Detection of SARS-CoV-2 B.1.351 (Beta) Variant through Wastewater Surveillance before Case Detection in a Community, Oregon, USA

Melissa Sutton, Tyler S. Radniecki, Devrim Kaya, Dana Alegre, Matthew Geniza, Anne-Marie Girard, Katherine Carter, Mark Dasenko, Justin L. Sanders, Paul R. Cieslak, Christine Kelly, Brett M. Tyler

Genomic surveillance has emerged as a critical monitoring tool during the SARS-CoV-2 pandemic. Wastewater surveillance has the potential to identify and track SARS-CoV-2 variants in the community, including emerging variants. We demonstrate the novel use of multilocus sequence typing to identify SARS-CoV-2 variants in wastewater. Using this technique, we observed the emergence of the B.1.351 (Beta) variant in Linn County, Oregon, USA, in wastewater 12 days before this variant was identified in individual clinical specimens. During the study period, we identified 42 B.1.351 clinical specimens that clustered into 3 phylogenetic clades. Eighteen of the 19 clinical specimens and all wastewater B.1.351 specimens from Linn County clustered into clade 1. Our results provide further evidence of the reliability of wastewater surveillance to report localized SARS-CoV-2 sequence information.

Since its emergence in late 2019, more than 481 million COVID-19 cases have been confirmed worldwide (*1*) and >79 million cases reported in the United States (*2*). Numerous variants of the causative virus, SARS-CoV-2, have emerged; variants of concern have demonstrated characteristics of public health concern, including increased transmissibility or clinical severity, reduced vaccine or therapeutic effectiveness, or diagnostic escape (*3*). SARS-CoV-2 genomic surveillance has quickly become an essential tool for tracking transmission and coordinating response (*4*,*5*). Individual-level genomic surveillance relies on the testing of infected persons, which, in turn, requires testing access and acceptance. In the United States, testing access has improved dramatically over the course of the pandemic but remains limited, particularly in disproportionately affected communities (*6*), and testing acceptance remains an obstacle to effective disease mitigation (*7*).

SARS-CoV-2 is shed in feces, and wastewater surveillance has emerged as complementary cost-effective community-level surveillance independent of testing access and acceptance or symptomatic infection (*8*–*10*). Using wastewater testing for SARS-CoV-2 genomic surveillance avoids the testing and symptomatic biases inherent to the sequencing of individual specimens; however, interpreting sequence data from complex mixtures of viruses at a population-level remains challenging (*11*). Multilocus sequence typing (MLST) is a method traditionally used to identify species or variants in environmental samples, including rivers, urban streams, hospital sewage, and wastewater treatment plant influents and effluents (*12*,*13*). MLST is well-suited for the analysis of wastewater RNA because it detects a set of mutations unique to a variant and does not require these mutations to be present on a single molecule of RNA (*12*–*18*).

During the COVID-19 pandemic, Oregon has been among the US states with the lowest cumulative case rates and among those with the highest proportion of cumulative molecular specimens sequenced in the United States (*19*). As of March 31, 2021, a total of 159,455 confirmed cases of COVID-19 had been identified in Oregon, and specimens from 5,674 (3.6%) of the cases had been sequenced and published in the GISAID database (https://www.gisaid.org) (*20*). At that time, the dominant SARS-CoV-2 variant circulating in Oregon was B.1.427/B.1.429 (Epsilon), followed by B.1.2 and B.1.1.7 (Alpha); only 25 B.1.351 (Beta) and 8 P.1 (Gamma) variants had been identified.

Author affiliations: Oregon Health Authority, Portland, Oregon, USA (M. Sutton, P.R. Cieslak); Oregon State University, Corvallis, Oregon (T.S. Radniecki, D. Kaya, D. Alegre, M. Geniza, A.-M. Girard, K. Carter, M. Dasenko, J.L. Sanders, C. Kelly, B.M. Tyler)

DOI: https://doi.org/10.3201/eid2806.211821

On April 19, 2021, Oregon mandated reporting of all SARS-CoV-2 variants of concern to public health authorities; before this, genomic surveillance relied largely on deidentified data submitted to GISAID. Statewide sequencing partners have been asked to submit all individual specimen sequencing results to the State of Oregon phylodynamics resource in GISAID (https://www.gisaid.org/phylodynamics/oregon-usa).

In September 2020, the Oregon Health Authority launched wastewater surveillance in collaboration with the Oregon State University (OSU) Team-Based Rapid Assessment of Community-Level Coronavirus Epidemics (TRACE) project; >40 communities comprising ≈60% of Oregon's population currently participate. Through this program, wastewater samples from the influent of all wastewater treatment facilities are collected at least weekly and sent to OSU for SARS-CoV-2 viral RNA quantification and sequencing of all positive samples with sufficient viral loads. Through this statewide SARS-CoV-2 wastewater surveillance platform, we demonstrate the use of MLST to detect the emergence of the SARS-CoV-2 B.1.351 (Beta) variant in rural Oregon in late March 2021, before its detection in reported cases, illustrating the ability of wastewater-based epidemiology to detect emerging variants of concern.

