medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Tracing the origin of SARS-CoV-2 Omicron-like Spike sequences detected in wastewater

Martin M. Shafer, Devon Gregory, Max J. Bobholz, Adelaide Roguet, Luis A. Haddock Soto, Clayton Rushford, William C. Vuyk, Kayley H. Janssen, Hunter J. Ries, Hannah E. Pilch, Paige A. Mullen, Rebecca B. Fahney, Wanting Wei, Matthew Lambert, Jeff Wenzel, Peter Halfmann, Yoshihiro Kawaoka, Nancy A. Wilson, Thomas C. Friedrich, Ian W. Pray, Ryan Westergaard, David H. O'Connor*, Marc C. Johnson*

Author	Affiliation(s)	Email	ORCID	Highest degrees
Martin M. Shafer	University of Wisconsin-Madison, Wisconsin State Laboratory of Hygiene, School of Medicine & Public Health	mmshafer@wisc.edu	0000-0002-0810-298X	Ph.D.
Devon Gregory	University of Missouri-School of Medicine	gregoryde@missouri.edu	0000-0002-7865-3891	Ph.D.
Max J. Bobholz	University of Wisconsin-Madison Department of Pathology and Laboratory Medicine	mbobholz@wisc.edu	0000-0002-6455-3902	B.S.
Adelaide Roguet	University of Wisconsin-Madison, Wisconsin State Laboratory of Hygiene, School of Medicine & Public Health	Adelaide.Roguet@slh.wisc.edu	0000-0003-3688-4525	Ph.D.
Luis A. Haddock Soto	University of Wisconsin–Madison Department of Pathobiological Sciences	haddocksoto@wisc.edu	0000-0001-8916-5693	M.S.
Clayton Rushford	University of Missouri-School of Medicine	car539@mail.missouri.edu		M.S.
William C. Vuyk	University of Wisconsin-Madison Department of Pathology and Laboratory Medicine	wvuyk@wisc.edu	0000-0002-9461-4788	B.S.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Kayley H. Janssen	University of Wisconsin-Madison, Wisconsin State Laboratory of Hygiene, School of Medicine & Public Health	Kayley.Janssen@slh.wisc.edu	0000-0001-9546-5799	Ph.D.
Hunter J. Ries	University of Wisconsin–Madison Department of Pathobiological Sciences	hries@wisc.edu	0000-0001-7280-4399	B.S.
Hannah E. Pilch	University of Wisconsin-Madison, Wisconsin State Laboratory of Hygiene, School of Medicine & Public Health	Hannah.Pilch@slh.wisc.edu	0000-0001-5656-329X	M.S.
Paige A. Mullen	University of Wisconsin-Madison, Wisconsin State Laboratory of Hygiene, School of Medicine & Public Health	Paige.Mullen@slh.wisc.edu	0000-0003-4690-4374	B.S.
Rebecca B. Fahney	University of Wisconsin-Madison, Wisconsin State Laboratory of Hygiene, School of Medicine & Public Health	Rebecca.Fahney@slh.wisc.edu	0000-0003-1629-4392	B.S.
Wanting Wei	University of Wisconsin–Madison Department of Pathobiological Sciences	wwei54@wisc.edu	0000-0001-5978-5883	B.S.
Matthew Lambert	University of Wisconsin-Madison Department of Medicine; Wisconsin Department of Health Services	matthew.lambert@dhs.wisconsin.gov	0000-0002-5302-4650	MD
Jeff Wenzel	Missouri Department of Health and Senior Services	Jeff.Wenzel@health.mo.gov		B.S.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Peter Halfmann	University of Wisconsin–Madison Department of Pathobiological Sciences	peter.halfmann@wisc.edu	0000-0002-1648-1625	Ph.D.
Yoshihiro Kawaoka	University of Wisconsin–Madison Department of Pathobiological Sciences	yoshihiro.kawaoka@wisc.edu	0000-0001-5061-8296	DVM, Ph.D.
Nancy A. Wilson	University of Wisconsin-Madison Department of Pathology and Laboratory Medicine	nwilson@wisc.edu	0000-0002-9472-6026	Ph.D.
Thomas C. Friedrich	University of Wisconsin–Madison Department of Pathobiological Sciences	tfriedri@wisc.edu	0000-0001-9831-6895	Ph.D.
Ian W. Pray	Career Epidemiology Field Officer, CDC Wisconsin Department of Health Services, Madison, WI	ian.pray@dhs.wisconsin.gov	0000-0001-6935-5123	Ph.D.
Ryan Westergaard	University of Wisconsin-Madison Department of Medicine; Wisconsin Department of Health Services	ryan.westergaard@dhs.wisconsin.gov	0000-0001-5701-4516	MD, PhD
David H. O'Connor	University of Wisconsin-Madison Department of Pathology and Laboratory Medicine	dhoconno@wisc.edu	0000-0003-2139-470X	Ph.D.
Marc C. Johnson	University of Missouri-School of Medicine	marcjohnson@missouri.edu		Ph.D.

* Contributed equally

Corresponding author

Marc C. Johnson
Phone: 573-882-1519
marcjohnson@missouri.edu

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

471c Bond Life Sciences Center
1201 Rollins St
Columbia, MO USA 65211

Word Count

1250

Disclaimer

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.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Key points

Question: What is the source of unusual SARS-CoV-2 Omicron-like Spike variants detected in wastewater but not in clinical samples?

Findings: We identified a cryptic SARS-CoV-2 lineage in wastewater collected at a central wastewater treatment facility and traced its source to a single wastewater outlet serving six restrooms. The virus in this sample resembled a 2020-2021 lineage except for the Spike protein, in which Omicron-like variants were observed.

Meaning: Prolonged shedding from the human gastrointestinal tract is the most likely source for evolutionarily advanced SARS-CoV-2 variant sequences found in wastewater.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Abstract

Importance: The origin of highly divergent "cryptic" SARS-CoV-2 Spike sequences, which appear in wastewater but not clinical samples, is unknown. These wastewater sequences have harbored many of the same mutations that later emerged in Omicron variants. If these enigmatic sequences are human-derived and transmissible, they could both be a source of future variants and a valuable tool for forecasting sequences that should be incorporated into vaccines and therapeutics.

