomes from Michigan, 2157 genomes from the United States outside of Michigan, and 1699 global genomes.

We masked the 5' and 3' ends of the alignment along with other sites commonly affected by sequencing errors and homoplasies. We inferred a maximum likelihood phylogeny with IQ-TREE with a generalized time reversible model [24]

and used TreeTime to generate a time-scaled phylogeny rooted on Wuhan/Hu-1/2019 with a clock rate of 0.0008 with a standard deviation of 0.0004 substitutions/site/day [25]. We used TempEst to fit and plot a root-to-tip regression of divergence over time [26]. We removed genomes that exceeded 3 interquartile ranges from the root-to-tip regression (n = 25, including 2 genomes we generated). We generated a new time-scaled tree with this filtered alignment (n = 7149). We used it as the basis for discrete-trait ancestral state reconstruction with TreeTime and BEAST 1.10.4 [27, 28]. We inferred introductions from nonstudent to student populations at nodes with a BEAST probability of >0.9. To determine whether the contextual genomes were biasing our results, we generated a total of 10 random subsamples of the global data using the schema described above and analyzed the data in the same manner (Supplementary Figure 2).

Availability of Data and Materials

Consensus genomes generated for this study are available on GISAID. Accessions for all genomes are listed in Supplementary Table 1. Analysis code is available at https://github.com/lauringlab/SARS2_Fall_2020.

RESULTS

We initiated prospective genomic surveillance in Southeastern Michigan in August 2020 with the goal of capturing SARS-CoV-2 transmission dynamics at the University of Michigan–Ann Arbor and the surrounding community. We obtained and sequenced all available SARS-CoV-2-positive specimens from the Michigan Medicine Clinical Microbiology Laboratory and the University Health Service (UHS) on a daily basis, from August 16, 2020, through November 24, 2020. November 24 corresponded to the end of in-person instruction for the semester and the peak of the November surge of new cases in Washtenaw County. The Clinical Microbiology Laboratory has performed COVID-19 testing for all inpatient and ambulatory clinical settings associated with Michigan Medicine, a large academic medical center with ~2.3 million patient clinic visits annually. Testing of students presenting to UHS was either performed on site or sent to the Clinical Microbiology Laboratory. These 2 specimen sources allowed us to broadly sample the university student population as well as individuals from the community.

University Setting and Epidemic Course

The COVID-19 epidemic in Southeastern Michigan evolved over the Fall 2020 semester. Washtenaw County (population 367 000) is part of Michigan Public Health Preparedness Region 2 South, which includes Monroe and Wayne counties and the city of Detroit. The State of Michigan was under a face mask requirement throughout fall of 2020. In addition, gathering sizes at private residences were restricted through state and county orders to no more than 10 people indoors and no more than 25 outdoors, and restaurant capacity was limited to 50%. From August 16 through November 24, there were 6707 laboratory-confirmed COVID-19 cases in Washtenaw County. COVID-19 cases increased in Washtenaw County over the fall semester (Figure 1A), rising from 130 cases during the week of August 23 (63 daily cases per million) to 1125 cases during the week of November 15 (405 daily cases per million).

The University of Michigan–Ann Arbor, a public university with undergraduate, graduate, and professional students, is centrally located in the city of Ann Arbor. In Fall 2020, the university opened for on-campus student residence and hybrid in-person and remote learning. There were 12 812 students residing in Washtenaw County for Fall term, with 1736 in campus

