Article

Linking contact tracing with genomic surveillance to deconvolute SARS-CoV-2 transmission on a university campus

Jacquelyn Turcinovic, Kayla Kuhfeldt, Madison Sullivan, ..., Catherine Klapperich, Hannah E. Landsberg, John H. Connor

jhconnor@bu.edu

Highlights

Contact tracing and sequencing provide more information than either approach alone

Primary exposures in an athletic group occurred outside structured athletic events

Genomic and contact tracing data can inform effective public health decisions

iScience

Article

Linking contact tracing with genomic surveillance to deconvolute SARS-CoV-2 transmission on a university campus

Jacquelyn Turcinovic,[1,2,3] Kayla Kuhfeldt,[4] Madison Sullivan,[4] Lena Landaverde,[5,6,7] Judy T. Platt,[4] Lynn Doucette-Stamm,[7] William P. Hanage,[8] Davidson H. Hamer,[2,6,9,10,11] Catherine Klapperich,[5,6] Hannah E. Landsberg,[4] and John H. Connor[1,2,3,11,12,]*

SUMMARY

Contact tracing and genomic data, approaches often used separately, have both been important tools in understanding the nature of SARS-CoV-2 transmission. Linked analysis of contact tracing and sequence relatedness of SARS-CoV-2 genomes from a regularly sampled university environment were used to build a multilevel transmission tracing and confirmation system to monitor and understand transmission on campus. Our investigation of an 18-person cluster stemming from an athletic team highlighted the importance of linking contact tracing and genomic analysis. Through these findings, it is suggestive that certain safety protocols in the athletic practice setting reduced transmission. The linking of traditional contact tracing with rapid-return genomic information is an effective approach for differentiating between multiple plausible transmission scenarios and informing subsequent public health protocols to limit disease spread in a university environment.

INTRODUCTION

SARS-CoV-2, the causative agent of COVID-19, has posed significant risks to typical university campus functions including classroom teaching, congregate living in dormitories, and athletic events. After an initial pause to in-person instruction in the spring of 2020, many colleges and universities began a return to in-person classes and certain activities in the fall of that year. This return to on-campus life has been successful overall but has been associated with repeated SARS-CoV-2 infection clusters and outbreaks of varying sizes (Doyle et al., 2021; Wilson et al., 2020). Some of the more dramatic clusters have led universities to revert to remote instruction with cessation of on-campus activities (Wan et al., 2022; Wilson et al., 2020).

Universities have taken multiple approaches to limit SARS-CoV-2 superspreading events. These have included limited in-person instruction, mask mandates, heating, ventilation, and air conditioning improvements, regular surveillance testing, and vaccination requirements (Denny et al., 2020; Hamer et al., 2021; Pollock et al., 2021). With the rollout of SARS-CoV-2 vaccination, the risk of severe disease and death following SARS-CoV-2 infection has lessened, but breakthrough infections and transmission continue across the United States and variant sweeps have had major impacts within university populations (Petros et al., 2022; Wan et al., 2022). Various mitigation strategies have been developed to prevent on-campus transmission in classrooms, workplaces, dormitories, and university-sponsored activities like athletic competitions and commencement (Hamer et al., 2021). These strategies have included traditional case investigation and bidirectional contact tracing, the process of identifying who exposed a case in addition to who the case exposed, to provide insight into factors associated with transmission and defining areas where transmission is minimal (Currie et al., 2021; Kuhfeldt et al., 2022; Rebmann et al., 2021).

In response to the COVID-19 pandemic, Boston University (BU) developed and implemented a robust prevention, surveillance, and control strategy to allow most students, faculty, and staff to return to campus and participate in near-normal educational activities. These multifaceted strategies, which included a robust in-house contact tracing effort, were successful at limiting disease transmission in the fall of 2020 (Hamer et al., 2021). In 2021, the University added sequencing of all SARS-CoV-2-positive samples from BU campus

[1]Department of Microbiology, Boston University School of Medicine, Boston, MA 02118, USA

[2]National Emerging Infectious Diseases Laboratories, Boston University, Boston, MA 02118, USA

[3]Program in Bioinformatics, Boston University, Boston, MA 02215, USA

[4]Student Health Services, Boston University, Boston, MA 02215, USA

[5]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

[6]Precision Diagnostics Center, Boston University, Boston, MA 02215, USA

[7]BU Clinical Testing Laboratory, Research Department, Boston University, Boston, MA 02215, USA

[8]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[9]Department of Global Health, Boston University School of Public Health, Boston, MA 02118, USA

[10]Section of Infectious Disease, Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA

[11]Center for Emerging Infectious Disease Policy and Research, Boston University, Boston, MA 02118, USA

[12]Lead contact

*Correspondence:
jhconnor@bu.edu

https://doi.org/10.1016/j.isci.2022.105337

surveillance testing. This sequencing brought important visibility to the level of variant of concern (VoC) transmission on the campus. In addition, the sequencing data opened a new opportunity for genomics-aided epidemiology.