Objective: To determine whether enigmatic SARS-CoV-2 lineages detected in wastewater have a human or non-human (i.e., animal) source.

Design: On January 11, 2022, an unusual Spike sequence was detected in municipal wastewater from a metropolitan area. Over the next four months, more focused wastewater sampling resolved the source of this variant.

Setting: This study was performed in Wisconsin, United States, which has a comprehensive program for detecting SARS-CoV-2 in wastewater.

Participants: Composite wastewater samples were used for this study; therefore, no individuals participated.

Main Outcome(s) and Measure(s): The primary outcome was to determine the host(s) responsible for shedding this variant in wastewater. Both human and non-human hosts were plausible candidates at the study's outset.

Results: The presence of the cryptic virus was narrowed from a municipal wastewater sample (catchment area >100,000 people) to an indoor wastewater sample from a single facility (catchment area ~30 people), indicating the human origin of this virus. Extraordinarily high concentrations of viral RNA (~520,000,000 genome copies / L and ~1,600,000,000 genome copies / L in June and August 2022, respectively) were detected in the indoor wastewater sample. The virus sequence harbored a combination of fixed nucleotide substitutions previously observed only in Pango lineage B.1.234, a variant that circulated at low levels in Wisconsin from October 2020 to February 2021.

Conclusions and Relevance: High levels of persistent SARS-CoV-2 shedding from the gastrointestinal tract of an infected individual likely explain the presence of evolutionarily advanced "cryptic variants" observed in some wastewater samples.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Compared to SARS-CoV-2 viruses that circulated in 2020, the Omicron lineage that first emerged in South Africa in November 2021 had a highly divergent Spike receptor binding domain (RBD), with 10 lineage-defining amino acid substitutions between residues 412 and 579 (K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, and N501Y) [1–3]. Though Omicron did not begin its spread until late 2021, variants at 8 of these residues were observed in New York City wastewater samples collected in the first half of 2021. One wastewater sample, collected in May 2021, had more extensive mutation than current Omicron lineages, with one haplotype containing sixteen amino acid substitutions between Spike residues 412 and 579 [4]. Subsequent testing of wastewater samples from Missouri and California resulted in sporadic detection of additional "cryptic" lineages that are not commonly detected in clinical samples [5]. This raises the question of whether divergent "cryptic" sequences in wastewater will continue to forecast viral variants that will later emerge and spread in people. If so, these sequences could help inform medical countermeasure approaches to control spread and transmission and assess how cross-protective existing vaccines and therapeutics are against divergent, potentially "evolutionarily advanced" viruses.

There are two leading, non-mutually-exclusive hypotheses for the source of these "cryptic" sequences. The first is that unsampled infections in people may have harbored viruses with these sequences, as has been shown for less divergent wastewater sequences [6]. Individuals with prolonged SARS-CoV-2 infections in the gastrointestinal tract could be a significant subgroup of unsampled infections. Individuals with immunocompromising conditions are at high risk for such prolonged infections, and suboptimal immune responses in the gastrointestinal tract of such individuals could select for antigenic variation over the course of infection [7,8]. Human waste is the predominant source of genetic material in wastewater, making it the most parsimonious explanation for the origin of SARS-CoV-2 sequences in this material. An alternate hypothesis is that an un- or under-sampled animal reservoir is shedding these viruses into wastewater, e.g., through defecation in combined sewer systems, via inflow/infiltration, or livestock processing. SARS-CoV-2 exhibits a broad host range, including infection of household pets [9]. One study detected SARS-CoV-2 antibodies in 40% of wild white-tailed deer sampled in four U.S. states [10], suggesting wide natural circulation. SARS-CoV-2 has caused large-scale outbreaks in farmed mink, some of which have included transmission of novel viral variants back to humans [11–13]. Thus, it is plausible that wild animal reservoirs already exist undetected; it is also plausible that prolonged infections occur in a subset of animal infections.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

On January 11, 2022, a cryptic lineage containing at least six unusual Spike RBD variants was first detected in a composite wastewater sample from a metropolitan area in Wisconsin (Figure 1). The initial sample contained raw influent from five interceptor districts in the metropolitan area sewershed, effectively sampling a population of more than 100,000 people. The source of the variant was narrowed by testing each of the five interceptors; the cryptic lineage was only detected in one. Additional testing of manholes and substations within the interceptor district in March 2022 revealed the ongoing persistence of the cryptic lineage and further refined the source of the variant. By May 2022, this investigation traced the source of the lineage to a single manhole accessing a lateral collecting wastewater from a single building. Subsequent testing of wastewater from service lines within the building in June 2022 further narrowed the source to one side of the building serving six restrooms, supporting a human source of this cryptic lineage (Figure 2). The positive service line was retested in August 2022 and remained positive for the cryptic lineage.

As quantified by digital PCR, unprecedented wastewater RNA viral loads were observed in samples collected in June (~520,000,000 genome copies per liter undiluted wastewater) and August (~1,600,000,000 copies per liter), though viable virus could not be cultured. This resolved a paradox from earlier cryptic lineage studies: if cryptic lineages comes from only a single source, how could they be detected in a dilute municipal wastewater sample? The amount of viral RNA present in this single source would be sufficiently detected even after extensive dilution in wastewater from other sources.

Sequencing the genome of this unusual SARS-CoV-2 using overlapping PCR amplicons hints at its origins and evolution. Both the June and August timepoints had identical, nearly fixed (>95%) variants at 45 different sites throughout the genome (Supplemental Figure 1). The consensus sequences for both genomes were classified as lineage B.1.234 by Pangolin. B.1.234 viruses were first detected in Wisconsin on September 2, 2020, and were last detected on March 30, 2021 [14]. The simplest explanation of this data is that a single individual, originally infected when B.1.234 was in circulation, excreted viruses with the cryptic lineage in 2022. Notably, mutations have accumulated in this lineage faster than expected based on the substitution rate that prevailed when B.1.234 viruses were circulating (Figure 3a). We also detect a substantially elevated rate of nonsynonymous substitutions relative to synonymous ones in the spike gene, but not in other viral genes. These observations suggest that Spike variation in this virus is driven by diversifying selection (Figure 3).