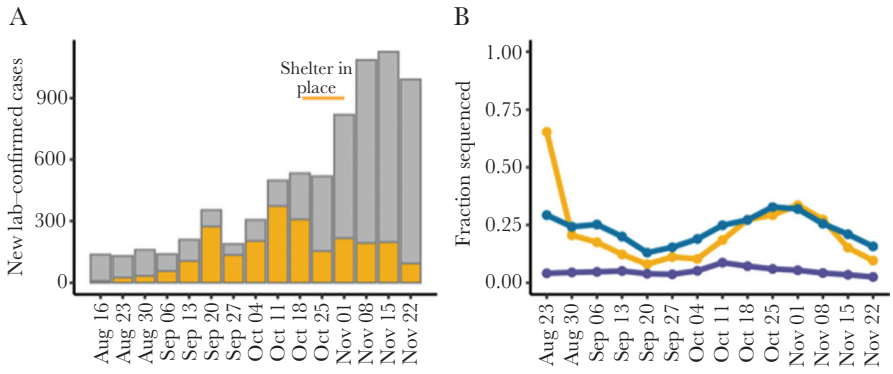

Figure 1. Case curves and sequencing density. A, New lab-confirmed cases of COVID-19 in Washtenaw County, Michigan, from the week of 8/16/2020 through 11/23/2020, displayed by day of symptom onset (as reported by MDHHS). New cases per week are shown on the y-axis and time in weeks on the x-axis. The fraction of new lab-confirmed cases in University of Michigan students is shown in yellow. The Washtenaw County "shelter-in-place" order for undergraduates is indicated (10/20/2020 through 11/03/2020). B, Sampling density is displayed as the fraction of new lab-confirmed COVID-19 cases with complete genome sequences (y-axis) per week during the fall term (x-axis). The fraction of student cases sequenced is shown in yellow, all Washtenaw County cases in blue, and all cases in Region 2S in violet (includes Washtenaw, Wayne, and Monroe counties). Abbreviations: COVID-19, coronavirus disease 2019; MDHHS, Michigan Department of Health and Human Services.

housing. Campus mitigation measures included August pre-arrival testing of students, daily symptom checks and reporting through a phone app, asymptomatic testing of a subset of students, symptomatic testing, isolation of cases and quarantine of close contacts, reduced residence hall and classroom capacity, and reduced occupancy in university studios and laboratories. In-person instruction was held from August 31 through November 24.

Out of 72 798 tests performed from the week of August 16 through the week of November 22, there were 2374 COVID-19 cases in students, 1064 (44.8%) of which had testing performed at Michigan Medicine laboratories. There was no syndromic case definition for obtaining COVID-19 testing at that time, and testing was available for both symptomatic and asymptomatic cases depending on clinical circumstances. Cases in University of Michigan–Ann Arbor students constituted the majority of cases in Washtenaw County by the end of September 2020 (Figure 1A). In response to this spike, additional mitigation efforts were implemented, including stay-in-place orders targeted to specific residence halls, mass testing events, and a broad stay-in-place order for all undergraduates from October 20 to November 3. The fraction of campus-associated cases in the county declined after mid-October 2020 during a wave of new infections in Washtenaw County in November 2020.

Genomic Surveillance in Southeastern Michigan

We assembled complete SARS-CoV-2 genomes from 1659 individuals. This represents a median of 24% of the confirmed cases per week from Washtenaw County and a median of 4.5% of cases per week from Michigan Region 2 South (Figure 1B). We sequenced 468 complete genomes from University of Michigan–Ann Arbor students, representing 20% of cases from August 16 through November 24 (Supplementary Figure 1). We were able to determine on-campus residences for 131 of these students; the on- or off-campus residence of the remainder is unknown. The genomes presented here consisted of several different viral clades, mostly Nextstrain clades 20A, 20C, and 20G (Supplementary Figure 1).

We used phylogenetic analysis to characterize the influx of viruses into the student population. We generated a time-calibrated maximum likelihood phylogenetic tree with our sequenced genomes and additional contextual genomes (see the "Methods" section; Supplemental Figure 2). To optimize the inference of transmission lineages in Southeastern Michigan, we included all available genomes from Michigan on GISAID that were collected from July to December 2020. We inferred traits of ancestral nodes on this time-calibrated phylogenetic tree using a binary discrete trait model of student vs nonstudent, as has been performed in related studies [27, 29–31]. Genomes from students that shared the same ancestral "student" node were considered part of the same introduced transmission lineage. Genomes from students that were not preceded by a "student" node were considered singleton introductions. We verified that our results were not substantially biased by contextual genomes using the same analysis on 10 random subsamples of non-Michigan genomes [29].

Using this approach, we inferred 203 distinct transmission introductions into students (Supplementary Figure 2). There were 2 large transmission lineages in students, which we denote as Cluster A (n = 115 students) and Cluster B (n = 73 students). These were the predominant source of cases in students during the middle of the semester, representing >50% of genomes from students from the week of September 20 through the week of October 18. The rest of the introductions were phylogenetic singletons (n = 171) or small clusters of 2–8 students (n = 30 introductions). Small transmission lineages (2–8 students) were often short in duration, lasting a median (IQR [interquartile range]) of 3.5 (1–8) days. The frequency of singletons increased during the latter half of the semester (Figure 2B), indicating new introductions into the student population rather than spread from older transmission clusters, such as Clusters A and B. These data suggest a shifting epidemiology in the student population, characterized by 2 dominant lineages during the early and middle portions of the semester followed by many small introductions later in the semester as community incidence increased.