Methods

Wastewater RNA Extraction

Participating facilities collected wastewater composite samples from Albany(Linn County), Corvallis (Benton County), and Dallas (Polk County), Oregon, USA, during March 26–April 21, 2021, according to routine practice for the Oregon Wastewater Surveillance Program (Table). In brief, 24-hour time-weighted

Table. SARS-CoV-2 variant B.1.351 mutations detected in clinical specimens and wastewater samples in Linn County, Oregon, and surrounding jurisdictions, March-May 2021

			Mutations specific to‡							
			B.1.351			Clade 1			Subclade§	
Sample*	Collection site†	Date	+	–	?	+	–	?	1a	1b
Clinical specimens										
EPI_ISL_1866415	Linn Co.	2021 Mar 29¶	9	0	0	5	1	0	–	–
EPI_ISL_1736521	Linn Co.	2021 Apr 5	9	0	0	6	0	0	+	–
EPI_ISL_1736532	Linn Co.	2021 Apr 5	9	0	0	6	0	0	+	–
EPI_ISL_1737841	Linn Co.	2021 Apr 7	9	0	0	0#	0	0	–	–
EPI_ISL_1964160	Linn Co.	2021 Apr 9	8	0	1	6	0	0	–	+
EPI_ISL_1999265	Linn Co.	2021 Apr 12	9	0	0	5	1	0	–	–
EPI_ISL_2202145	Linn Co.	2021 Apr 16	8	0	1	4	0	2	+	–
EPI_ISL_2139637	Linn Co.	2021 Apr 26	9	0	0	6	0	0	+	–
EPI_ISL_2139638	Linn Co.	2021 Apr 26	9	0	0	6	0	0	+	–
EPI_ISL_2139639	Linn Co.	2021 Apr 26	9	0	0	6	0	0	+	–
EPI_ISL_2139644	Linn Co.	2021 Apr 27	9	0	0	6	0	0	+	–
EPI_ISL_2250177	Linn Co.	2021 Apr 27	8	0	1	6	0	0	–	+
EPI_ISL_2086679	Linn Co.	2021 Apr 28	9	0	0	6	0	0	+	–
EPI_ISL_2086678	Linn Co.	2021 Apr 28	9	0	0	6	0	0	+	–
EPI_ISL_2139636	Linn Co.	2021 Apr 30	9	0	0	6	0	0	+	–
EPI_ISL_2086694	Linn Co.	2021 Apr 30	9	0	0	6	0	0	+	–
EPI_ISL_2339336	Linn Co.	2021 May 10	9	0	0	6	0	0	+	–
EPI_ISL_2382524	Linn Co.	2021 May 12	9	0	0	6	0	0	+	–
EPI_ISL_2382527	Linn Co.	2021 May 14	9	0	0	6	0	0	+	–
Wastewater samples										
ALB-Inf-03-26-21-A	Albany, Linn Co.	2021 Mar 26	7	1	1	3	1	2	–	+
ALB-Inf-03-31-21-A	Albany, Linn Co.	2021 Mar 31	9	0	0	5	0	1	+	+
COR-25th-04-04-21-A	Corvallis, Benton Co.	2021 Apr 4	9	0	0	6	0	0	+	–
COR-26th-04-04-21-A	Corvallis, Benton Co.	2021 Apr 4	6	2	1	5	0	1	–	–
ALB-Inf-04-07-21-A	Albany, Linn Co.	2021 Apr 7	8	1	0	6	0	0	+	+
DAL-Inf-04-19-21-A	Dallas, Polk Co.	2021 Apr 19	9	0	0	6	0	0	–	+
ALB-Inf-04-21-21-A	Albany, Linn Co.	2021 Apr 21	5	3	1	2	3	1	–	–

*Sequences from individual clinical specimens are identified by their GISAID accession numbers (https://www.gisaid.org). Wastewater sequences are identified by their field collection identifier. Co., County; +, mutation detected; –, mutation not detected; ?, inadequate sequence data for a determination.
†Cities where individual clinical specimens were collected are not provided to reduce identifiability of case-patients.
‡Number of mutations matched by the sequences from each sample. Mutations specific to B.1.351 are G174T, A2692T, G5230T, A21801C, 22283Δ9, G22813T, C25904T, C26456T, and C28253T. Mutations specific to clade 1 are A1763G, C5100T, G13045A, C19524T, 28027Δ129, and C29741T.
§A single mutation defines each of clades 1a (A11875G) and 1b (C15928T). Absence of both mutations defines clade 1c in the case of individual specimens; in the case of wastewater samples, determining whether a mutation is truly absent from the RNA molecules present, or if the mutations have simply not been detected, is not possible.
¶This sample was retrospectively identified as B.1.351 late in April 2021 after routine sequencing of historical specimens.
#This specimen falls into clade 2 (Figure 1).