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Interestingly, in addition to the highly divergent Spike, there was a cluster of fixed variants in the region that encodes the ectodomain of the viral Membrane protein. This region is exposed outside of the SARS-CoV-2 virion and is a known target for binding antibodies [15,16]. In one analysis, Membrane-binding antibodies were present at a higher level than anti-Spike antibodies. The cluster includes a 15-nucleotide insertion (GCAACAACTCAGAGT, encoding the amino acids SNNSEF); interestingly, this is identical to the sequence found between positions 11,893 and 11,907 in ORF1ab. Additionally, the cryptic lineage has M:A2E, M:G6C, and M:L17V amino acid substitutions; the phenotypic impact of these substitutions, if any, is unclear.

Together, our results indicate that a SARS-CoV-2 cryptic lineage, persistently detected in Wisconsin wastewater, was most likely derived from a single individual with prolonged infection. The mutational profile of this virus appears to bear the imprints of diversifying selection on Spike, and perhaps other gene products, consistent with reports of immunocompromised individuals with prolonged SARS-CoV-2 infections [7,8]. Our findings suggest that SARS-CoV-2 cryptic lineages in wastewater may be derived largely, or even entirely, from human, and not animal, sources. This has several important implications. Such lineages likely exist wherever people are infected with SARS-CoV-2, i.e., worldwide. More frequent global wastewater surveillance of catchment areas would likely detect many more examples of such lineages. Due to the extremely large number of Omicron infections, we speculate that Omicron-derived cryptic lineages will be detectable in wastewater in the future. Given the extensive spread of Omicron, the number of prolonged infections that give rise to these cryptic lineages is also expected to increase, making the emergence and detection of cryptic lineages more common. The fact that the Wisconsin cryptic lineages appear to be derived from a prolonged infection with an ancestral B.1.234 virus also highlights the importance of prolonged infections in the emergence of highly divergent viruses and emphasizes the importance of ascertaining such infections. We note that individuals with immunocompromising conditions are at increased risk for prolonged infections but may not be the only population in which such infections occur. SARS-CoV-2 cryptic lineage sequences could aid in forecasting the future evolutionary trajectory of SARS-CoV-2 to evaluate the cross-protection of existing and future vaccines and monoclonal antibodies.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Data availability

Sequencing data is available in NCBI SRA and Genbank. Additional data is available from https://go.wisc.edu/4134pl. All sequences used for the phylogenetic inferences were obtained from GISAID and can be accessed using the identifier EPI_SET_221024fg (doi: 10.55876/gis8.221024fg).

Acknowledgments

This study was made possible by the generous support of the Rockefeller Foundation's Regional Accelerators for Genomics Surveillance (DHO/TCF), Wisconsin Department of Health Services Epidemiology and Laboratory Capacity funds (www.dhs.wisconsin.gov, 144 AAJ8216) to DHO, CDC contract 75D30121C11060 (DHO/TCF), Wisconsin Department of Health Services ELC Wastewater Surveillance funds (www.dhs.wisconsin.gov, 130:AAI8627) to the UW-Madison Wisconsin State Laboratory of Hygiene (WSLH), and NIDA contract 1U01DA053893-01 (MJ), the Center for Research on Influenza Pathogenesis and Transmission (75N93021C00014) from the National Institutes of Allergy and Infectious Diseases to YK. The authors thank Roger Wiseman, Nick Minor, David Baker, and CDC SPHERES for helpful discussions. They would also like to acknowledge and thank the wastewater engineers from the city wastewater utility for their sewershed sampling prowess. Additional thanks to Katia Koelle and Michael Martin of Emory University for helpful discussions on the quantitative analysis of viral evolution.

Competing interests

YK has received unrelated funding support from Daiichi Sankyo Pharmaceutical, Toyama Chemical, Tauns Laboratories, Shionogi, Otsuka Pharmaceutical, KM Biologics, Kyoritsu Seiyaku, Shinya Corporation and Fuji Rebio.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

