

The Molecular Biology of SARS-CoV-2 and its Evolution within a College Community

Jesse Brewer, BA; Qiqi Yang, MS; Irina Aranovich, MS; Wei Wang, PhD;
Daniel Notterman, MD, MA; Irini Daskalaki, MD
Princeton University, Princeton, New Jersey



Jesse Brewer, BA
Princeton University
Princeton, NJ 08544
609.937.5074 (phone)
Email: jtbrewer22@gmail.com
For information about COVID Resources at Princeton, please visit <https://covid.princeton.edu/>

Abstract
SARS-CoV-2, a novel and highly pathogenic coronavirus, has caused unprecedented global disruption following its introduction into the human population. Beginning in January 2021, a NJ university invited all students to campus and initiated an asymptomatic testing protocol using weekly to twice-weekly PCR-based detection of human saliva samples. RNA extracted from PCR-positive human saliva samples was sequenced for surveillance purposes.

Methods
Positive samples were submitted for RNA-Seq analysis (ARTIC amplicon sequencing protocol, illumina MiSeq) and analyzed using Nextclade and USHER (comparison data from GISAID). Using sequencing data, the evolution, transmission, and emergence of SARS-CoV-2 variants were monitored over time in the campus community. Using sequencing data from NY, PA, and NJ in combination with University data, we performed an IQ-TREE based phylogenetic analysis.

Results
Analyzing sequencing data of 1,011 University positive samples we demonstrate that SARS-CoV-2 variants Delta (B.1.617.2) and Omicron (BA.1 and BA.2) were first to emerge following widespread vaccination and, quickly, became predominant. These trends witnessed on campus preceded those same variants emerging in New Jersey, providing evidence of local campus spread distinct from the state-wide pandemic. The analysis of 2,359 total sequences from NY, PA, and NJ in combination with University data, provided evidence of the SARS-CoV-2 transmission chain on campus evolving from out-of-state (January 2021) to local (January 2022) spread over one year of the virus circulating within the community at large. Upon performing a Ct value analysis of 2,822 Princeton University sequences, no significant differences were discovered between N gene Ct values when grouped by age or vaccination status.

Conclusions
Sequencing of positive SARS-CoV-2 samples from population screening of a highly vaccinated University campus community allowed the detection of emergence of new variants that became predominant on campus irrespective of the circulation of variants in the surrounding area.

Introduction

Setting
Princeton University's campus contains multiple dormitory style living spaces and dining halls in 6 college 'houses' in which nearly 100% of the 5,240 **undergraduate** students live and eat.

In addition, nearly 70% of the 3,157 **graduate** students live in dormitory style housing on and near campus.

Students interact daily with **faculty** (1,285) and **staff** (7,400). With over 200 **international internship** opportunities, most students travel internationally on an annual basis.

