

Genomic surveillance of SARS-CoV-2 in wastewater reveals a highly divergent cryptic lineage



Alexis Russell¹, Ian Vasconcellos Caldas², Victoria Schnurr¹, Sara B. Griesemer¹, Aishwarya Jadhav¹, Lindsey Rickerman¹, Yasir H. Ahmed-Braimah², Marc C. Johnson³, Daryl M. Lamson¹, Kirsten St. George^{1,4}

¹ Wadsworth Center, New York State Department of Health, Albany NY; ² Department of Biology, Syracuse University, Syracuse NY; ³ Department of Molecular Microbiology & Immunology, University of Missouri-School of Medicine, Columbia MO; ⁴ Department of Biomedical Science, University at Albany, SUNY, Albany NY



Department of Health

Wadsworth Center

Background

- Cryptic lineages are unique, highly-divergent SARS-CoV-2 lineages found in wastewater but not closely related to any clinical patient sequences.¹ Current evidence suggests that cryptic lineages arise from single human hosts with chronic SARS-CoV-2 infection leading to prolonged gastrointestinal viral shedding.
- Unrelated cryptic lineages tend to convergently evolve several of the same, otherwise extremely rare mutations, which may be adaptations to prolonged infection in the gastrointestinal tract.
- The New York State Department of Health conducts routine genomic surveillance of SARS-CoV-2 in wastewater samples collected statewide, allowing for the tracking of SARS-CoV-2 variants across time and geography at a population level.
- Results of whole genome sequencing from multiple wastewater samples collected in Bergen Point, Suffolk County, Long Island in April 2023 contained a large proportion of reads indicating an unusual, highly-mutated SARS-CoV-2 lineage, distantly related to B.1.1.7 (Alpha), a variant which last circulated two years prior in 2021.
- An investigation was initiated to determine if this highly-mutated sequence could be from a cryptic lineage.

Methods

- Wastewater samples were extracted and quantified at Stony Brook University and sent to New York State Department of Health for whole genome sequencing with the Ion AmpliSeq SARS-CoV-2 Insight Research Assay for the Genexus Integrated Sequencer.
- Relative proportions of SARS-CoV-2 lineages within each sample were determined using Freyja.²
- Consensus sequences at 20x coverage were generated from samples determined to be >70% proportion B.1.1.7-like lineage and were placed in the SARS-CoV-2 phylogenetic tree using Nextclade.³ The unique mutation profile of the lineage was compared to concurrently circulating lineages, all publicly-available clinical sequences, and other known cryptic lineages.
- The relative proportions of four unique single nucleotide polymorphism (SNP) mutations characteristic of the lineage were tracked in wastewater collections from the Bergen Point sewershed from January through October 2023.
- A tiled mutation analysis method was developed to track the presence of five SNP mutations characteristic of the lineage in all wastewater collections statewide January through October 2023. A logsum value was used to weigh the number of reads containing the mutations (sensitive to viral concentration and sequencing quality) and the co-occurrence of nearby mutations on the same read (tile) to assess the likelihood that the signal was generated by a single lineage within the mixture.

Results

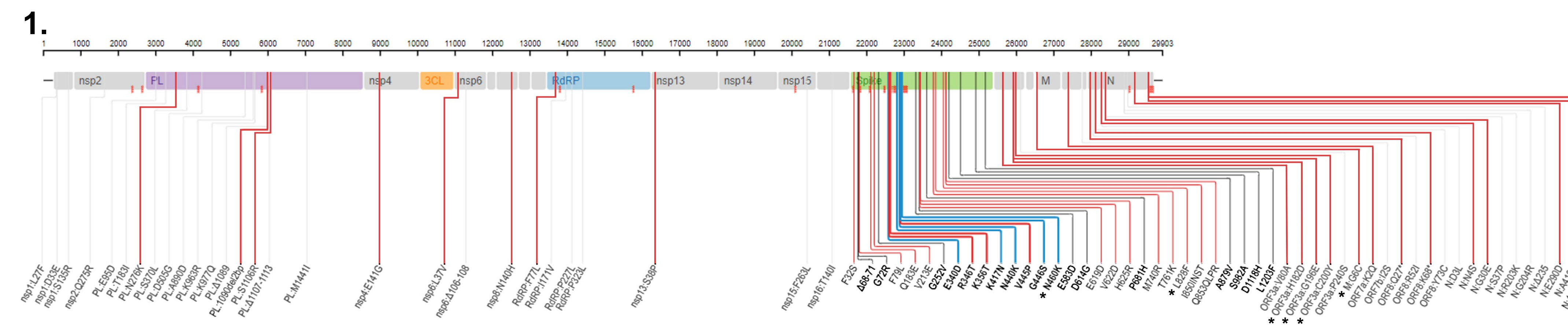


Figure 1: Mutation profile of consensus sequence from Bergen Point wastewater sample collected 4/24/23. Mutations in reference to Wuhan-Hu-1 are listed. Red denotes a rare mutation with a global prevalence below 0.01%. Blue denotes a known monoclonal antibody resistance mutation. * denotes a mutation that has been identified in other cryptic lineages. Created using the Coronavirus Resistance Database.⁴