We further investigated the 2 largest inferred lineages, which indicate sustained local transmission of SARS-CoV-2 within the student population. Genomes from Cluster A were part of Nextstrain clade 20B and Pango lineage B.1.1.304 (TMRCA, August 26; 95% CI, August 3–September 26). Genomes from Cluster B were part of Nextstrain clade 20C and Pango lineage B.1.593 (TMRCA, July 29; 95% CI, July 3–August 28). We detected cases from Cluster A from October 8 through November 14 and Cluster B from September 11 through October 23 (Figure 2C). These clusters were genetically distinct from other lineages in our data and were identified in all 10 independent subsamples (Supplementary Figure 2). Lineages B.1.1.304 and B.1.593 are rare in the GISAID database (0.07% and 0.02%, respectively, out of 532 715 genomes from the United States as of June 7, 2021). Outside of the genomes presented here, B.1.1.304 has been detected in Michigan only 3 times (238 in the United States), and B.1.593 has not been detected elsewhere in Michigan (only twice in the United States). The rarity of these lineages suggests that these outbreaks among students did not spark significant transmission in Southeastern Michigan or in other parts of the country.

We examined the on-campus residence locations for students in these transmission lineages. There was no obvious association with a single residence hall in any transmission lineage in students (Figure 2C). Of the 131 individuals with known on-campus residences, 31% were singleton introductions (n = 40) and 56% (n = 73) were part of Cluster A or B. Individuals from many different residence halls dispersed across the campus were

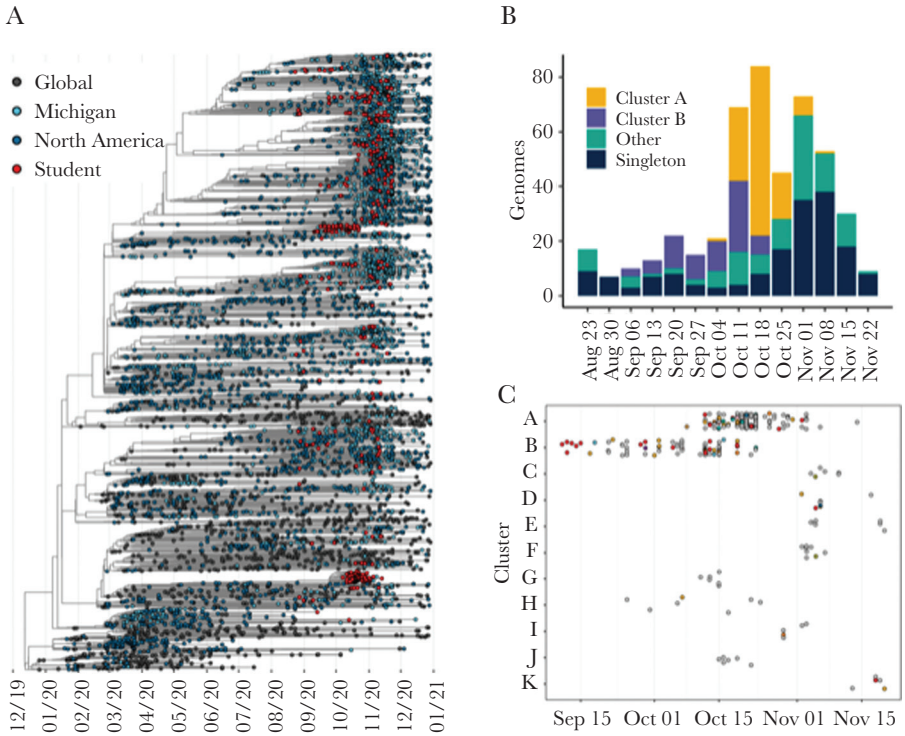

Figure 2. Introductions into the student population. A, Time-calibrated maximum likelihood phylogenetic tree of 7149 SARS-CoV-2 genomes, including 1657 genomes sequenced for this study. Month is shown on the x-axis. Genomes from students are shown in red, other genomes from Michigan in light blue, other genomes from North America in dark blue, and global genomes in gray. B, Counts of genomes from students (y-axis) per week during the fall term (x-axis) by inferred transmission lineage group. Singleton introductions are shown in dark blue, genomes from Cluster A in yellow, genomes from Cluster B in violet, and genomes from smaller clusters (2–8 individuals) in teal. C, Inferred transmission lineages in students, with genomes from each lineage or cluster (y-axis) shown as points with time on the x-axis. Only inferred transmission lineages with ≥5 students are shown. Point colors reflect on-campus residences, if known (unknown residence is light gray). Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