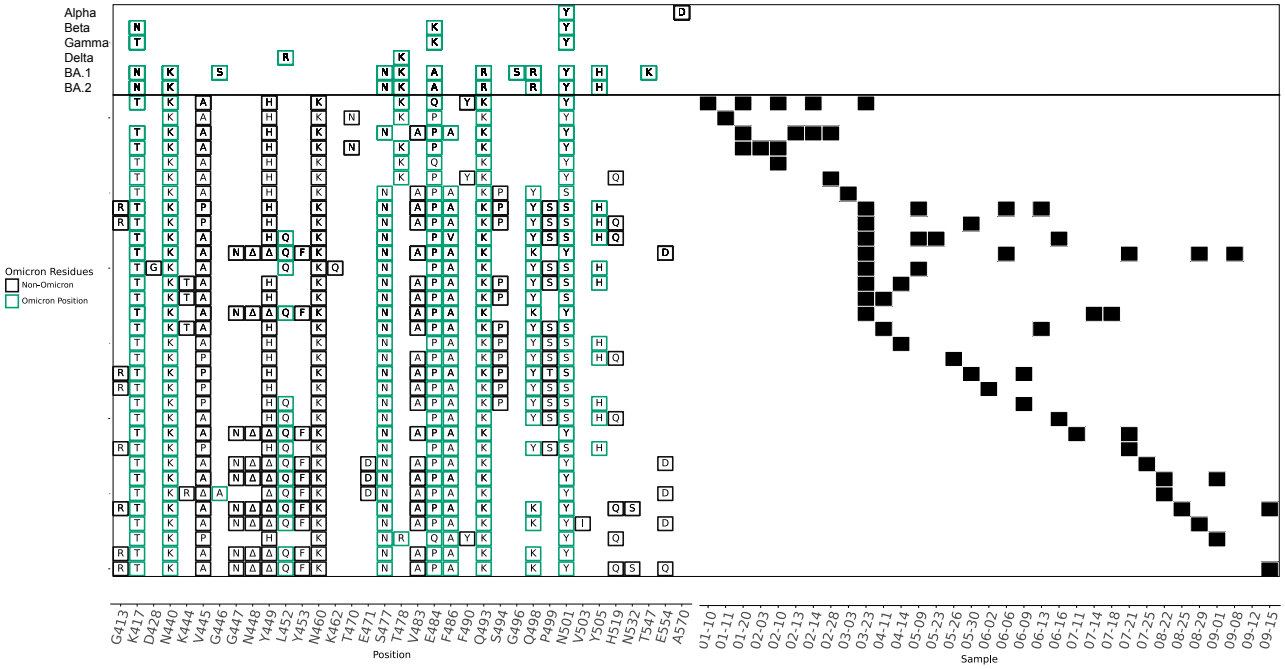

Figure 1. Representative Haplotypes of Cryptic Sequences. SARS-CoV-2 Spike RBD was amplified from wastewater samples using primers designed to exclude Omicron lineages. Haplotypes are displayed which represented at least 25% of the total sequences in at least one sample. Green boxes indicate residues that are also altered in Omicron (BA.1 or BA.2). Δ indicates an in-frame amino acid deletion.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

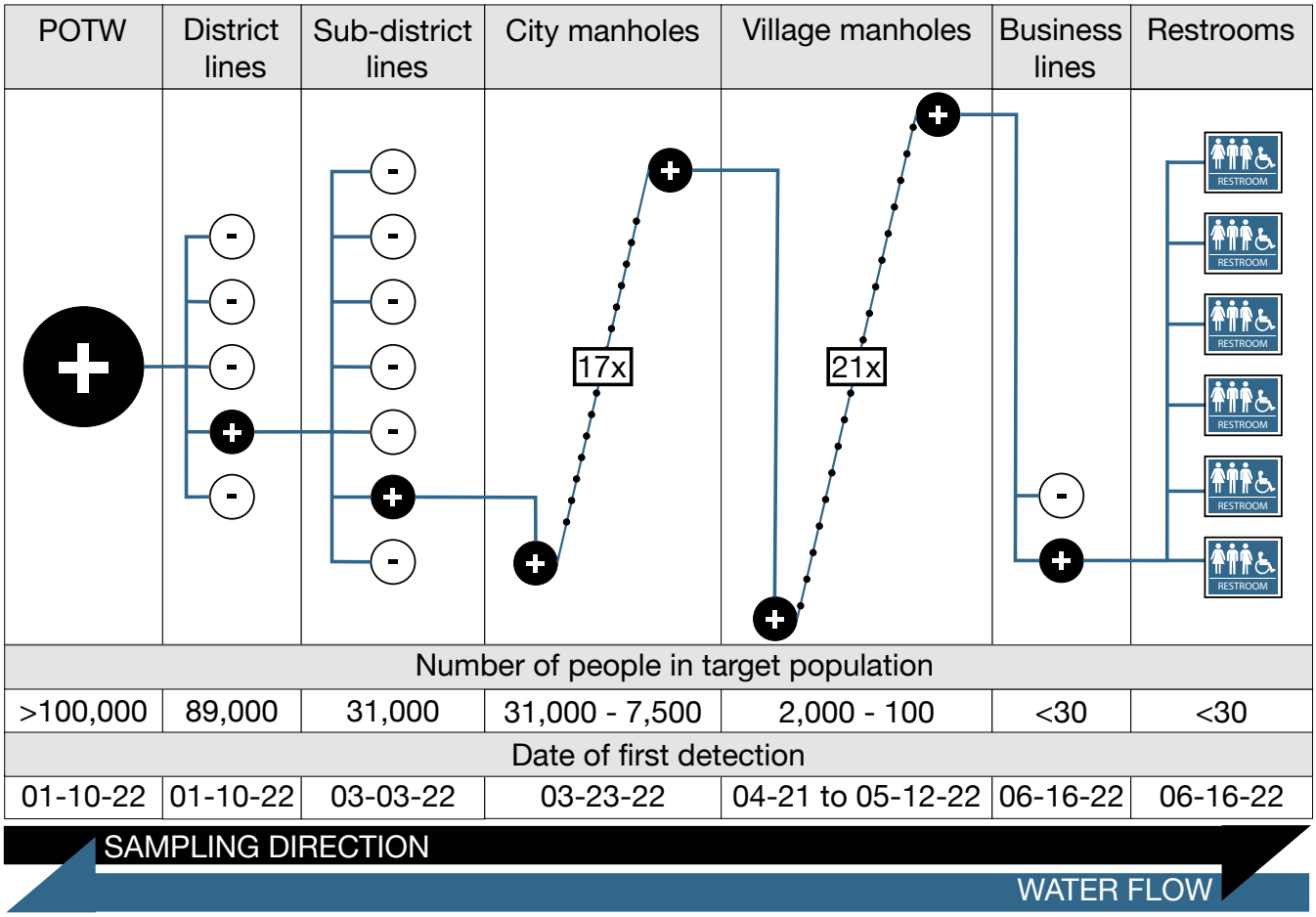

Figure 2. Wastewater sampling program used to track the source of the cryptic SARS-CoV-2 variant. The cryptic variant was first detected at the publicly owned treatment works (POTW) in wastewater from one of the five primary district lines that service the POTW sewershed. Follow-up wastewater sampling at seven pumping stations in this district isolated the variant source to one sub-district line. Further source-tracking required detailed wastewater sampling from manholes, first from 17 city-owned manholes, and then from 21 manholes within a village that sends its wastewater to the city collection system. This exhaustive manhole sampling isolated the variant source to a specific lateral line servicing a single place of business. Sampling within the place of business pinpointed the source of the variant to a wastewater line collecting sewage from 6 toilets used by employees of the facility.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