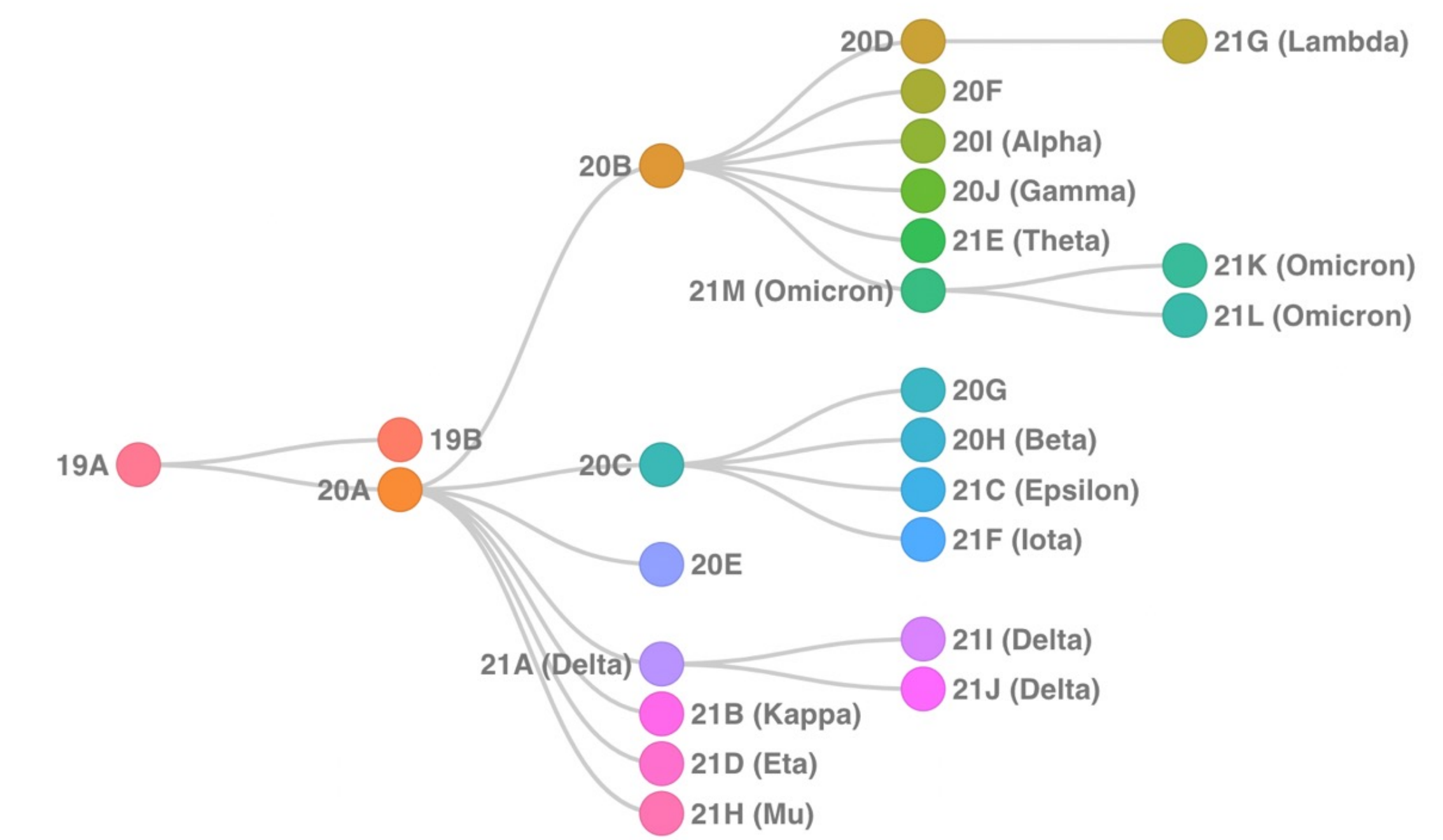
The introduction of a communicable disease on campus could result in an outbreak of disease affecting a large majority of the nearly 17,000 persons who live, work, and eat daily on campus.



SARS-CoV-2 Mutation and Evolution
SARS-CoV-2, a novel and highly pathogenic coronavirus, has caused unprecedented global disruption following its introduction into the human population. Since its emergence in December 2019, SARS-CoV-2 has diversified into different co-circulating variants grouped into clades defined by combinations of signature mutations.

Viral evolution occurs rapidly, and the presence of mutations that alter either the transmissibility or virulence of a virus lead to the potential for variants of significant public health concern to emerge.