present in Clusters A and B, including students from 9 and 7 residences, respectively. Of the 188 students in Clusters A and B, 61% (n = 115) did not have an identifiable on-campus residence and likely lived off-campus (Figure 2C, light gray points). Besides Clusters A and B, there was only 1 other student lineage that had >1 individual from the same campus residence (2 of 3 students in the lineage were from the same residence). The first individuals detected in Cluster B resided in the same residence hall (Figure 2C), potentially reflecting transmission in a congregate setting. However, without detailed contact tracing to complement the genomic data, it is unclear whether clusters A and B originated in specific residence halls and then spread further among students, or whether the clusters originated in off-campus gatherings. Overall, these data demonstrate local transmission among the student population with intermixing among students from multiple on-campus residences and students residing off-campus.

Limited Spillover From Student Clusters Into the Broader Community

Clusters A and B, defined by Pango lineages B.1.1.304 and B.1.593, waned by mid-November, when new COVID-19 cases increased in Washtenaw County (Figure 3), which suggested that these large, student-dominated lineages did not significantly contribute to the rise in community cases. To further examine the extent of COVID-19 spread from students into the broader Southeastern Michigan community, we used the same ancestral trait reconstruction method as above (see the "Methods" section). There were very few nonstudents in

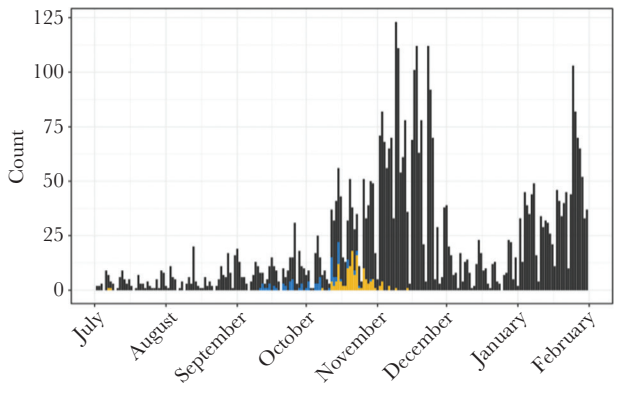

Figure 3. Frequency of lineages B.1.1.304 and B.1.593 in Michigan. All genomes from Michigan available on GISAID from July 2020 to January 2021 are shown. Count of genomes is shown on the y-axis with time on the x-axis (binwidth = 1 day). Genomes from lineage B.1.593 are shown in blue, genomes from B.1.1.304 in yellow, and genomes from all other lineages are shown in gray.