Linked genomic and contact tracing have been deployed during this pandemic to effectively assess transmission chains, including those in different athletic settings (Jang et al., 2020; Moreno et al., 2021; Teran et al., 2020). We began integrating traditional case contact tracing investigations with genomic sequencing to broadly confirm contact-tracing-identified links, eliminate others, and identify transmission events not initially identified through contact tracing. These efforts were especially informative in helping to identify likely transmission drivers in athletic interactions, an area where prior work has suggested significant transmission potential prior to the implementation of widespread vaccination. Here, we describe the linking of traditional contact tracing with molecular epidemiology to investigate an athletics-associated cluster of 18 cases identified from a fully vaccinated team.

RESULTS

COVID-19 vaccination was mandated for student attendance and faculty/staff employment on the Boston University campus as of September 2, 2021, with allowances for religious or medical exemption. Vaccination compliance at BU was defined as either completion of a SARS-CoV-2 vaccination series, or as a documented medical or religious exemption. Full vaccination rates for faculty and students were 98.5% and 98.7%, respectively.

During the fall 2021 semester, SARS-CoV-2 transmission was active. SARS-CoV-2 transmission in the Boston area was significant over the study period (7-day rolling average in Suffolk County between 100 and 200 cases per day, in Middlesex: 200–600 per day). Within the University campus, BU mandated regular weekly testing of the entire on-campus population as a means of maintaining real-time information about campus-related infections (Hamer et al., 2021). Approximately 4,500 COVID-19 surveillance diagnostic tests were run daily during the period between August 1 and October 31, 2021. This testing identified 587 SARS-CoV-2-positive samples from a pool of 414,000 total tests (0.14%).

Within the pool of 587 positives, there were several identified instances of clustered transmission. Clustered transmission was defined as a series of infections with initial evidence of linkage by contact tracing. Twelve instances of clustered transmission were defined in this time ranging from 2 to 18 individuals. The largest cluster of transmission between the August to November time frame was initially identified as an 18-individual outbreak that spanned 14 days. Contact tracing rapidly identified an athletics affiliation where more than 50 individuals had regular close contact through team practices and games. No athletic employees affiliated with the practice or games tested positive. Interviews with positive individuals identified 4 key events as potential sources of transmission: athletic practices, a university-sponsored volunteer event, and two non-university-sponsored social gatherings (Figure 1A).

Out of the 18 cases, 13 were identified as having had a potential household exposure. There was also a large non-university-sponsored social gathering among approximately 20 individuals of the athletic team (Team 1) with 20–30 individuals of another athletic team (Team 2). Additionally, there were multiple smaller group social interactions among the 18 individuals, including unmasked car rides, indoor dining, and dorm or apartment socialization, including watching television and playing video games in close proximity.

Athletic practice

The contact tracing cluster involved 16 cases from one team (Team 1) and 2 cases from a second athletic team (Team 2). The teams did not practice together. Team 1 athletic practices consisted of outdoor practices, indoor conditioning sessions, and team video review sessions. Practices were held outdoors 2 times a week. Masks were either not worn consistently or at all during these outdoor practices. Conditioning sessions of 20–50 individuals were held indoors 4 times a week for approximately 1 h each, and masks were not always worn during these sessions. Finally, video review sessions lasting 25–30 min for approximately three times per week were held in a smaller, indoor room. Masks were worn during video review sessions. Online video review options were offered. During the second week of the cluster, all video review sessions were moved online. Athletic practices included full team use of the locker rooms before and after practices. In the locker rooms, there was no social distancing, and masks were typically worn but removed while
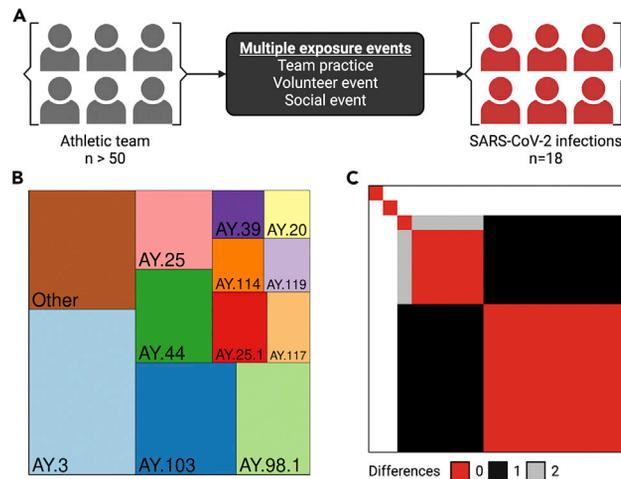
Figure 1. Initial cluster investigation and background genetic diversity
Contact tracing identified 18 COVID-19 cases linked to two athletic teams at Boston University (A). Sequencing of all positive COVID-19 surveillance tests at BU showed substantial background genetic diversity despite classification of all samples as variant of concern Delta (B). Pairwise comparison following viral genome sequencing (C) showed that 16 of the 18 had highly similar viral genomes (≤2 nucleotide changes). The final 2 cases were not genetically linked (≥3 nucleotide changes) and were eliminated from the transmission cluster.