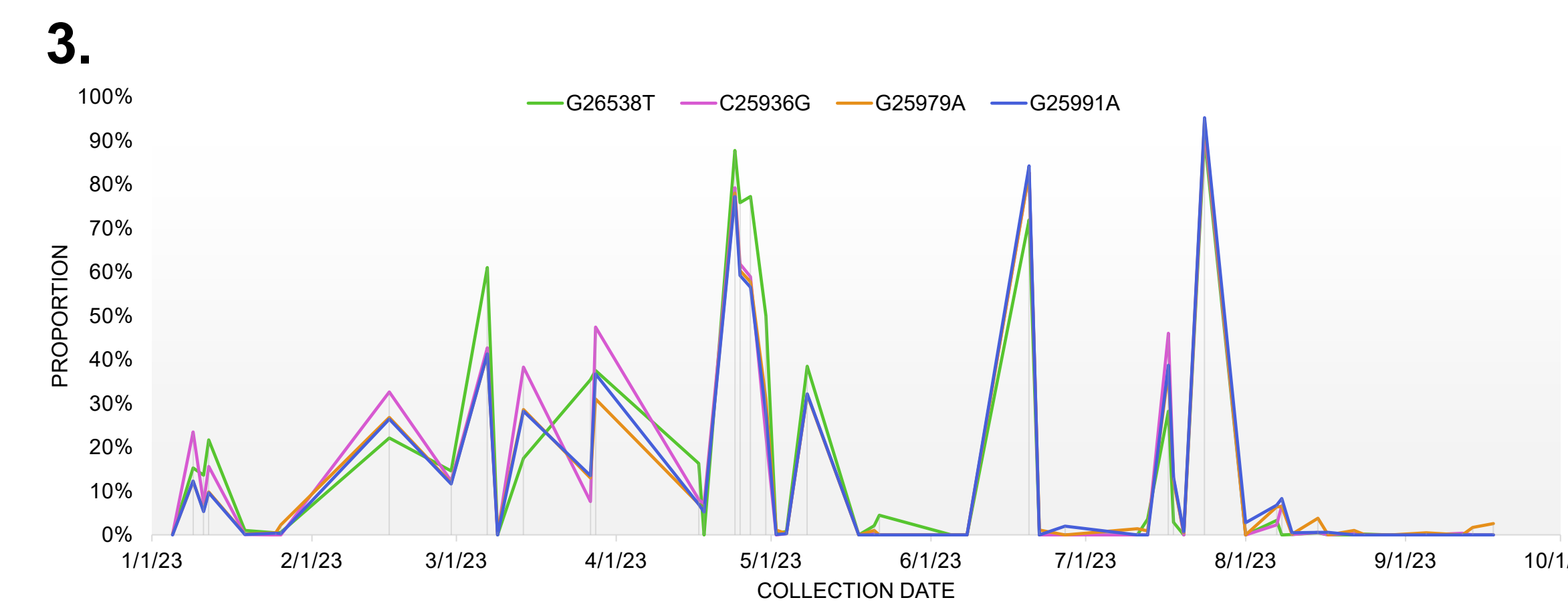