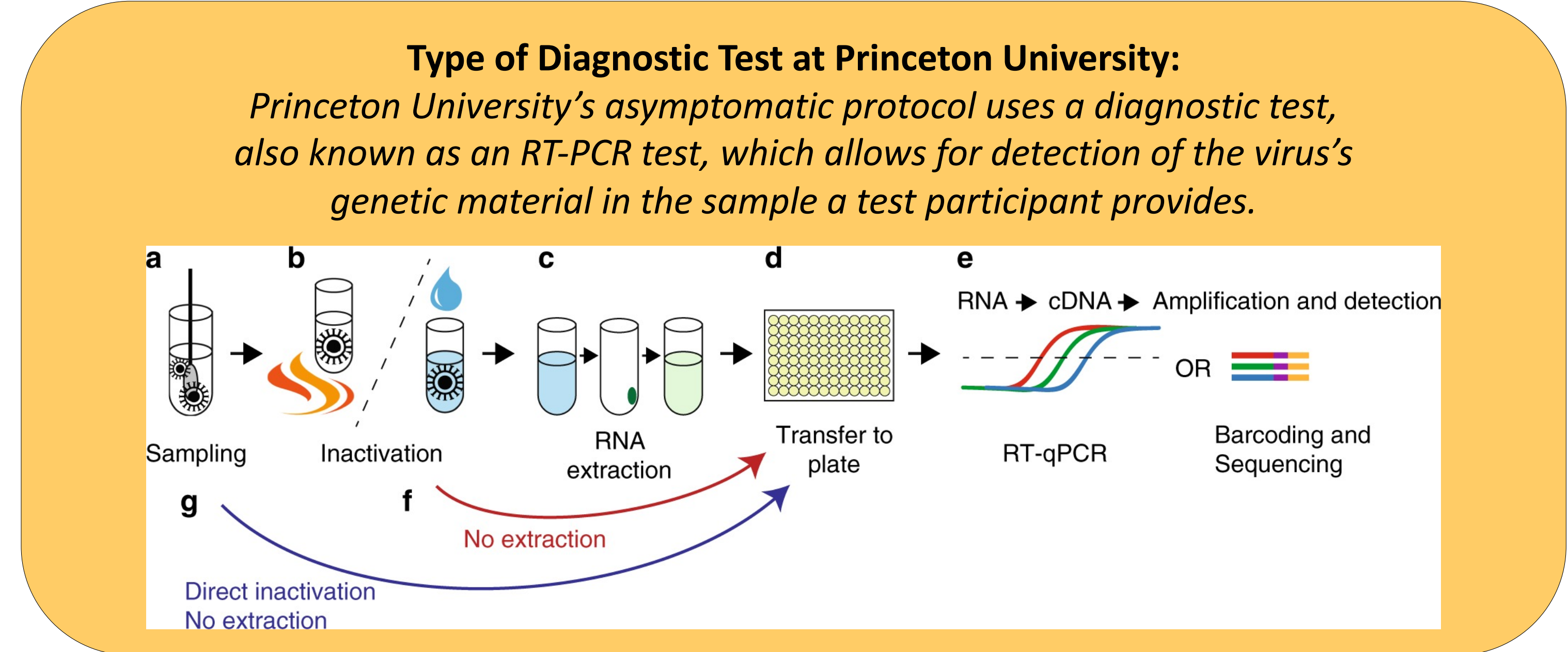
The accumulation of S protein mutations is associated with viral escape from vaccine-induced immunity and prior-infection-induced immunity, resulting in the potential for both breakthrough cases and SARS-CoV-2 re-infections.



Nextstrain Clades — Evolution of SARS-CoV-2 Variants. Illustration of the phylogenetic relationships between SARS-CoV-2 clades, as defined by Nextstrain. As the distance from 19A (original strain identified in Wuhan, China in December 2019) increases, the total number of mutations present in the SARS-CoV-2 genome increases. Adapted from Nextstrain.

