

1 **Continued selection on cryptic SARS-CoV-2 observed in Missouri wastewater**

2

3 Devon A. Gregory¹, Clayton Rushford¹, Chung-Ho Lin², Christie Darby³, Nicole Niehues³,

4 Elizabeth Semkiw³, Melissa Reynolds³, Jeff Wenzel³, Marc C. Johnson^{1*}

5

6 ¹Department of Molecular Microbiology and Immunology, University of Missouri-School of

7 Medicine, Columbia, MO, USA

8 ²Center of Agroforestry, School of Natural Resources, University of Missouri, Columbia, MO,

9 USA

10 ³Bureau of Environmental Epidemiology, Division of Community and Public Health, Missouri

11 Department of Health and Senior Services, Jefferson City, MO, USA

12 *marcjohanson@missouri.edu

13

14 **Short title:** SARS-CoV-2 wastewater surveillance

15 Author Contributions: DAG , CR , CL , MR , JW, ES, NN , CD, MCJ

16

17

18 **Abstract**

19 Deep sequencing of wastewater to detect SARS-Cov-2 has been used during the COVID-19

20 pandemic to monitor viral variants as they appear and circulate in communities. SARS-CoV-2

21 lineages of an unknown source that have not been detected in clinical samples, referred to as

22 cryptic lineages, are sometimes repeatedly detected in specific locations. We have continued to

23 detect one such lineage previously seen in a Missouri site. This cryptic lineage has continued to

24 evolve, indicating continued selective pressure similar to that observed in Omicron lineages.

25

26 **Author Summary**

27 Monitoring sewage for SARS-CoV-2 has been an important part of understanding the dynamics
28 of the virus's spread and persistence within and across communities during the pandemic. We
29 and others have also observed variants appearing in wastewater that do not appear in clinical
30 sampling. Many of these variants not only possess genomic changes identical to or at the
31 same position as those that have been observed in variants of concern, particularly currently
32 circulating Omicron variants, but often acquire the changes before they have been observed in
33 clinical samples. We report here the continued observation of a variant in Missouri wastewater,
34 but not in clinical sampling, that has continued to evolve, gaining genomic changes that often
35 are the same and predate changes seen in clinical samples. These observation add to our
36 understanding of the selective pressures driving the evolution of SAR-CoV-2.

37

38 **Introduction**

39 Surveillance of wastewater for SARS-CoV-2 has been used to detect and track community
40 circulating variants^{1,2}. In addition to the variants of concern (VOCs) and other common variants,
41 wastewater surveillance has also detected variants that have not been otherwise observed^{3,4}.
42 These novel variants, which we call cryptic lineages, often persist in a sewershed for months or
43 years and show signs of continued positive selection. The specific sources of cryptic lineages
44 are unknown, though recent efforts have provided evidence of a human source⁵. Sequencing of
45 immunocompromised individuals with persistent infection has also detected sequences with
46 some similarities to the cryptic lineages⁶. However, a non-human source for some of the cryptic
47 lineages observed in wastewater can not be ruled out. We have previously reported on a
48 cryptic lineage found in a Missouri metropolitan area (MO45) in June of 2021⁴. Since the initial
49 observation of this cryptic lineage, it has been sporadically detected with evolving genotypes.

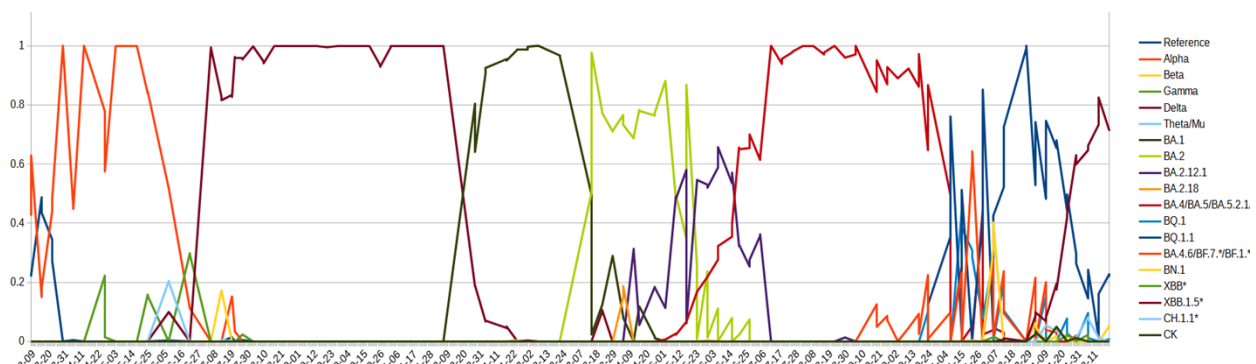
50

51 **Results and Discussion**

52 We use next generation sequencing of SARS-CoV-2's RBD to monitor variants present in
53 Missouri wastewater. Monitoring of MO45 began in March 2021 and continues to the present
54 with roughly weekly sampling (Fig 1). Initially this sewershed was observed to primarily have
55 the Alpha variant with some ancestral sequences. Beta, Gamma, Delta and Mu/Theta
56 sequences were all observed later with Delta becoming the only variant detected by August
57 2021. Delta was then rapidly replaced by Omicron in December 2021. Since, various Omicron
58 variants have circulated, generally with newer variants displacing older ones, resulting in a
59 mixture of variants co-circulating in late 2022.