showering. The two separate athletic teams did not interact with one another at practice, conditioning, or in the locker rooms.

Volunteer event

In addition to practice, another potential source of transmission was a university-sponsored volunteer event for Team 1. This event was held off campus, and many individuals traveled together in private cars for longer than 30 min without masks. This event was mostly outdoors, and individuals were separated throughout the day. Mask use was minimal while outdoors.

Social gatherings

A third potential source of transmission was a non-university-sponsored athletic social gathering involving both Team 1 and Team 2. During this event, individuals socialized indoors and outdoors, and masks were not worn during this time. The event lasted 4–5 h, with people mingling in close proximity throughout the event space. Two individuals from Team 2 who were subsequently positive following the event had extended interactions before and after this gathering with cases from Team 1.

The fourth event was a non-university social event at an area nightclub. This event linked two cases from Team 1 and 2 non-athletic-affiliated cases. This event was indoors in a crowded space for multiple hours. Masks were not worn.

Genomic analysis of the cluster

Whole genome sequencing of SARS-CoV-2 was attempted on all positive samples from this cluster, as well as on all positive BU tests during the study period. All 18 genomes initially identified as part of this cluster were successfully sequenced, with an average coverage of >200 reads at each nucleotide of the genome. All samples were identified as belonging to Delta sublineage AY.3 by Pangolin analysis. AY.3 was one of more than a dozen lineages present on campus (Figure 1B). It was the most common sequenced lineage but constituted less than 20% of total sequenced cases.

Pairwise comparison (Figure 1C) of all 18 genomes showed that the genomes from individuals initially associated with the outbreak had differing levels of similarity. There were two highly similar sets of identical AY.3 lineage genomes (A and B) that together accounted for 15 of the 18 cases. Set A was made up of 5 genomes; set B was composed of 10. Groups A and B differed by a single substitution (nucleotide 24,410 G > A) that created a D950N amino acid change in the spike protein. One of the remaining 3
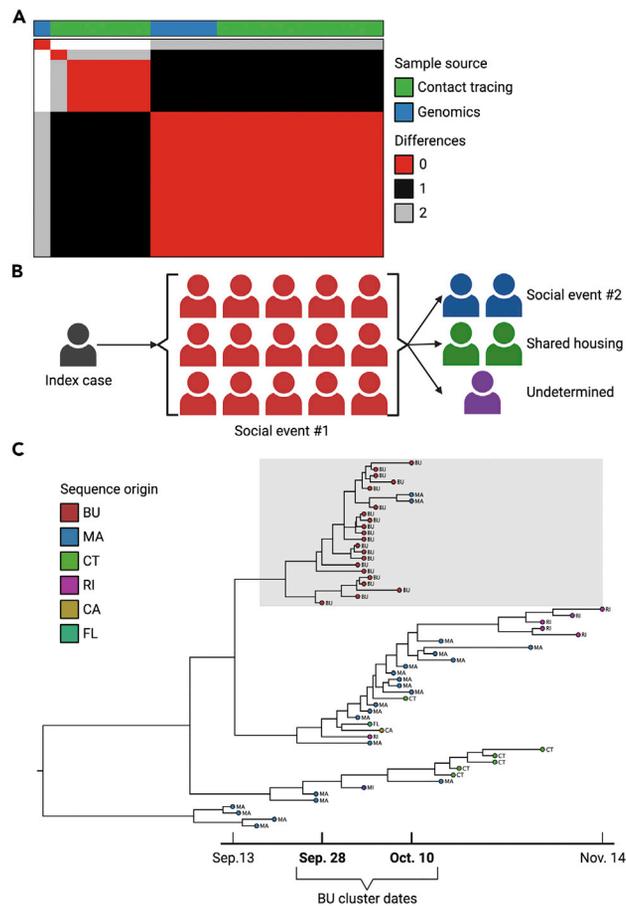
Figure 2. Additional cases in cluster and community time tree
Pairwise comparison heatmap of all genetically clustered cases (A). Schematic of likely transmission chain (B). A Bayesian time tree incorporating community samples shows it is probable that the BU cases are a transmission cluster (C).