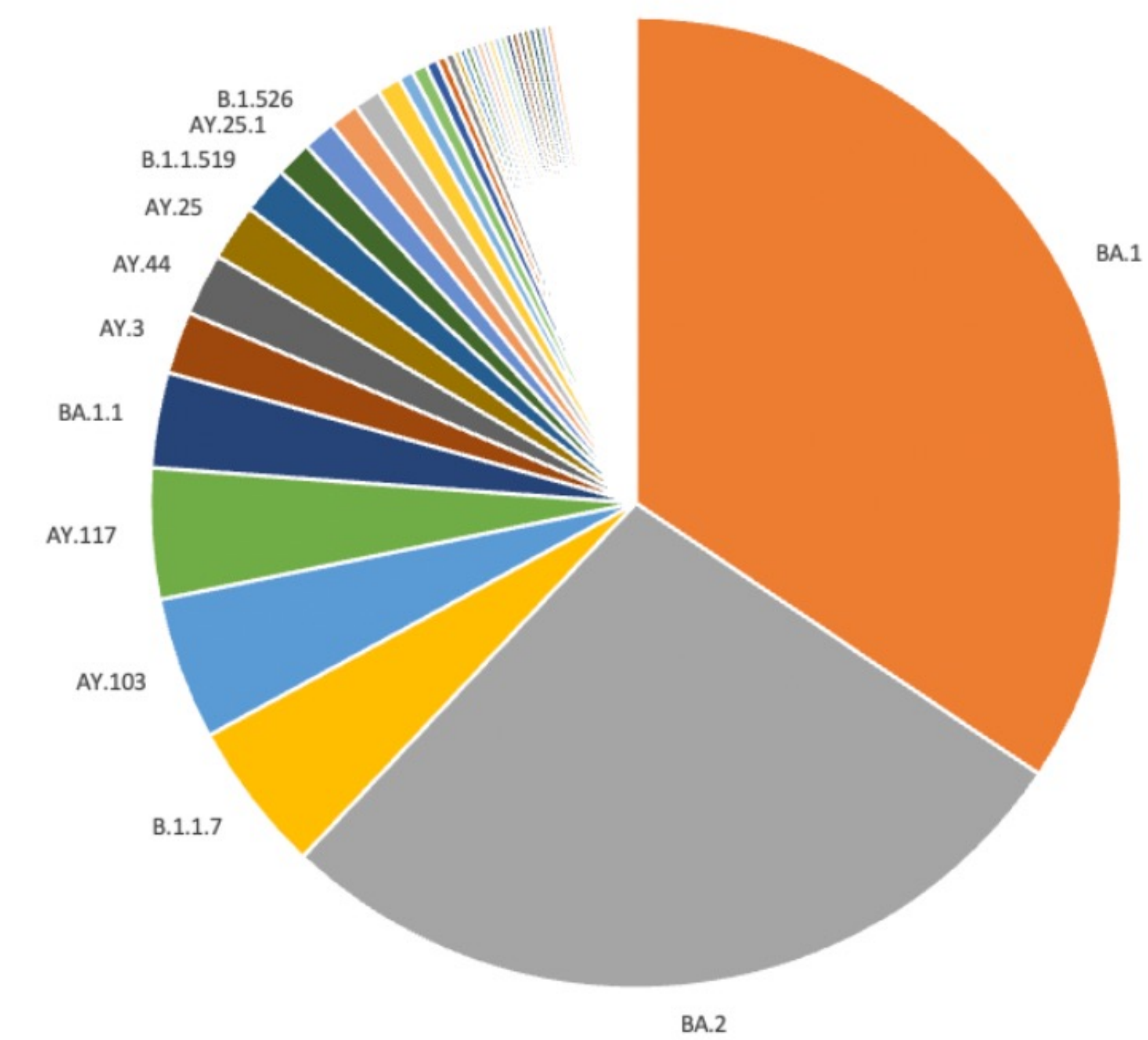
Methods

- Evolutionary Analysis Methodology**
 - Beginning in the 2021 spring semester, Princeton University invited all students to campus and initiated an asymptomatic testing protocol using PCR-based detection of human saliva samples. RNA extracted from PCR-positive human saliva samples was sequenced for surveillance purposes.
 - Positive samples were submitted for RNA-Seq analysis (ARTIC amplicon sequencing protocol, illumina MiSeq) and analyzed using Nextclade and USHER (comparison data from GISAID).
 - Using sequence data, the evolution, transmission, and emergence of SARS-CoV-2 variants were monitored over time in the campus community. Using sequencing data from NY, PA, and NJ in combination with University data, an IQ-TREE based phylogenetic analysis was performed.
- CT Value Analysis Methodology**
 - Cycle threshold (Ct) Values are obtained from reverse-transcriptase quantitative polymerase chain reactions (RT-qPCR). RNA extracted from PCR-positive human saliva samples obtained through Princeton University's asymptomatic testing protocol was sequenced, and Ct value data was obtained for the N, S, MS2, and ORF1ab genes of 2,822 SARS-CoV-2 samples. Box plots of N Gene Ct value grouped by either patient age or vaccination status were generated.