60

61

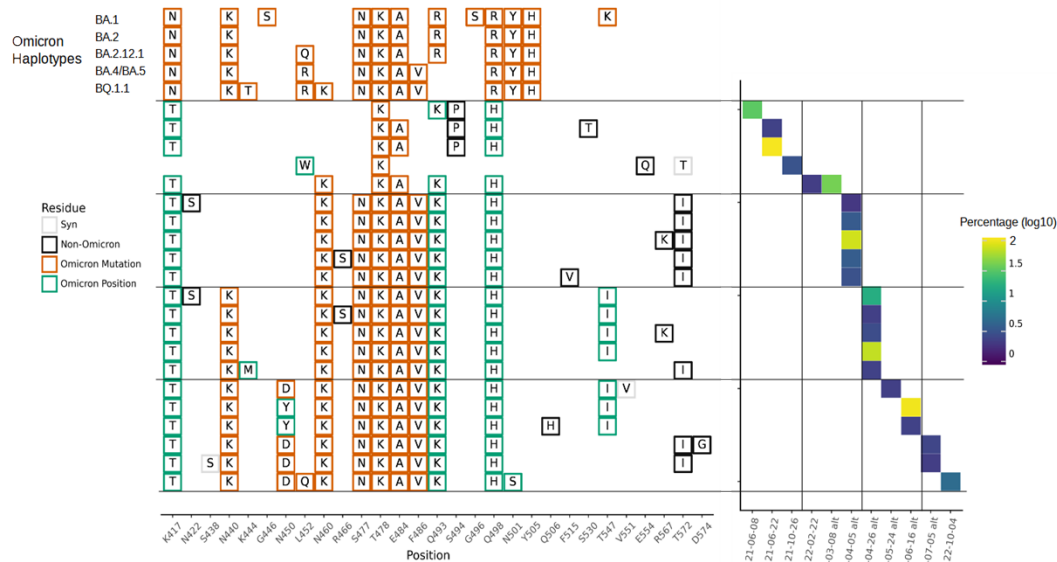


62

63 **Fig 1. Presence of Variants Over Time.** Plot of relative defined SARS-CoV-2 variant
64 abundance detected in MO45 sewershed. Wastewater samples were processed for viral RNA
65 extraction and SARS-CoV-2 RBD sequences were amplified and then deep sequenced.
66 Genotype and number of reads were used to estimate the relative abundance of defined
67 variants present in the sewershed.
68

69 In addition to the defined variants, a cryptic lineage has also been sporadically detected, first in
70 June 2021 and last in October 2022 (Fig 2). Initial sequences of this variant had K417T T478K
71 E484A Q493K S494P Q498H amino acid changes relative to the ancestral sequence, with
72 E484A and Q493K only appearing in one of the two first detections. E484A and Q493K were
73 both observed in all subsequent sequences of this cryptic lineage, while S494P was not

74 observed again. K417T and T478K had previously been observed in the Gamma and Delta
75 variants respectively, but the other mutations had not yet appeared in any major VOCs.
76



77
78 **Fig 2. Cryptic Haplotypes Detected in MO45.** Cryptic haplotypes are shown with the dates of
79 their detection and relative abundance in the sample amplification. Select Omicron haplotypes
80 are shown for comparison. Detection dates followed by 'alt' indicate the amplification was
81 carried out with primers designed to exclude Omicron lineages for better detection of cryptic
82 sequences.
83

84 Several amino acid changes occurred subsequent to the initial observation and appeared to
85 become fixed in the lineage. On February 2, 2022 N460K was first observed in the cryptic
86 lineage and was thereafter fixed. Likewise S477N and F486V were first observed in the cryptic
87 lineage on April 5, 2022, and N440K on April 26, 2022, and in all detections since. N450D was
88 first observed in the cryptic lineage on May 24, 2022. Though the lineage had N450Y on June
89 16, 2022 instead, the two subsequent detections of the lineage had N450D again. Several
90 other changes were observed in the lineage over time, though none could be concluded to have
91 become fixed. Of note, T547I and T572I both were observed in 3 samples each, but were not
92 observed in the most recent detection. At the last detection of the MO45 cryptic lineage in
93 October 2022, the lineage appeared remarkably similar to an Omicron lineage with 12 amino