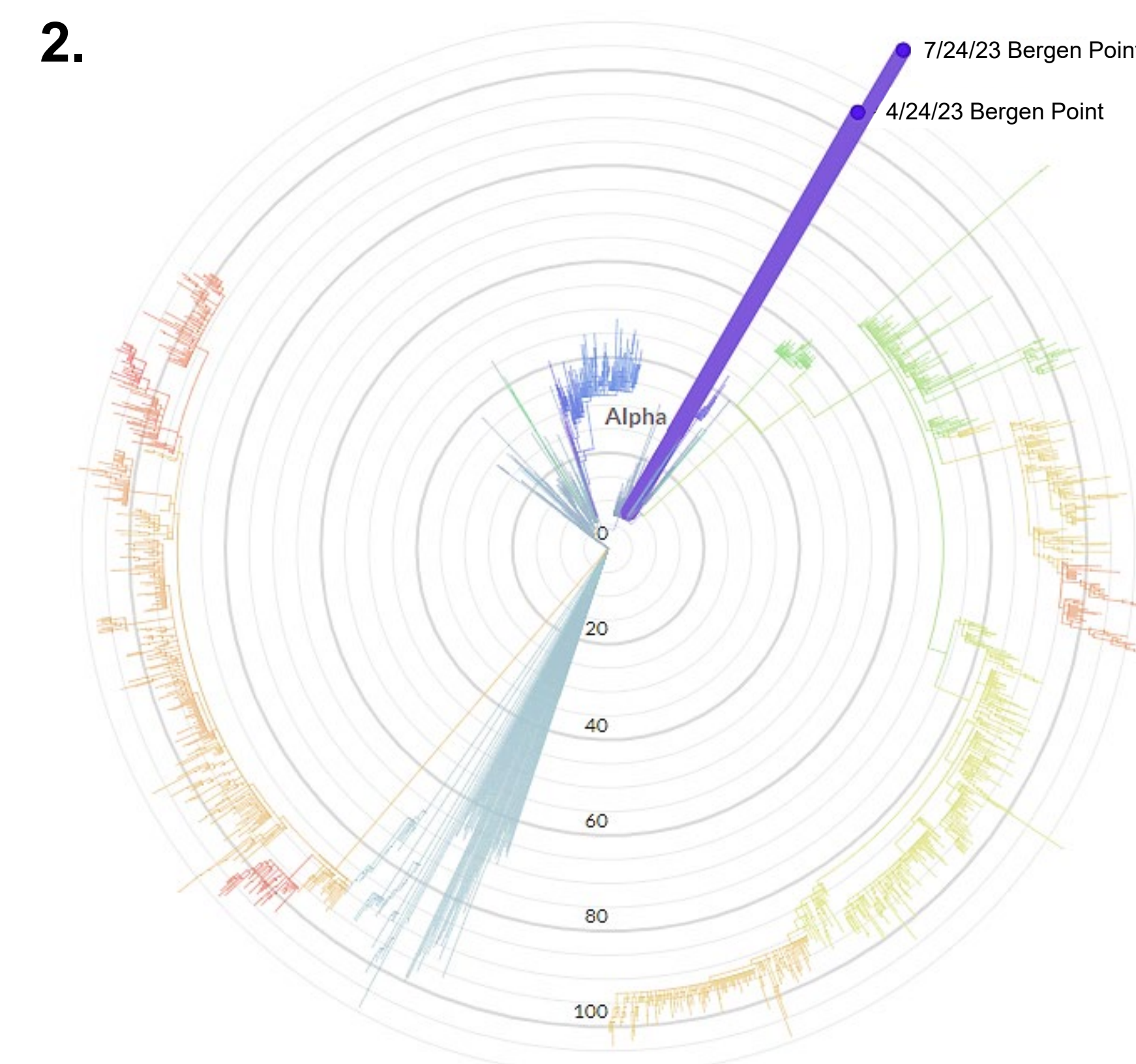