genomes differed from set B by a single nucleotide change as well. The extremely high relatedness of these 16 viral genomes indicated that they were all part of the same interaction-related transmission chain.

Importantly, two of the 18 cases that were originally identified as part of the outbreak appeared to be unrelated to the main transmission cluster. These cases were pivotal in understanding potential transmission networks. Both individuals shared all practice/game interactions and tested positive at times that were epidemiologically consistent with the main outbreak. The genomes from these individuals differed from the others by ≥ 3 and ≥ 25 nucleotide changes and were classified as lineages AY.3 and AY.25.1.1, respectively. The significant number of mutations compared to the other samples within the outbreak suggests acquisition outside of the cluster (Bendall et al., 2022). Sequence analysis of all positive cases on the BU campus showed that neither of these genomes was transmitted to others within the team or elsewhere on campus. This suggested that clustered transmission was unlikely to be tied to athletic practices themselves but to outside factors. If athletics-associated interactions had been a significant driver of transmission, multiple clusters from each of the lineages identified would have been expected.

Identification of additional transmission links

To examine the possibility that there were additional transmission events not originally identified as part of contact tracing interviews, we examined whether any of the 290 genomes sequenced between September 13 and October 25, 2021, as part of our surveillance sequencing efforts were related to this cluster. Pairwise comparison identified five cases of interest that had not been originally identified as part of the cluster (Figure 2A). This information prompted contact tracing reinvestigation of interactions of these cases. Two non-athletic affiliated individuals exclusively had only dorm interactions with one individual in the cluster from

Team 1. Two additional identical genomes tracked to individuals that had been involved in an extended unmasked social interaction with members of the main cluster. These cases are consistent with the hypothesis that transmission was associated with longer term, repeated social interactions.

While most of the cases that showed high genomic relatedness were associated with reported long-term interactions with other positive cases, this was not universal. For the fifth sequencing-identified genome in the expanded search, the individual reported no clear epidemiological link of living space, classroom, or known social encounters among the remaining samples could be identified. This individual did report using public transportation on the same subway line as the other individuals within the cluster, which is a possible source of transmission (Figure 2B).

Comparison to community cases

We also investigated whether any of the 21 genetically linked cases in our cluster could be linked to other sequenced cases in the community. Using the GISAID database, we identified 37 samples with ≤2 nucleotide differences from at least one sample in our cluster. Construction of a phylogenetic time tree using these sequences and those associated with the BU outbreak identified multiple different likely introduction/spread among these samples (Figure 2C). All 25 BU cases clustered together on one arm of this tree with a second distinct arm indicated a large cluster with samples from MA/CT/RI/FL and CA. There were two additional smaller clusters associated with MA and MA/CT cases, respectively. Interestingly, 2 MA sequences were associated with the BU to the BU cluster suggesting a small amount of transmission outside of our recognized cluster, but there was no evidence for ongoing transmission after October 10 (Figure 2C).

DISCUSSION

Through our combined genomic and epidemiological analysis, we refined an initial cluster of 18 cases thought to be associated with athletic events to a final 21-case cluster of infections with the variant of concern Delta, Pangolin lineage AY.3. All individuals had been fully vaccinated with a WHO-listed vaccine. Our findings suggest a possible evolution of SARS-CoV-2 infection risk from early stages of the pandemic. Reports that covered disease transmission prior to the appearance of widespread vaccination highlight the significant propensity for transmission in athletic or workout settings. This included documenting transmission in gymnastics facilities studios (Dougherty et al., 2021), through soccer practice (Teran et al., 2020) and athletic competitions (Moreno et al., 2021). Our analysis of the cluster described here were less consistent with an at-practice or at-competition disease transmission scenario.