94 acid changes in its RBD that were all identical to, or at the same position as, changes found in
95 Omicron lineages.
96 Most of the residue changes observed in the cryptic lineage predate the changes observed in
97 Omicron. The initial detections of the cryptic lineage, months before the emergence of Omicron,
98 already had two changes that were to be seen ubiquitously in Omicron lineages, T478K and
99 E484A, and three changes at the same residues as changes common in Omicron lineages,
100 K417T, Q493K and Q498H. Likewise, N460K, which appeared in the cryptic lineage in
101 February of 2022, did not become prevalent in an Omicron background until six months later.
102 The convergence of the cryptic lineage and Omicron variants suggest similar selection
103 pressures. The origin of Omicron and the origin of the MO45 cryptic lineage are unknown. At
104 least in some cases, cryptic lineages appear to be derived from individuals with chronic SARS-
105 CoV-2 infections. However, as the MO45 cryptic lineage hasn't been traced, a non-human
106 source can not be ruled out. Since the cryptic lineage in some cases acquired changes prior to
107 Omicron, continued monitoring of waste water for such cryptic lineages may provide
108 foreknowledge of changes, or at least the position of changes, likely to be selected for in the
109 circulating Omicron variants.

110

111

112 **Materials & Methods**

113 **Sample collection and RNA extraction.**

114 Collection and processing of samples were as previously described. Twenty-four-hour
115 composite samples were collected at the MO45 wastewater treatment facility and maintained at
116 4 °C until they were delivered to the analysis lab, generally within 24 h of collection. Samples
117 were then centrifuged at 3000× *g* for 10 min and followed by filtration through a 0.22 μM
118 polyethersulfone membrane (Millipore, Burlington, MA, USA). Approximately 37.5 mL of
119 wastewater was mixed with 12.5 mL solution containing 50% (*w/vol*) polyethylene glycol 8000

120 and 1.2 M NaCl, mixed, and incubated at 4 °C for at least 1 h. Samples were then centrifuged at
121 12,000× g for 2 h at 4 °C. Supernatant was decanted and RNA was extracted from the
122 remaining pellet (usually not visible) with the QIAamp Viral RNA Mini Kit (Qiagen, Germantown,
123 MD, USA) using the manufacturer's instructions. RNA was extracted in a final volume of 60 µL.

124 **MiSeq**

125 Similar to our previous protocol, the primary RBD RT-PCR was performed using the Superscript
126 IV One-Step RT-PCR System (Thermo Fisher Scientific, 12594100). Primary RT-PCR
127 amplification was performed as follows: 25°C (2:00) + 50°C (20:00) + 95°C (2:00) + [95°C (0:15)
128 + 55°C (0:30) + 72°C (1:00)] × 25 cycles using the MiSeq primary PCR primers
129 CTGCTTTACTAATGTCTATGCAGATTC and NCCTGATAAAGAACAGCAACCT. Secondary
130 PCR (25 µL) was performed on RBD amplifications using 5 µL of the primary PCR as template
131 with MiSeq nested gene specific primers containing 5' adapter sequences (0.5 µM each)
132 acactctttcctacacgacgctctccgatctGTRATGAAGTCAGMCAAATYGC and
133 gtgactggagttcagacgtgtgctctccgatctATGTCAAGAATCTCAAGTGTCTG, dNTPs (100 µM each)
134 (New England Biolabs, N0447L) and Q5 DNA polymerase (New England Biolabs, M0541S).
135 Secondary PCR amplification was performed as follows: 95°C (2:00) + [95°C (0:15) + 55°C
136 (0:30) + 72°C (1:00)] × 20 cycles.

137 For Omicron exclusion amplification, the primary RBD RT-PCR was performed using the MiSeq
138 primary PCR primers ATTCTGTCCTATATAATTCCGCAT and
139 CCCTGATAAAGAACAGCAACCT (the first primer was changed to
140 TATATAATTCCGCATCATTTTCCAC starting in May, 2022 to adapt to changing Omicron
141 lineages) and secondary PCR used MiSeq nested gene specific primers containing 5' adapter
142 sequences (0.5 µM each) acactctttcctacacgacgctctccgatctGTGATGAAGTCAGACAAATCGC
143 and gtgactggagttcagacgtgtgctctccgatctATGTCAAGAATCTCAAGTGTCTG.