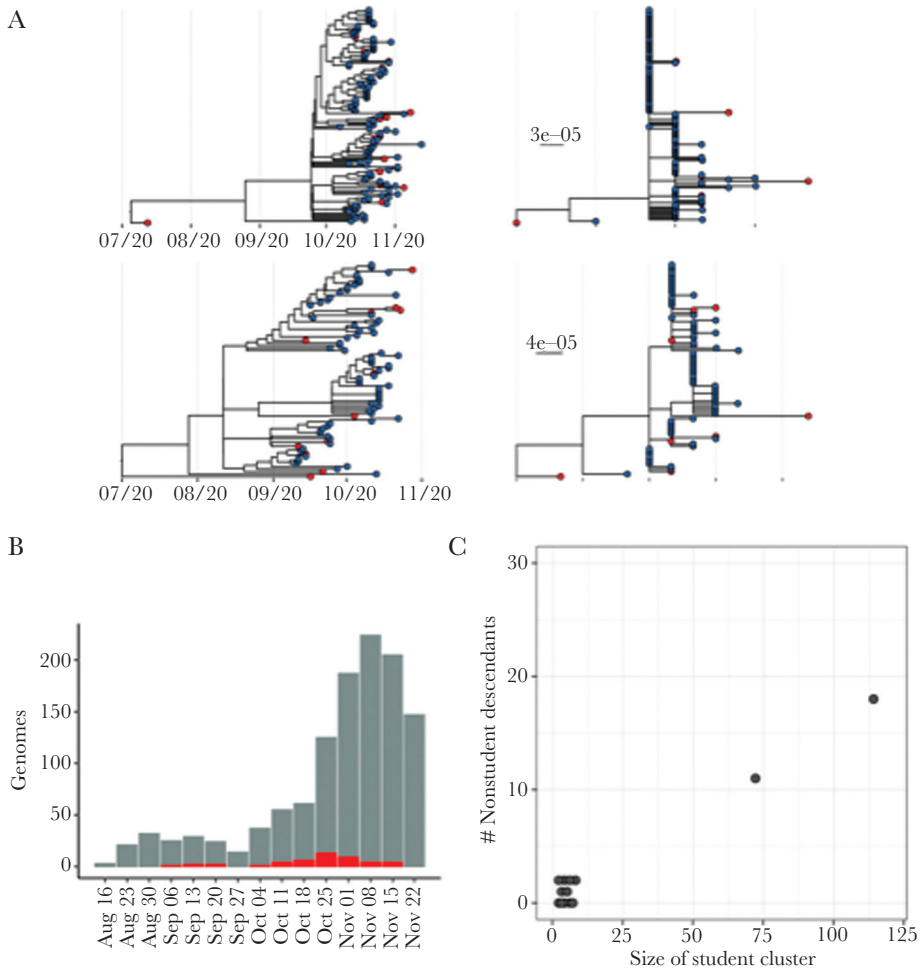

Figure 4. Spillover from student-associated clusters. A, Maximum likelihood phylogenetic trees of student Cluster A (top) and Cluster B (bottom). Time-calibrated trees are displayed on the left and divergence trees on the right. Tip colors reflect genomes from students (blue) and nonstudents (red). B, Bar plot of the number of nonstudent genomes sequenced (y-axis, n = 1191) per week over the fall term (x-axis). Genomes that are derived from inferred "student" nodes are shown in red, and genomes not derived from "student" nodes are shown in gray. C, For each nonsingleton transmission lineage in students, the number of nonstudent descendants is shown (y-axis) by the number of students in the cluster (x-axis).

the community with genomes that descended from student-associated lineages (n = 53), and most were nested within either Cluster A (n = 18) or B (n = 11) (Figure 4A). A total of 24 nonstudents descended from the rest of the inferred transmission lineages (2–8 students each). Out of the 1191 genomes we sequenced from nonstudents, 96% (n = 1138) were not genetic descendants of detected clusters in students (Figure 4B). Larger clusters of students had greater numbers of nonstudent descendants (Figure 4C). The median age of nonstudents descending from Clusters A and B (IQR) was 47 (20–61) years. We do not have epidemiologic information on the association of these individuals with students, so it is not possible to determine the circumstances of transmission. It is also possible that student status was misclassified for these individuals or that some of these individuals were campus faculty or staff and therefore had differential exposure compared with the broader community.

DISCUSSION

We conducted a prospective study of SARS-CoV-2 genomic surveillance focused on a large public university and the surrounding community. A major strength of our study is high-density sampling of the student population and the surrounding community by sequencing all available specimens from 2 major testing laboratories. These data illustrate the rapid transmission of SARS-CoV-2 within a large public university population with remarkably little spread into the community.

Our analysis demonstrates that the COVID-19 epidemic among students at the University of Michigan–Ann Arbor was not derived from a single introduction. Most student cases in early Fall were derived from 1 of 2 dominant viral lineages, which coexisted for several weeks and circulated throughout several on-campus residences. We think it is unlikely that these 2 Pango lineages (B.1.1.304 and B.1.593) have

enhanced transmissibility or other notable intrinsic properties. They did not disseminate widely and do not exhibit the same mutations as other highly transmissible variants. The influx of new singleton introductions late in the semester may have been driven by "outside-in" transmission from elsewhere in the community as incidence spiked in Washtenaw County. This emphasizes the importance of reducing overall county and regional COVID-19 incidences, in addition to preventing outbreaks at IHE.

A key finding was that very few genomes from nonstudents were genetically linked to student clusters. Notably, lineages B.1.1.304 and B.1.593 have not persisted in Michigan since this time, providing additional support that these outbreaks did not spur widespread transmission in the community. We did not sequence viruses from every infection and therefore cannot exclude that other spillover events may have occurred. Nevertheless, given our notably high depth of sampling, it is unlikely that large outbreaks among students were the source of most COVID-19 infections in the community. If this was the case, we would expect to see vastly different patterns, particularly a higher frequency of B.1.1.304 and B.1.593 in nonstudents during November. Although there are differences in context, setting, and mitigation measures here compared with other IHE, this study suggests that previous findings with limited community surveillance may generalize more broadly [4, 6, 18].