Our results are mostly consistent with the hypothesis that SARS-CoV-2 transmission between vaccinated individuals was associated with extended, close contact exposure outside of the practice environment. This is supported by the presence of SARS-CoV-2-positive individuals in the practice environment that did not appear to transmit to others within that same environment. Instead, viral genomics tracked transmission to individuals that had social contact outside of athletic events and to individuals sharing housing with infected individuals. One individual who reportedly only attended athletics events with Team 1 and no additional social exposures outside of his household was originally identified as part of the epidemiologic cluster but later eliminated based on genomic analysis. Moreover, the additional genetically linked cases included those with prolonged, unmasked social exposure. This indicated that the athletic association of the original 18 cases was not a strong driver of transmission risk. Instead, repeated and prolonged unmasked contact outside of athletic activities correlated most effectively with transmission.

In our efforts to understand factors driving transmission of SARS-CoV-2 within a university setting, we found that the combination of comprehensive contact tracing and all-campus viral genomic surveillance was synergistic in providing information on transmission scenarios in a highly vaccinated population. An initial assessment of transmission scenarios based on contact tracing alone was equivocal in establishing the cluster described here, as there were multiple potential transmission site possibilities. The addition of viral genomics to this cluster investigation provided important clarification on the source of transmission, and the ability to sequence-search all other cases at the university during this time added additional insights beyond cluster-focused sequencing.

The initiation of both bidirectional contact tracing and genomic sequence analysis at the same time is not a constant in university disease surveillance but offers distinct advantages. When surveillance or symptomatic testing leads to contact tracing first, followed by requests for whole genome sequencing

on high-interest samples (i.e., targeted sequencing), it is often too late to ensure sample quality and/or adequate sample volume for successful sequencing. Additionally, sequencing only samples associated with an epi-identified outbreak can result in incomplete identification of the full scope of transmission events. From the other end, surveillance sequencing is often difficult to connect to contact tracing since these efforts often span multiple institutions with different freedom to acquire or analyze samples of interest. The ability to initiate both contact tracing and genomic sequencing from a central-ized testing system can be critical to understanding viral transmission when done proactively and synonymously.

Limitations of the study

Delays in both reinvestigation of cases and sequencing results on samples can challenge the ability to make real-time decisions for public health strategies. Delays in reinvestigation may have led to incomplete recall from a case. For example, the university volunteer event had minimal information regarding travel to the event in private cars. In addition, the personnel and resources required to complete reinvestigations for all linked samples may be limiting in some university settings. Further data analysis on other athletic team interactions at the university may have improved the analysis to put the findings in the context of col-legiate athletics as a whole instead of two teams. However, we believe that the insight into transmission links provided by looking into even a limited number of clusters adds critical insight into the behavior of the virus. The ability to have answers on the social context and the biologic context of the virus outweighs the limitations presented by the approach.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - COVID-19 PCR testing
  - Contact tracing
  - SARS-CoV-2 sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Sequence data processing
  - Pairwise comparison of SARS-CoV-2 genomes
  - Time tree with community sequences
  - Linkage of contact tracing and genomics

AUTHOR CONTRIBUTIONS

J.T. and J.H.C. conducted the genomic sequencing and analysis. M.S., K.K., and H.L. conducted the case investigation and contact tracing analysis. L.D.S. oversaw the COVID-19 PCR testing and workflow between the clinical testing laboratory and the NEIDL. L.L., J.T.P., C.K., W.P.H., and D.H. reviewed the process, anal-ysis, and manuscript writing.

DECLARATION OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Aksamentov, I., Roemer, C., Hodcroft, E., and Neher, R. (2021). Nextclade: clade assignment, mutation calling and quality control for viral genomes. J. Open Source Softw. 6, 3773. https://doi.org/10.21105/JOSS.03773.

Bendall, E., Paz-Bailey, G., Santiago, G.A., Porucznik, C.A., Stanford, J.B., Stockwell, M.S., Duque, J., Jeddy, Z., Veguilla, V., Major, C., et al. (2022). SARS-CoV-2 genomic diversity in households highlights the challenges of sequence-based transmission inference. Preprint at medRxiv. https://doi.org/10.1101/2022.08.09.22278452.

Currie, D.W., Moreno, G.K., Delahoy, M.J., Pray, I.W., Jovaag, A., Braun, K.M., Cole, D., Shechter, T., Fajardo, G.C., Griggs, C., et al. (2021). Interventions to disrupt coronavirus disease transmission at a university, Wisconsin, USA, August–October 2020 - volume 27, number 11—November 2021 - emerging infectious diseases journal - CDC. Emerg. Infect. Dis. 27, 2776–2785. https://doi.org/10.3201/EID2711.211306.