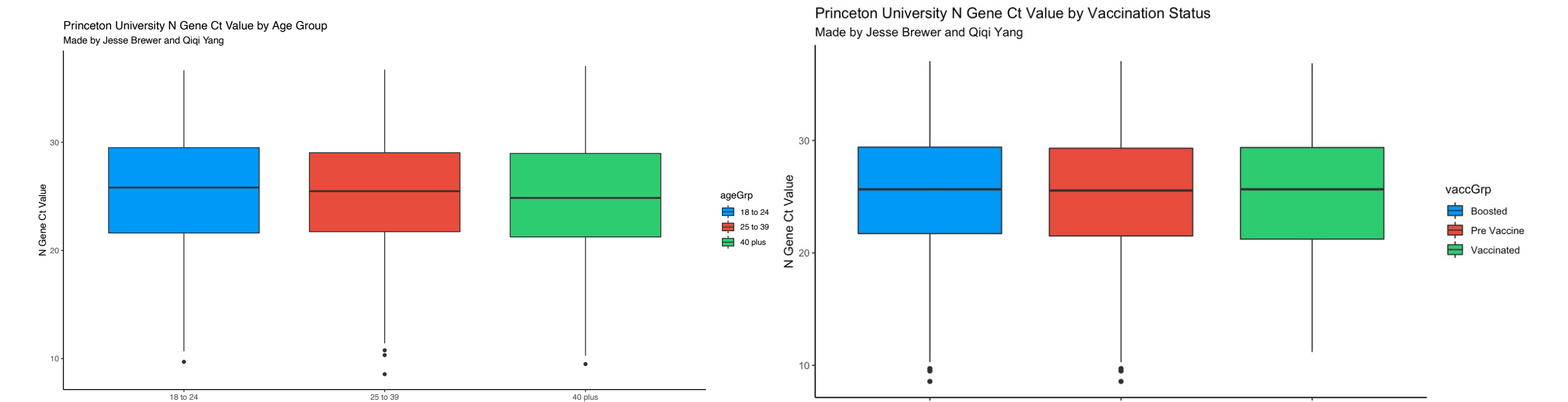


Results

Princeton University Samples — Breakdown by Pango Lineage



Analyzing sequencing data of 1,011 University positive samples we demonstrate that SARS-CoV-2 variants Delta (B.1.617.2) and Omicron (BA.1 and BA.2) were first to emerge following widespread vaccination and, quickly, became predominant. These trends witnessed on campus preceded those same variants emerging in New Jersey, providing evidence of local campus spread distinct from the state-wide pandemic.



Upon performing a Ct value analysis of 2,822 Princeton University sequences, no significant differences were discovered between N gene Ct values when grouped by age or vaccination status. However, there were significant differences in Ct values between strains.