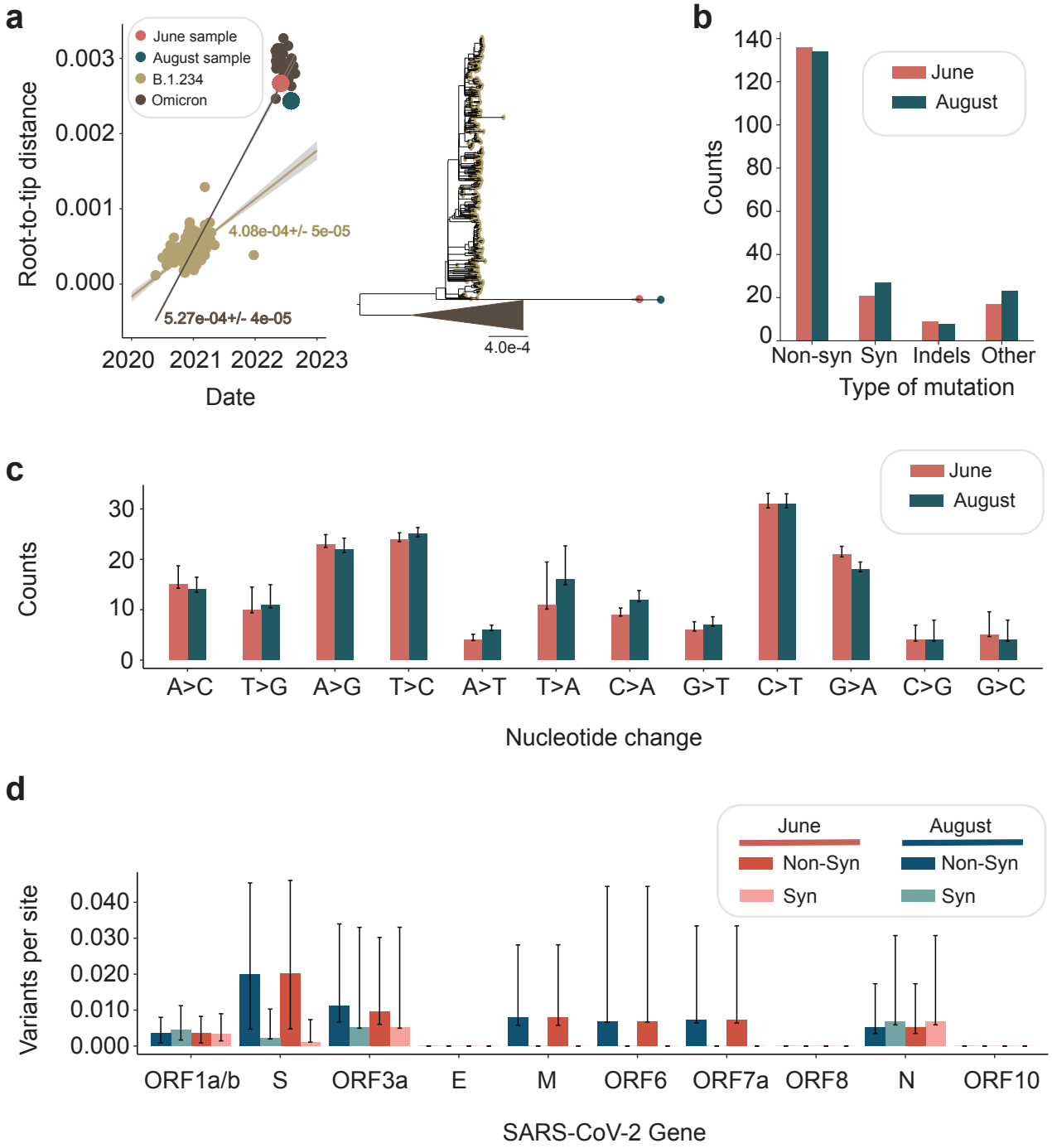

Figure 3. Diversity analysis of wastewater genomic sequences from June and August 2022. (**a**) Root-to-tip regression analysis (distance) via TreeTime based on the maximum likelihood phylogenetic tree inferred with iqtree (displayed on the right) and aligned to the most ancestral sequence. Regression is shown for B.1.234 sequences only (gold timepoints and rate) as well as both B.1.234 and Omicron samples (brown timepoints and rate). All sequences were obtained from GISAID and can be accessed using the identifier EPI_SET_221024fg (doi: 10.55876/gis8.221024fg). Estimated divergence for the June and August timepoints are shown in red and blue, respectively. (**b**) The number of

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

intra-host single nucleotide variants (iSNVs; y-axis) for the June and August timepoints for each mutation type following alignment to the reference genome MN908947.3 (colored as in panel a). Variants were classified as non-synonymous (Non-syn), synonymous (Syn), insertions-deletions (indels), or others (including nonsense and frameshift mutations). (**c**) The number of nucleotide transitions and transversions from both timepoints. The 95% confidence intervals were obtained from the relative risk (RR) of every nucleotide substitution (i.e. RR= `A>C` / `C>A`). (**d**) Shown are counts of nonsynonymous substitutions per nonsynonymous site (darker bars) and synonymous substitutions per synonymous site (lighter bars) for each SARS-CoV-2 gene. The 95% confidence intervals were obtained using a binomial probability distribution.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Position	Jun-22	Aug-22
378	1	1
orf1ab		
p.V38A		
2258	1	0.99
orf1ab		
p.V665I		
3037	1	1
orf1ab		
p.F924F		
3967	1	1
orf1ab		
p.A1234A		
5648	1	1
orf1ab		
p.K1795Q		
6860	1	1
orf1ab		
p.A2199T		
8140	1	1
orf1ab		
p.S2625S		
9204	1	1
orf1ab		
p.D2980G		
11081	0.99	0.99
orf1ab		
p.L3606V		
13647	1	1
orf1ab		
p.T4461I		
14408	0.96	0.97
orf1ab		
p.L4715L		
20268	1	1
orf1ab		
p.*6668Wext*?		