There are other important limitations. First, we were not able to access specimens from commercial testing sites, and the number of detected transmission introductions is certainly an underestimate. Next, our study is not an epidemiologic investigation with contact tracing and individual behavioral information. It is difficult to reliably assess the effectiveness of any single mitigation measure from these data alone. Thorough contact tracing investigations of IHE-associated outbreaks with dense genomic surveillance across populations may be able to resolve these questions in greater detail. It is also possible that these dynamics could have played out differently with earlier emergence of a highly transmissible variant, such as B.1.1.7 [27, 32]. This work will be a valuable point of comparison for future studies examining the effects of more transmissible variants and vaccination on COVID-19 spread within IHE.

Our phylogenetic analysis of well-sampled genomic surveillance data provides insight into the spread of SARS-CoV-2 at a large public university in the United States. The small number of dominant lineages that circulated in the early and middle portions of the semester did not significantly contribute to the rise in county-level cases in November 2020. We emphasize that even rare transmission events can disproportionally impact vulnerable populations [17]. Additionally, SARS-CoV-2 infection can have severe clinical manifestations even in populations with generally lower risk [33, 34]. Therefore, it is critical that every effort be made to prevent and mitigate IHE-associated outbreaks in conjunction with measures at broader geographic levels.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

We thank Teodora Jevtic for assistance in obtaining individual metadata. We thank Nathan Grubaugh, Anderson Brito, and Simon Dellicour for helpful discussion on phylogenetic analysis. We thank the University of Michigan Clinical Microbiology Laboratory, COVID-19 Central Biorepository, and University Health Service for sharing specimens. We thank the University of Michigan Microbiome Core facility for their assistance in genome sequencing.

Financial support. This work was supported by the Centers for Disease Control and Prevention (contract 75D30120C09870 to T.C.F., D.H.O., and A.S.L.). The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. The funder had no role in the study design.

Potential conflicts of interest. All authors: no reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Prior presentations. This work has been presented as a poster at the Virus Genomics and Evolution Meeting (Wellcome Trust, UK; virtual) and at IDWeek 2021 (IDSA; virtual).

References

1. Andersen MS, Bento AI, Basu A, et al. College openings, mobility, and the incidence of COVID-19. medRxiv 2020.09.22.20196048 [Preprint]. 23 February **2021**. Available at: https://doi.org/10.1101/2020.09.22.20196048. Accessed 10 August 2021.
2. Leidner AJ. Opening of large institutions of higher education and county-level COVID-19 incidence — United States, July 6–September 17, 2020. MMWR Morb Mortal Wkly Rep **2021**; 70:14–19.
3. Vang KE. Participation in fraternity and sorority activities and the spread of COVID-19 among residential university communities — Arkansas, August 21–September 5, 2020. MMWR Morb Mortal Wkly Rep **2021**; 70:20–23.
4. Currie DW, Moreno GK, Delahoy MJ, et al. Description of a University COVID-19 outbreak and interventions to disrupt transmission, Wisconsin, August - October 2020. medRxiv 2021.05.07.21256834 [Preprint]. 10 May **2021**. Available at: https://doi.org/10.1101/2021.05.07.21256834. Accessed 10 August 2021.
5. Fox MD. Response to a COVID-19 outbreak on a University Campus — Indiana, August 2020. MMWR Morb Mortal Wkly Rep **2021**; 70:118–22.
6. Weil AA, Sohlberg SL, O'Hanlon JA, et al. SARS-CoV-2 epidemiology on a public university campus in Washington State. Open Forum Infect Dis **2021**:ofab464. doi:10.1093/ofid/ofab464
7. Wilson E. Multiple COVID-19 clusters on a university campus — North Carolina, August 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:1416–8.
8. Doyle K, Teran RA, Reefhuis J, et al. Multiple variants of SARS-CoV-2 in a university outbreak after spring break - Chicago, Illinois, March-May 2021. MMWR Morb Mortal Wkly Rep **2021**; 70:1195–200.
9. Bradley EH, An M-W, Fox E. Reopening colleges during the coronavirus disease 2019 (COVID-19) pandemic—one size does not fit all. JAMA Netw Open **2020**; 3:e2017838.
10. Paltiel AD, Zheng A, Walensky RP. Assessment of SARS-CoV-2 screening strategies to permit the safe reopening of college campuses in the United States. JAMA Netw Open **2020**; 3:e2016818.
11. Yamey G, Walensky RP. Covid-19: re-opening universities is high risk. BMJ **2020**; 370:m3365.
12. Bharti N, Lambert B, Exten C, et al. Large university with high COVID-19 incidence did not increase risk to non-student population. medRxiv 2021.04.27.21255023 [Preprint]. 29 April **2021**. Available at: https://doi.org/10.1101/2021.04.27.21255023. Accessed 10 August 2021.