Denny, T.N., Andrews, L., Bonsignori, M., Cavanaugh, K., Datto, M.B., Deckard, A., DeMarco, C.T., DeNaeyer, N., Epling, C.A., Gurley, T., et al. (2020). Implementation of a pooled surveillance testing program for asymptomatic SARS-CoV-2 infections on a college campus — Duke university, Durham, North Carolina, August 2–October 11, 2020. MMWR Morb. Mortal. Wkly. Rep. 69, 1743–1747. https://doi.org/10.15585/MMWR.MM6946E1.

Dougherty, K., Mannell, M., Naqvi, O., Matson, D., and Stone, J. (2021). SARS-CoV-2 B.1.617.2 (Delta) variant COVID-19 outbreak associated with a gymnastics facility — Oklahoma, April–may 2021. MMWR Morb. Mortal. Wkly. Rep. 70, 1004–1007. https://doi.org/10.15585/MMWR.MM7028E2.

Doyle, K., Teran, R.A., Reefhuis, J., Kerins, J.L., Qiu, X., Green, S.J., Choi, H., Madni, S.A., Kamal, N., Landon, E., et al. (2021). Multiple variants of SARS-CoV-2 in a university outbreak after spring break — Chicago, Illinois, March–may 2021. MMWR Morb. Mortal. Wkly. Rep. 70, 1195–1200. https://doi.org/10.15585/MMWR.MM7035A3.

Hamer, D.H., White, L.F., Jenkins, H.E., Gill, C.J., Landsberg, H.E., Klapperich, C., Bulekova, K., Platt, J., Decarie, L., Gilmore, W., et al. (2021). Assessment of a COVID-19 control plan on an urban university campus during a second wave of the pandemic. JAMA Netw. Open 4, e2116425. https://doi.org/10.1001/JAMANETWORKOPEN.2021.16425.

Jang, S., Han, S.H., and Rhee, J.Y. (2020). Cluster of coronavirus disease associated with fitness dance classes, South Korea - volume 26, number 8—August 2020 - emerging infectious diseases journal - CDC. Emerg. Infect. Dis. 26, 1917–1920. https://doi.org/10.3201/EID2608.200633.

Khare, S., Gurry, C., Freitas, L., Schultz, M.B., Bach, G., Diallo, A., Akite, N., Ho, J., Lee, R.T., Yeo, W., et al. (2021). GISAID's role in pandemic response. China CDC Wkly. 3, 1049–1051. https://doi.org/10.46234/CCDCW2021.255.

Kuhfeldt, K., Turcinovic, J., Sullivan, M., Landaverde, L., Doucette-Stamm, L., Hamer, D.H., Platt, J., Klapperich, C., Landsberg, H.E., and Connor, J.H. (2022). Minimal SARS-CoV-2 classroom transmission at a large urban university experiencing repeated into campus introduction. Preprint at medRxiv.

Landaverde, L., McIntyre, D., Robson, J., Fu, D., Ortiz, L., Chen, R., Oliveira, S.M.D., Fan, A., Barrett, A., Burgay, S.P., et al. (2022). Buildout and integration of an automated high-throughput CLIA laboratory for SARS-CoV-2 testing on a large urban campus. SLAS Technol. 27, 302–311. https://doi.org/10.1016/J.SLAST.2022.06.003.

Primer-schemes/nCoV-2019/V4 at master · artic-network/primer-schemes https://github.com/artic-network/primer-schemes/tree/master/nCoV-2019/V4

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Moreno, G.K., Braun, K.M., Pray, I.W., Segaloff, H.E., Lim, A., Poulsen, K., Meiman, J., Borcher, J., Westergaard, R.P., Moll, M.K., et al. (2021). Severe acute respiratory syndrome coronavirus 2 transmission in intercollegiate athletics not fully mitigated with daily antigen testing. Clin. Infect. Dis. 73, S45–S53. https://doi.org/10.1093/CID/CIAB343.

Petros, B.A., Turcinovic, J., Welch, N.L., White, L.F., Kolaczyk, E.D., Bauer, M.R., Cleary, M., Dobbins, S.T., Doucette-Stamm, L., Gore, M., et al. (2022). Early introduction and rise of the Omicron severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variant in highly vaccinated university populations. Clin. Infect. Dis. 13, ciac413. https://doi.org/10.1093/CID/CIAC413.