Position	Jun-22	Aug-22
21603	0.99	0.99
S		
p.Q14L		
22012	0.98	0.99
S		
p.K150N		
22202	1	1
S		
p.R214S		
22206	1	1
S		
p.D215G		
22713	0.99	0.99
S		
p.P384L		
22812	0.98	1
S		
p.K417T		
22907	0.98	1
S		
p.Y449H		
22942	1	1
S		
p.N460K		
22992	0.98	0.98
S		
p.S477N		
23012	1	0.98
S		
p.E484P		
23018	0.99	0.99
S		
p.F486A		
23039	1	1
S		
p.Q493K		
23403	1	1
S		
p.D614G		
24380	0.98	1
S		
p.S940T		
25020	1	1
S		
p.D1153A		
25088	1	1
S		
p.V1176F		
25116	0.99	1
S		
p.R1185H		
25254	1	1
S		
p.T1231I		

Position	Jun-22	Aug-22
25846	1	1
ORF3a		
p.N152H		
25936	1	1
ORF3a		
p.H182D		
25979	1	1
ORF3a		
p.G196E		
25991	1	1
ORF3a		
p.C200Y		
26527	0.99	1
M		
p.A2E		
26538	0.99	1
M		
p.G6C		
26544	0.96	0.96
M		
p.I8delinsSNNSEF		
26571	1	1
M		
p.L17V		
27285	0.99	1
ORF6		
p.N28K		
27634	0.97	1
ORF7a		
p.S81P		
28321	1	1
N		
p.T16T		
28382	1	1
N		
p.S37P		
28854	1	1
N		
p.S194L		
29249	0.99	1
N		
p.P326S		
29758	1	1
non-coding		

Supplemental Figure 1. Shared SARS-CoV-2 variants in the June and August 2022 facility waste-water sequences. Illumina sequencing data of overlapping PCR amplicons were processed using the nf-core viral-recon workflow. Variants present in both timepoints at an allele frequency of at least 95% are shown along with their location in the genome and predicted protein impact.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Supplemental online methods

Ethics statement

This activity was reviewed by CDC and the Wisconsin Department of Health Services, and was conducted consistent with applicable federal law and CDC policy (see, eg, 45 C.F.R. part 46, 21 C.F.R. part 56; 42 U.S.C. §241(d); 5 U.S.C. §552a; 44 U.S.C. §3501 et seq).

Collection of wastewater

Wastewater samples for this study (January 2022 through September 2022) were collected in collaboration with experienced wastewater engineers from the city wastewater utility. The Wisconsin State Laboratory of Hygiene (WSLH) worked closely with the city engineering staff to identify sampling locations within the wastewater collection system that would provide the required delineation of service area for each round of sampling as the cryptic variant source region was incrementally narrowed to its source. Sewage lift-stations, manholes and facility line access points were sampled with compositing autosamplers (ISCO 6712 and 6712c). Appropriate safety precautions (e.g. toxic gas and oxygen testing) and PPE (gloves, face shield and protective suit) were donned before entering a manhole. Depending upon manhole depth, the autosampler was either placed on a shelf adjacent to the wastestream or suspended from the manhole opening, with weighted collection lines placed into the wastewater stream. The autosamplers were programmed to collect 24-hr composites, on typically a time-based mode, with wastewater composited into a 10L polypropylene container. The composite was kept cool during collection with ice packed around the collection container. Composite samples were transported to the analytical laboratory within a few hours of sample retrieval.

Isolation of viral RNA from wastewater

Two approaches were used to isolate viral RNA from wastewater.

For samples processed at WSLH, laboratory wastewater samples (homogenized and unfiltered) were spiked with 20 µL/250 mL Calf-Guard® (Zoetis, Parsippany, NJ, USA), a cattle vaccine containing Bovine Coronavirus (BCoV) (as a virus recovery control), and briefly stored at 4°C until the viral targets were isolated and concentrated, typically on the day of receipt. A total of 10 mL (2x5mL) of wastewater was concentrated using Nanotrap Magnetic Virus Particles, Microbiome A and Enhancing Reagent 2 (Ceres Nanosciences, Manassas, VA, USA), using a KingFisher Apex automation platform. Total nucleic acids (TNA) were extracted using Maxwell(R) HT Environmental TNA kits (Promega, Madison,

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

WI, USA) and eluted in 200 µL of 25 mM Tris HCl (pH 8.0) buffer. The extraction was automated using a KingFisher Flex (ThermoFisher Scientific, Waltham, MA, USA).

For samples processed at the University of Missouri, samples were processed as previously described [17]. Briefly, wastewater samples were centrifuged at 3000×g for 10 min and filtered through a 0.22 µM polyethersulfone membrane (Millipore, Burlington, MA, USA). Approximately 37.5 mL of wastewater was mixed with 12.5 mL solution containing 50% (w/vol) polyethylene glycol 8000 and 1.2 M NaCl, mixed, and incubated at 4°C for at least 1 h. Samples were then centrifuged at 12,000×g for 2 h at 4°C. Supernatant was decanted and RNA was extracted from the remaining pellet (usually not visible) with the QIAamp Viral RNA Mini Kit (Qiagen, Germantown, MD, USA) using the manufacturer's instructions. RNA was extracted in a final volume of 60 µL.

Quantification of viral RNA by RT-dPCR

Quantification of SARS-CoV-2, BCoV, and PMMoV (fecal marker) was achieved using reverse transcriptase digital PCR (RT-dPCR). Master mix was prepared using the One-Step Viral PCR kit (4x) (Qiagen, Germantown, MD, USA) and GT dPCR SARS-CoV-2 Wastewater Surveillance Assay Kit (GT Molecular, Fort Collins, CO, USA) with quantification of the following viral targets: N1, N2, BCoV, and PMMoV included with the GTMolecular dPCR SARS-CoV-2 Wastewater Surveillance Assay Kit, and BRSV (our spiked inhibition assay target) primers and probes from IDT [18]. The samples were run in triplicate on a QIAcuity Four Digital PCR System (Qiagen, Germantown, MD, USA). N1, N2, and BCoV were multiplexed on QIAcuity Nanoplate 26k 24-well plates while PMMoV and BRSV were single-plexed on 8.5k 96-well nanoplates. Cycling and exposure conditions are detailed in the table shown below. Analysis of the RT-dPCR results was performed with the QIAcuity Software Suite version 2.1.7.182. Thresholds were manually set to separate negative and positive partitions.

dPCR Thermocycling Conditions

Thermocycling Conditions:			
Step		**Time**	**Temp ºC**
Reverse Transcription		30 min	50
DNA polymerase activation		2 min	95
45 cycles	Denaturation	10 sec	95
	Anneal/Extend	30 sec	55

Target	Channel	Exposure	Gain

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

N1	Red (ROX)	500	4
N2	Green (FAM)	300	6
BCoV	Yellow (HEX)	300	6
PMMoV	Green (FAM)	300	6
BRSV	Yellow (HEX)	500	6