13. Lu H, Weintz C, Pace J, et al. Are college campuses superspreaders? A data-driven modeling study. Comput Methods Biomech Biomed Eng **2021**; 24:1136–45.
14. Martin MA, VanInsberghe D, Koelle K. Insights from SARS-CoV-2 sequences. Science **2021**; 371:466–7.
15. Moreno GK, Braun KM, Riemersma KK, et al. Revealing fine-scale spatiotemporal differences in SARS-CoV-2 introduction and spread. Nat Commun **2020**; 11:5558.
16. da Silva Filipe A, Shepherd JG, Williams T, et al; COVID-19 Genomics UK (COG-UK) Consortium. Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland. Nat Microbiol **2021**; 6:112–22.
17. Richmond CS, Sabin AP, Jobe DA, et al. SARS-CoV-2 sequencing reveals rapid transmission from college student clusters resulting in morbidity and deaths in vulnerable populations. medRxiv 2020.10.12.20210294 [Preprint]. 14 October **2020**. Available at: https://doi.org/10.1101/2020.10.12.20210294. Accessed 10 August 2021.
18. Arnold CRK, Srinivasan S, Herzog CM, et al. SARS-CoV-2 seroprevalence in a university community: a longitudinal study of the impact of student return to campus on infection risk among community members. medRxiv 2021.02.17.21251942 [Preprint]. 17 September **2021**. Available at: https://doi.org/10.1101/2021.02.17.21251942. Accessed 10 August 2021.
19. Aggarwal D, Warne B, Jahun AS, et al. Genomic epidemiology of SARS-CoV-2 in a UK university identifies dynamics of transmission. ResSquare rs-520627 [Preprint]. 19 May **2021**. Available at: https://www.researchsquare.com/article/rs-520627/v1. Accessed 10 August 2021.
20. Valesano AL, Rumfelt KE, Dimcheff DE, et al. Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. PLoS Pathog **2021**; 17:e1009499.
21. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv1303.3997 Q-Bio. **2013**. Available at: http://arxiv.org/abs/1303.3997. Accessed 3 September 2020.
22. Grubaugh ND, Gangavarapu K, Quick J, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol **2019**; 20:8.
23. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics **2018**; 34:4121–3.
24. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **2015**; 32:268–74.
25. Sagulenko P, Puller V, Neher RA. TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol **2018**; 4:vex042.
26. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol **2016**; 2:vew007.
27. Alpert T, Brito AF, Lasek-Nesselquist E, et al. Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. Cell **2021**; 184:2595–604.e13.
28. Drummond AJ, Rambaut A. BEAST: Bayesian Evolutionary Analysis by Sampling Trees. BMC Evol Biol **2007**; 7:214.
29. Lemieux JE, Siddle KJ, Shaw BM, et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science. **2021**; 371:eabe3261. doi:10.1126/science.abe3261
30. Müller NF, Wagner C, Frazar CD, et al. Viral genomes reveal patterns of the SARS-CoV-2 outbreak in Washington State. Sci Transl Med. **In press**.
31. du Plessis L, McCrone JT, Zarebski AE, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science **2021**; 371:708–12.
32. Washington NL, Gangavarapu K, Zeller M, et al. Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. Cell **2021**; 184:2587–94.e7.
33. Logue JK, Franko NM, McCulloch DJ, et al. Sequelae in adults at 6 months after COVID-19 infection. JAMA Netw Open **2021**; 4:e210830.
34. Tenforde MW. Symptom duration and risk factors for delayed return to usual health among outpatients with COVID-19 in a multistate health care systems network — United States, March–June 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:993–8.