Pollock, B.H., Kilpatrick, A.M., Eisenman, D.P., Elton, K.L., Rutherford, G.W., Boden-Albala, B.M., Souleles, D.M., Polito, L.E., Martin, N.K., and Byington, C.L. (2021). Safe reopening of college campuses during COVID-19: the University of California experience in Fall 2020. PLoS One 16, e0258738. https://doi.org/10.1371/JOURNAL.PONE.0258738.

Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., and Pybus, O.G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5, 1403–1407. https://doi.org/10.1038/s41564-020-0770-5.

Rebmann, T., Loux, T.M., Arnold, L.D., Charney, R., Horton, D., and Gomel, A. (2021). SARS-CoV-2 transmission to masked and unmasked close contacts of university students with COVID-19 — St. Louis, Missouri, January–may 2021. MMWR Morb. Mortal. Wkly. Rep. 70, 1245–1248. https://doi.org/10.15585/MMWR.MM7036A3.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4, vey016. https://doi.org/10.1093/VE/VEY016.

Teran, R.A., Ghinai, I., Gretsch, S., Cable, T., Black, S.R., Green, S.J., Perez, O., Chlipala, G.E., Maienschein-Cline, M., Kunstman, K.J., et al. (2020). COVID-19 outbreak among a university's Men's and women's soccer teams — Chicago, Illinois, July–August 2020. MMWR Morb. Mortal. Wkly. Rep. 69, 1591–1594. https://doi.org/10.15585/MMWR.MM6943E5.

Wan, J., Cazer, C.L., Clarkberg, M.E., Henderson, S.G., Lee, S.E., Meredith, G., Osman, M., Shmoys, D.B., and Frazier, P.I. (2022). Boosters protect against SARS-CoV-2 infections in young adults during an Omicron-predominant period. Preprint at medRxiv. https://doi.org/10.1101/2022.03.08.22272056.

Wilson, E., Donovan, C.v., Campbell, M., Chai, T., Pittman, K., Seña, A.C., Pettifor, A., Weber, D.J., Mallick, A., Cope, A., et al. (2020). Multiple COVID-19 clusters on a university campus — North Carolina, August 2020. MMWR Morb. Mortal. Wkly. Rep. 69, 1416–1418. https://doi.org/10.15585/MMWR.MM6939E3.

Wilm, A., Aw, P.P.K., Bertrand, D., Yeo, G.H.T., Ong, S.H., Wong, C.H., Khor, C.C., Petric, R., Hibberd, M.L., and Nagarajan, N. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 40, 11189–11201. https://doi.org/10.1093/nar/gks918.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
COVIDSeq Assay (96 samples)	Illumina	Cat# 20049393
COVIDSeq v4 Primer Pools	Illumina	Cat# 20065135
Quick-RNA Viral 96 Kit	Zymo Research	Cat# R1041
MagMAX™ Viral/Pathogen II (MVP II) Nucleic Isolation Kit	ThermoFisher	Cat# A48383
Applied Biosystems TaqPath Master Mix no ROX	ThermoFisher	Cat# A28523
CDC 2019-nCoV Real-Time RT-PCR Primers and Probes	IDT	Custom; see (Landaverde et al., 2022)
Deposited data		
Raw sequencing data	NCBI Bioproject	Bioproject: PRJNA867488
Community consensus sequences	GISAID EPI_SET	EPI_SET: EPI_SET_220804bm
Software and algorithms		
Bowtie2	bowtie-bio.sourceforge.net	v2.3.4.1
Samtools	htslib.org	v1.15.1
LoFreq	csb5.github.io/lofreq	v2.1.3.1
R	r-project.org	v4.0.2
Pangolin	pangolin.cog-uk.io	v4.1.1
Pango data	pangolin.cog-uk.io	v1.11
tidyverse	tidyverse.org	v1.3.0
argparse	github.com/trevorld/r-argparse	v2.1.5
BEAST	github.com/beast-dev/beast-mcmc	v1.10.4

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, John H. Connor (jhconnor@bu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw sequencing data has been deposited in the NCBI Sequence Read Archive are publicly available as of the date of publication. Accession numbers are listed in the key resources table.

- All original code has been deposited in GitHub and is publicly available as of the date of publication. DOIs are listed in the key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

The BU Charles River Campus Institutional Review Board reviewed and approved the protocol for sequencing to determine variant distribution and emergence on the BU campus (protocol #5693E).