Figure 2: Global SARS-CoV-2 phylogenetic tree placement of consensus sequences from Bergen Point wastewater samples collected 4/24/23 and 7/24/23. Numbered concentric circles denote mutational distance from Wuhan-Hu-1 reference. Created using Nextclade.³

Figure 3: Relative proportions of four unique single nucleotide polymorphism (SNP) mutations characteristic of the B.1.1.7-like cryptic lineage in raw sequencing data from Bergen Point wastewater samples by collection date.

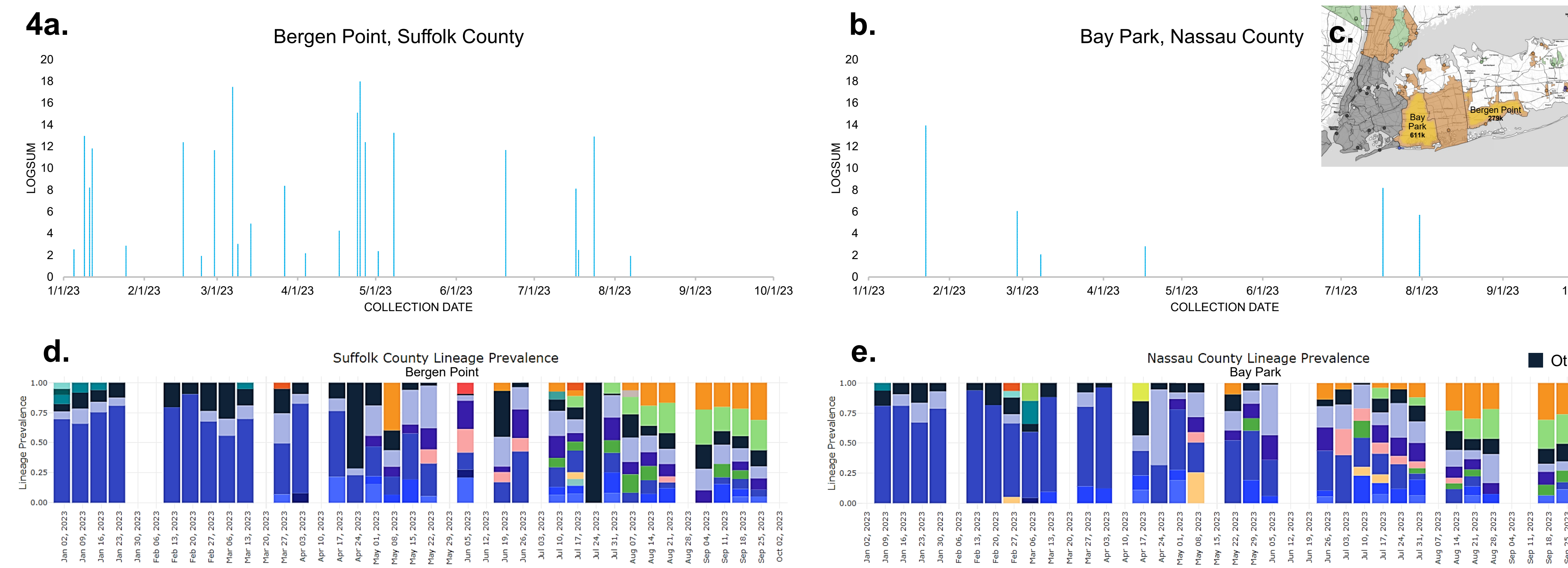


Figure 4: Tiled mutation analysis results from **a.** Bergen Point and **b.** Bay Park wastewater samples by collection date, with higher logsum value indicating higher likelihood of presence of the B.1.1.7-like cryptic lineage. **c.** Map of Bay Park and Bergen Point sewersheds on Long Island, New York with total population served by each. Freyja lineage deconvolution results for **d.** Bergen Point and **e.** Bay Park wastewater samples by collection date, with the B.1.1.7-like cryptic lineage falling within the “other” lineage category (dark navy).

Results

- The wastewater lineage is a highly-divergent B.1.1.7 (Alpha) descendent containing extensive mutations across the genome, several of which are extremely rare in clinical samples but are characteristic of cryptic lineages.
- SNP mutation analysis of Bergen Point wastewater samples indicates that the lineage was present in January 2023, when wastewater collections began, and varied in relative proportion within samples from that sewershed until August 2023. The lineage constituted >50% proportion of SARS-CoV-2 reads within Bergen Point wastewater multiple times, including in April, June, and July 2023.
- Tiled mutation analysis results were concordant with SNP proportion results in Bergen Point, and identified an additional sewershed in Bay Park, Nassau County, Long Island where the lineage was also present several times during the same period.

Conclusions

- The wastewater lineage likely represents a cryptic lineage, shed by an individual originally infected in mid-2021 and periodically shedding virus for over two years.
- Very high concentration viral shedding is likely, given the periodic predominance of the cryptic lineage in wastewater from sewersheds serving large populations.
- Tiled mutation analysis is an effective tool and will continue to be used to monitor for cryptic lineages and emerging variants in wastewater.

References

- Gregory DA, Trujillo M, Rushford C, Flury A, Kannoly S, et al. Genetic diversity and evolutionary convergence of cryptic SARS-CoV-2 lineages detected via wastewater sequencing. *PLoS Pathogens*. 2022 Oct 14;18(10):e1010636. doi: 10.1371/journal.ppat.1010636.
- Karthikeyan S, Levy JI, De Hoff P, Humphrey G, Birmingham A, et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature*. 2022 Jul 7;609(7925):101-108. doi: 10.1038/s41586-022-05049-6.
- Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software*. 2021 Nov 30, 6(67), 3773. doi: 10.21105/joss.03773.
- Tzou PL, Tao K, Pond SLK, Shafer RW. Coronavirus Resistance Database (CoV-RDB): SARS-CoV-2 susceptibility to monoclonal antibodies, convalescent plasma, and plasma from vaccinated persons. *PLoS One*. 2022 Mar 9;17(3):e0261045. doi: 10.1371/journal.pone.0261045.

Acknowledgements

The authors thank New York State Department of Health Wadsworth Center Laboratory of Viral Diseases, New York State Department of Health Center for Environmental Health, New York State Wastewater Surveillance Consortium Laboratories, Syracuse University, Stony Brook University, and Thermo Fisher Scientific. This work was funded by Epidemiology and Laboratory Capacity Grant 15-0775-10.