144 A tertiary PCR (50 µL) was performed to add adapter sequences required for Illumina cluster
145 generation with forward and reverse primers (0.2 µM each), dNTPs (200 µM each) (New

146 England Biolabs, N0447L) and Phusion High-Fidelity or (KAPA HiFi for CA samples) DNA
147 Polymerase (1U) (New England Biolabs, M0530L). PCR amplification was performed as follows:
148 98°C (3:00) + [98°C (0:15) + 50°C (0:30) + 72°C (0:30)] × 7 cycles +72°C (7:00). Amplified
149 product (10 µl) from each PCR reaction is combined and thoroughly mixed to make a single
150 pool. Pooled amplicons were purified by addition of Axygen AxyPrep MagPCR Clean-up beads
151 (Axygen, MAG-PCR-CL-50) or in a 1.0 ratio to purify final amplicons. The final amplicon library
152 pool was evaluated using the Agilent Fragment Analyzer automated electrophoresis system,
153 quantified using the Qubit HS dsDNA assay (Invitrogen), and diluted according to Illumina's
154 standard protocol. The Illumina MiSeq instrument was used to generate paired-end 300 base
155 pair reads. Adapter sequences were trimmed from output sequences using Cutadapt.

156 **Computational analysis**

157 Sequencing reads were processed similar to previously described. Briefly, BBTools (Bushnell
158 B. – <http://sourceforge.net/projects/bbmap/>) were used to merge paired reads, which were
159 dereplicated with a custom script (<https://github.com/degregory/Programs/blob/main/derep.py>).
160 Dereplicated sequences from RBD amplicons were mapped to the reference sequence of
161 SARS-CoV-2 (NC_045512.2) spike ORF using Minimap2. Mapped amplicon sequences were
162 then processed with SAM Refiner using the same spike sequence as a reference and the
163 command line parameters "--Alpha 1.6 --foldab 0.6".

164 For Fig 1, SAM Refiner covariant deconvolution outputs were matched to defined variants to
165 determine the relative abundance for each sample using a custom script (modified from
166 <https://github.com/istaves/covid-variant-counter>). For Fig 2, the same outputs of SAM Refiner
167 for MiSeq sequences were collected and were processed to determine core haplotypes of the
168 cryptic lineage. First sequences that contained fewer than 4 polymorphisms relative to the
169 reference Wuhan I sequence or matched officially named variants were discarded. Remaining
170 sequences were then processed to remove polymorphisms that never appeared in a sample at

171 an abundance greater than .5%. In-frame deletions bypassed this removal. Condensed
172 sequences that appear in at least two samples or had a summed abundance of at least 2%
173 across all samples were passed on to further steps. All were sequences rendered into the
174 figures using plotnine.

175 **Author Contributions**

176 DAG, MCJ, MR, CL, ES, and JW conceptualized and designed the study. DAG and MCJ
177 performed formal analysis. JW, ES, MR, CR, CL, NN, CD, DAG and MCJ conducted the
178 investigation. CR performed the sequencing. DAG wrote all the code. MCJ, CL, MR, and JW
179 acquired funding. MCJ, DAG, JW, and MR writing and review.

180 **Acknowledgements**

181 We would like to acknowledge the University of Missouri Bioinformatics and Analytics Core for
182 their services with MiSeq sequencing.

183

184

185 **References**

- 186 1. Trujillo M, Cheung K, Gao A, Hoxie I, Kannoly S, Kubota N, et al. (2021) Protocol for
187 safe, affordable, and reproducible isolation and quantitation of SARS-CoV-2 RNA from
188 wastewater. PLoS ONE 16(9): e0257454. <https://doi.org/10.1371/journal.pone.0257454>
- 189 2. Gregory, D.A.; Wieberg, C.G.; Wenzel, J.; Lin, C.-H.; Johnson, M.C. Monitoring SARS-
190 CoV-2 Populations in Wastewater by Amplicon Sequencing and Using the Novel
191 Program SAM Refiner. *Viruses* **2021**, *13*, 1647. <https://doi.org/10.3390/v13081647>
- 192 3. Smyth, D.S., Trujillo, M., Gregory, D.A. *et al.* Tracking cryptic SARS-CoV-2 lineages
193 detected in NYC wastewater. *Nat Commun* **13**, 635 (2022).
194 <https://doi.org/10.1038/s41467-022-28246-3>

- 195 4. Gregory DA, Trujillo M, Rushford C, Flury A, Kannoly S, San KM, et al. (2022) Genetic
196 diversity and evolutionary convergence of cryptic SARS- CoV-2 lineages detected via
197 wastewater sequencing. PLoS Pathog 18(10): e1010636.
198 <https://doi.org/10.1371/journal.ppat.1010636>
- 199 5. Shafer MM, Bobholz MJ, et al. (2023) Tracing the origin of SARS-CoV-2 Omicron-like
200 spike sequences detected in wastewater. medRxiv
201 <https://doi.org/10.1101/2022.10.28.22281553>
- 202 6. S A J Wilkinson and others, The COVID-19 Genomics UK (COG-UK) consortium,
203 Recurrent SARS-CoV-2 mutations in immunodeficient patients, *Virus Evolution*, Volume
204 8, Issue 2, 2022, veac050, <https://doi.org/10.1093/ve/veac050>
205
206
207