Identification of cryptic lineages in wastewater with non-Omicron PCR amplification and amplicon sequencing

The primary RBD RT-PCR was performed using the Superscript IV One-Step RT-PCR System (Thermo Fisher Scientific,12594100). Primary RT-PCR amplification was performed as follows: 25 °C (2:00) + 50 °C (20:00) + 95 °C (2:00) + [95 °C (0:15) + 55 °C (0:30) + 72 °C (1:00)] × 25 cycles using the MiSeq primary PCR primers ATTCTGTCCTATATAATTCCGCAT and CCCTGATAAAGAACAGCAACCT (the first primer was changed to TATATAATTCCGCATCATTTTCCAC starting in May, 2022 to adapt to changing Omicron lineages). Secondary PCR (25 µL) was performed on RBD amplifications using 5 µL of the primary PCR as template with MiSeq nested gene specific primers containing 5′ adapter sequences (0.5 µM each) acactctttccctacacgacgctcttccgatctGTGATGAAGTCAGACAAATCGC and gtgactggagttcagacgtgtgctcttccgatctATGTCAAGAATCTCAAGTGTCTG, dNTPs (100 µM each) (New England Biolabs, N0447L) and Q5 DNA polymerase (New England Biolabs, M0541S). Secondary PCR amplification was performed as follows: 95 °C (2:00) + [95 °C (0:15) + 55 °C (0:30) + 72 °C (1:00)] × 20 cycles. A tertiary PCR (50 µL) was performed to add adapter sequences required for Illumina cluster generation with forward and reverse primers (0.2 µM each), dNTPs (200 µM each) (New England Biolabs, N0447L) and Phusion High-Fidelity or (KAPA HiFi for CA samples) DNA Polymerase (1U) (New England Biolabs, M0530L). PCR amplification was performed as follows: 98 °C (3:00) + [98 °C (0:15) + 50 °C (0:30) + 72 °C (0:30)] × 7 cycles +72 °C (7:00). Amplified product (10 µl) from each PCR reaction is combined and thoroughly mixed to make a single pool. Pooled amplicons were purified by the addition of Axygen AxyPrep MagPCR Clean-up beads (Axygen, MAG-PCR-CL-50) or in a 1.0 ratio to purify final amplicons. The final amplicon library pool was evaluated using the Agilent Fragment Analyzer automated electrophoresis system, quantified using the Qubit HS dsDNA assay (Invitrogen), and diluted according to Illumina's standard protocol. The Illumina MiSeq instrument was used to generate paired-end 300 base pair reads. Adapter sequences were trimmed from output sequences using Cutadapt.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Sequencing reads were processed as previously described. Briefly, VSEARCH tools were used to merge paired reads and dereplicate sequences [19]. Dereplicated sequences from RBD amplicons were mapped to the reference sequence of SARS-CoV-2 (NC_045512.2) spike ORF using Minimap2 [20]. Mapped amplicon sequences were then processed with SAM Refiner using the same spike sequence as a reference and the command line parameters "--Alpha 1.8 --foldab 0.6" [21].
The haplotypes representing at least 25% of the total sequences in at least one sample were rendered into figures using plotnine https://plotnine.readthedocs.io/en/stable/index.html.

Full-genome sequencing of wastewater

Whole genome sequencing of the wastewater samples was performed at the WSLH. Briefly, 13 µL of TNA from the wastewater extracts isolated as described above, were used as input to QIAGEN's Direct SARS-CoV-2 Kit A. Libraries were prepared on a Biomek i5 liquid handler (Beckman Coulter, Brea, CA, USA). Libraries were quantified using a High Sensitivity Qubit 1X dsDNA HS Assay Kit (ThermoFisher Scientific) and fragment size analyzed by a QIAxcel Advanced and the QX DNA Screening Kit (QIAGEN, Germantown, MD, USA).

The data were analyzed with the nf-core/viralrecon workflow (https://nf-co.re/viralrecon/2.5) using the SARS-CoV-2 Wuhan-Hu-1 reference genome (Genbank accession MN908947.3) and the QIAseq Direct SARS-CoV-2 primer .bed file (https://www.qiagen.com/us/products/next-generation-sequencing/rna-sequencing/qiaseq-direct-sars-cov-2-kits/). After creating a sample sheet as described on the nf-core/viralrecon website (https://nf-co.re/viralrecon/usage) the workflow was initiated using the following lines of code:

```
# create conda environment
conda create --name nextflow -c bioconda nextflow
# activate environment
conda activate nextflow
# process data
nextflow run nf-core/viralrecon \
--input resources/27660-samples.csv \
--outdir results \
--platform illumina \
--protocol amplicon \
--genome 'MN908947.3' \
--primer_bed resources/QIAseqDIRECTSARSCoV2primersfinal.bed \
--primer_left_suffix '_LEFT' \
--primer_right_suffix '_RIGHT' \
--ivar_trim_offset 5 \
--skip_assembly \
-profile docker
```

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Virus culture

To remove debris, samples were centrifuged twice at 3,500 rpm at 4oC for 15 minutes and then passed through a 0.8 μM syringe filter (Agilent) or left unfiltered. Samples (1ml) were incubated on nearly confluent Vero E6-TMPRSS2 (JCRB1819) or Vero E6-TMPRSS2/hACE2 cells (from NIH Barney Graham) seeded the day prior in TC252 cm flasks for 1 hour at 370C. After the incubation, cells were washed twice and media was added back to the cells. The media contained 2-times the normal concentration of penicillin, streptomycin and amphotericin along with chloramphenicol. Cells were monitored daily for potential virus-induced cytopathic effects. After 10 days, a blind passage was performed using the entire volume of media (~4 ml) to fresh, nearly confluent cells seeded the day prior in TC1752 cm flasks.