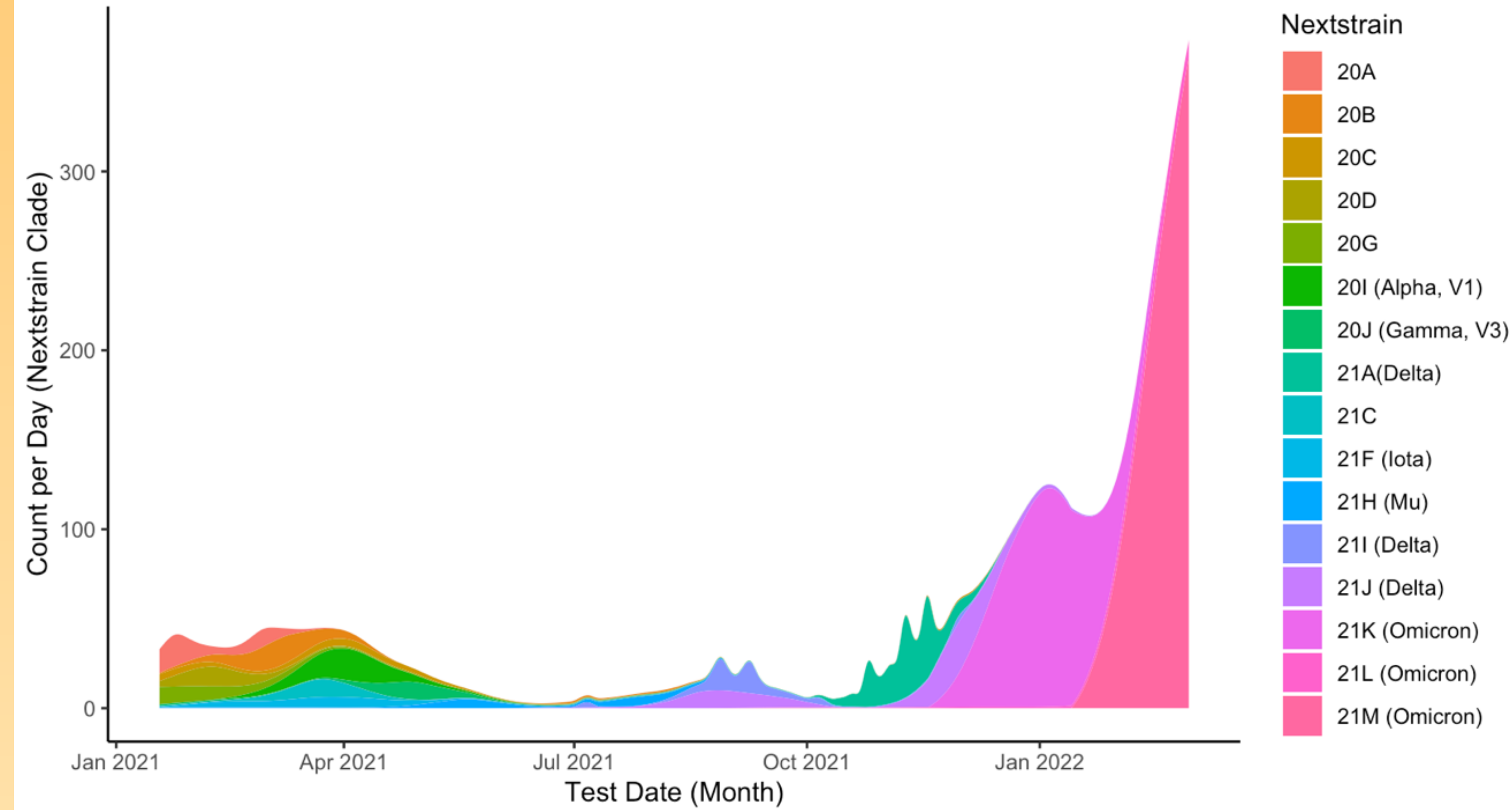
Conclusions

Sequencing of positive SARS-CoV-2 samples from population screening of a highly vaccinated University campus community allowed the detection of emergence of new variants that became predominant on campus irrespective of the circulation of variants in the surrounding area.

The highest frequency of SARS-CoV-2 cases detected at Princeton University included BA.1, followed by BA.2, supporting the literature that describes the high transmissibility of such variants.

Similar screening and sequencing protocols may be employed in the future throughout other College Communities in order to monitor the course of infectious disease outbreaks.

Princeton University Emerging SARS-CoV-2 Variants over Time



The Emergence of SARS-CoV-2 Variants at Princeton University. This plot represents the emergence of SARS-CoV-2 clades on campus, organized by Nextstrain clade and displayed as a proportion out of one. Dates of sample collection range from January 25, 2021, to March 1, 2022. X-axis represents SARS-CoV-2 sample test date (grouped by month); Y-axis represents the count per day organized by Nextstrain clade.