COVID-19 PCR testing

BU students, faculty, and staff performed routine testing up to twice a week and no less than once per week. Anterior nares samples were self-collected in sterile saline, and following clinical testing, the residual saline from positive samples was stored at -80°C. We identified all individuals with detectable SARS-CoV-2 from September 13 to October 25, 2021 by reverse transcriptase real-time PCR (rRT-PCR) using the Centers for Disease Control and Prevention (CDC) primers for N1, N2, and RNase P according to the diagnostic methods described in (Landaverde et al., 2022).

Contact tracing

Case investigation and bidirectional contact tracing were conducted for each case by BU contact tracers following adapted CDC and Massachusetts Department of Public Health (MDPH) protocols. Case investigations asked targeted questions about travel, housing, symptoms, large gatherings, public transportation, and on campus affiliations. Bidirectional contact tracing was used to identify where a case may have been exposed in the past 5 days in addition to who they may have exposed during their infectious period. In addition to traditional contact tracing, contact tracers utilized internal data monitoring systems such as class rosters and event registration systems to identify clusters quickly based on academic, living or student group affiliation and inform individuals of their potential exposure. Contact tracing was used to: 1) identify potential at-work, in-class, or school-sponsored event transmission; and 2) determine the size and likely duration of clustered transmission events.

SARS-CoV-2 sequencing

Total RNA from residual discarded saline from SARS-CoV-2 positive swabs collected during the study window was extracted using the Zymo Research Quick-RNA Viral 96 Kit according to the manufacturer's protocol. The ARTIC v4 primer set ("primer-schemes/nCoV-2019/V4 at master · artic-network/primer-schemes," n.d.) and Illumina COVIDSeq Assay kit were then used to amplify and pool the viral genome according to the manufacturer's protocol. The resulting libraries were sequenced on an Illumina NextSeq500.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sequence data processing

Raw reads were aligned to the Wuhan-Hu-1 reference sequence (NC_045512.2) using Bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012), and primers were soft-clipped using the *ampliconclip* tool from samtools v1.15.1 (Li et al., 2009). Nucleotide substitutions, insertions, and deletions were identified with LoFreq v2.1.3.1 (Wilm et al., 2012) with a minimum coverage of 10 reads, and single nucleotide variants (SNVs) at >0.5 frequency were used to assemble consensus sequences using a custom R script (v4.0.2). Lineage assignment for each genome was carried out using Pangolin v4.1.1 and PangoData v1.11 (Rambaut et al., 2020). All processing and analysis scripts are available on GitHub (https://github.com/neidl-connor-lab/athletic-cluster).

Pairwise comparison of SARS-CoV-2 genomes

Pairwise comparisons were performed in R v4.0.2 using LoFreq VCF files. To reduce sequence quality noise, SNVs in the 3′ and 5′ UTRs (bases <265 and >29,675) of each genome were removed. All samples with >7000 low-coverage bases were dropped. Genome relatedness was quantified by counting the number of nucleotide differences between genomes. Each genome was modeled as a binary SNV vector spanning all nucleotide changes found at >0.5 frequency in the dataset. Presence of an SNV in >50% of the aligned reads was coded as a one; otherwise, zero. After pairwise comparison identical (0 differences), similar (1–2 differences), and unrelated ($\geq$3 differences) sequence sets were defined.

Time tree with community sequences

A cluster of 21 samples with 4 unique genotypes was identified. Each genotype was used to query the GISAID database (Khare et al., 2021) using Audacity*Instant* v3.08 to identify other 37 potentially related samples spanning September 13 to October 25, 2021 (EPI_SET_220804bm). The 3′ and 5′ UTRs were clipped from all 58 samples, and multiple sequence alignment was performed with Nextstrain Nextalign (Aksamentov et al., 2021). A Bayesian time tree was constructed using BEAST v1.10.4 (Suchard et al., 2018).

Linkage of contact tracing and genomics

To complement contact tracing, a positive PCR test also triggered genomic analysis. Sequence relatedness was determined using pairwise comparison as described above. Samples with identical genomes (0 differences) were considered part of linked transmission, while those with 1 or 2 nucleotide differences were considered likely and possible, respectively. Genomes that differed by $\geq$3 nucleotides across >29,000 bases were not considered linked. Genomic and contact tracing information were combined during transmission tracking events using de-identified coding and a shared database. Sequences identified as related were prioritized for case investigation by the contact tracing team. Conversely, sequencing prioritized samples identified as clusters from the contact tracing team.