Root-to-tip regression

Full consensus genomes for SARS-CoV-2 specimens collected in the Midwest region (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin) were downloaded from GISAID under the GISAID identifier EPI_SET_221024fg (doi: 10.55876/gis8.221024fg). The dataset is composed of 9,585 individual genome sequences collected between 2022-05-22 to 2022-08-30 and classified as either Omicron or B.1.234. The dataset contained 1,438 sequences classified as B.1.234, representing all the available B.1.234 sequences for the Midwest region between the dates 2020-05-22 to 2021-12-24. The remaining sequences were Omicron (BA.*) sequences collected between 2022-05-01 and 2022-08-31. Dataset was filtered to contain no more than 50 sequences per state and the multiple alignment was generated using MAFFT (v7.505) using the most ancestral sequence of good quality (EPI_ISL_13597234). A maximum likelihood phylogenetic tree was inferred using iqtree (v.2.2.0.3) with a molecular clock and distances obtained through treetime (v0.9.3). Distances were obtained for B.1.234 sequences only, and then for both B.1.234 and Omicron sequences. Plot (Figure 3a) was generated in R Studio using ggplot2 and dplyr libraries. Phylogeny was visualized and annotated with FigTree (v.1.4.4). Scripts are available in the github repository accompanying this manuscript (https://github.com/tcflab/wisconsin_cryptic_lineages).

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Viral diversity analyses

Variants obtained through the nf-core/viralrecon workflow were processed using custom Python scripts (see Data Availability) to generate panels b-d in **Figure 3**. Variants differing from reference sequence MN908947.3 were classified as non-synonymous (Non-syn), synonymous (Syn), insertions-deletions (indels), or others (including nonsense and frameshift mutations). Synonymous and non-synonymous point mutations were quantified and compared between timepoints, and 95% confidence intervals obtained from the relative risk (RR) of every nucleotide substitution against its inverted change (i.e., RR= `A>C` / `C>A`) using SciPy's relative_risk function (v.1.9.3). To obtain the proportion of variants per site, we enumerated the synonymous and non-synonymous substitutions across the SARS-CoV-2 genome, and obtained the proportion against the number of synonymous and non-synonymous sites, respectively. A binomial probability distribution was implemented to obtain the 95% confidence intervals via SciPy's binomtest function (v.1.9.3). Scripts are available in the github repository accompanying this manuscript (https://github.com/tcflab/wisconsin_cryptic_lineages).

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.22281553; this version posted October 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

References

1. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;581(7807):215-220. doi:10.1038/s41586-020-2180-5
2. Abdool Karim SS, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *Lancet*. 2021;398(10317):2126. doi:10.1016/S0140-6736(21)02758-6
3. Cov-Lineages. Accessed July 27, 2022. https://cov-lineages.org/global_report_BA.1.html
4. Smyth DS, Trujillo M, Gregory DA, et al. Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. *Springer Nature*. Published online February 3, 2022:1-9. doi:10.1038/s41467-022-28246-3
5. Gregory DA, Trujillo M, Rushford C, et al. Genetic Diversity and Evolutionary Convergence of Cryptic SARS-CoV-2 Lineages Detected Via Wastewater Sequencing. *medRxiv : the preprint server for health sciences*. Published online June 3, 2022. doi:10.1101/2022.06.03.22275961
6. Pérez-Cataluña A, Chiner-Oms Á, Cuevas-Ferrando E, et al. Spatial and temporal distribution of SARS-CoV-2 diversity circulating in wastewater. *Water Res*. 2022;211:118007. doi:10.1016/j.watres.2021.118007
7. Corey L, Beyrer C, Cohen MS, Michael NL, Bedford T, Rolland M. SARS-CoV-2 Variants in Patients with Immunosuppression. *N Engl J Med*. 2021;385(6):562-566. doi:10.1056/NEJMsb2104756
8. Wilkinson SAJ, Richter A, Casey A, et al. Recurrent SARS-CoV-2 mutations in immunodeficient patients. *Virus Evol*. 2022;8(2):veac050. doi:10.1093/ve/veac050
9. Fritz M, Rosolen B, Krafft E, et al. High prevalence of SARS-CoV-2 antibodies in pets from COVID-19+ households. *One Health*. 2021;11. doi:10.1016/j.onehlt.2020.100192
10. Chandler JC, Bevins SN, Ellis JW, et al. SARS-CoV-2 exposure in wild white-tailed deer (Odocoileus virginianus). *Proc Natl Acad Sci U S A*. 2021;118(47). doi:10.1073/pnas.2114828118
11. Rasmussen TB, Fonager J, Jørgensen CS, et al. Infection, recovery and re-infection of farmed mink with SARS-CoV-2. *PLoS Pathog*. 2021;17(11):e1010068. doi:10.1371/journal.ppat.1010068
12. Lu L, Sikkema RS, Velkers FC, et al. Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. *Nat Commun*. 2021;12(1):6802. doi:10.1038/s41467-021-27096-9
13. Bb OM, Sikkema RS, Nieuwenhuijse DF, et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*. 2021;371(6525). doi:10.1126/science.abe5901
14. outbreak.info. outbreak.info. Accessed September 8, 2022. https://outbreak.info/
15. Jörrißen P, Schütz P, Weiand M, et al. Antibody Response to SARS-CoV-2 Membrane Protein in Patients of the Acute and Convalescent Phase of COVID-19. *Front Immunol*. 2021;0. doi:10.3389/fimmu.2021.679841
16. Heffron AS, McIlwain SJ, Amjadi MF, et al. The landscape of antibody binding in SARS-CoV-2 infection. *bioRxiv*. doi:10.1101/2020.10.10.334292
17. Gregory DA, Trujillo M, Rushford C, et al. Genetic diversity and evolutionary convergence of cryptic SARS- CoV-2 lineages detected via wastewater sequencing. *PLoS Pathog*. 2022;18(10):e1010636. doi:10.1371/journal.ppat.1010636
18. Boxus M, Letellier C, Kerkhofs P. Real Time RT-PCR for the detection and quantitation of bovine respiratory syncytial virus. *J Virol Methods*. 2005;125(2). doi:10.1016/j.jviromet.2005.01.008
19. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4. doi:10.7717/peerj.2584
20. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100. doi:10.1093/bioinformatics/bty191
21. Gregory DA, Wieberg CG, Wenzel J, Lin CH, Johnson MC. Monitoring SARS-CoV-2 Populations in Wastewater by Amplicon Sequencing and Using the Novel Program SAM Refiner. *Viruses*. 2021;13(8):1647. doi:10.3390/v13081647