composite wastewater samples were collected weekly from the influent of Albany and Dallas wastewater treatment plants and from wastewater conveyance lines in Corvallis because of micro-sewershed surveillance at a local university. Samples were vacuum-filtered (10–50 mL) onsite through a 0.45-µm pore size, 47-mm diameter mixed cellulose ester electronegative filter (MF-Millipore, https://www.emdmillipore.com). Filters were placed into a 2-mL tube containing garnet (0.5 mm) beads and DNA/RNA Shield (Zymo Research, https://www.zymoresearch.com) to stabilize the RNA during the shipment to OSU for processing.

Upon receipt of the samples, we subjected the filters to bead beating and extracted RNA by using the MagMAX Viral/Pathogen Nucleic Acid Isolation Kit (Thermo Fisher Scientific, https://www.thermofisher.com). We quantified 2 SARS-CoV-2 gene targets (nucleocapsid gene 1 and 2) and a human gene target (ribonuclease P, an internal control) through droplet digital reverse transcription PCR (ddRT-PCR) on a QX200 ddPCR system (Bio-Rad, https://www.bio-rad.com) using the 2019-nCoV CDC ddPCR Triplex Probe Assay (Bio-Rad) and the One-Step RT-ddPCR Advanced Kit for Probes (Bio-Rad), according to the manufacturer's protocols.

Amplicon-Based Sequencing

We performed amplicon-based sequencing to enable high coverage for the length of the genome, except 25 bp at each end. We synthesized cDNA by using the SuperScript IV First-Strand Synthesis System (Thermo Fisher) and sequenced it by using the Swift Amplicon SARS-CoV-2 Panel (Swift Biosciences, https://www.idtdna.com), together with Swift Amplicon Combinatorial Dual indexed adapters (Swift Biosciences), according to the manufacturer's protocols. The Swift Amplicon SARS-CoV-2 Panel spans the SARS-CoV-2 genome with 341 amplicons with an average length of 150 bp. We produced sequences on a HiSeq 3000 or NextSeq 2000 sequencer (Illumina, https://www.illumina.com) to a depth sufficient for confident identification of variants, typically 10–30 million sequence reads per wastewater sample (Appendix, https://wwwnc.cdc.gov/EID/article/28/6/21-1821-App1.pdf).

Bioinformatic Processing of Sequences

We demultiplexed the sequence reads with zero index mismatches by using bcl2fastq2 version 2.20 (Illumina) for samples sequenced on the HiSeq 3000 and BCL Convert version 1.2.1 (Illumina) for the NextSeq 2000, then trimmed them by using BBDuk (BBMap version 38.84 (US Department of Energy Joint Genome Institute, https://jgi.doe.gov). We aligned trimmed reads to the reference sequence (Wuhan-Hu-1, GenBank accession no. NC_045512.2) by using the BWA-MEM algorithm version 0.7.17-r1188 (https://github.com/lh3/bwa) and coordinate-sorted them with SAMtools version 1.10 (Genome Research Limited, https://www.sanger.ac.uk); we then removed primer sequences by using Primerclip version 0.3.8 (Swift BioSciences). We converted reads from sam files to bam files and coordinate-sorted and indexed them using SAMtools. We then used GATK version 4.2.0.0 (Broad Institute, https://www.broadinstitute.org) to identify mutations compared with the reference sequence (Appendix). We used Integrated Genomics Viewer version 2.8.7 (Broad Institute) to manually inspect sequence alignments and mutation calls (*21*,*22*).

Multilocus Sequence Typing

Through the well-established process of MLST (*12*–*18*), we matched sets of mutations unique to known SARS-CoV-2 variants to mutations found in the wastewater sequences to infer the presence of variants in the community's wastewater. Because of the heterogenous nature of wastewater and the potential presence of numerous SARS-CoV-2 variants in wastewater, we used a set of mutations identified as specific to B.1.351 in Oregon to screen for this variant (Table; Appendix Figure 1). To create this unique panel, we screened a set of 22 mutations associated with B.1.351 (H. Tegally et al., unpub. data, https://doi.org/10.1101/2020.12.21.20248640) (Appendix) against a database of mutations associated with individual clinical specimens sequenced in Oregon and deposited into GISAID (*20*) and from published reports of novel variants (*23*–*28*) (I. Ferreira, unpub. data, https://doi.org/10.1101/2021.05.08.443253; X. Deng et al., unpub. data, https://doi.org/10.1101/2021.03.07.212526 47; M.K. Annavajhala et al., unpub. data, https://doi.org/10.1101/2021.02.23.21252259). Of the 22 mutations associated with the B.1.351 variant, 9 were found only in identified B.1.351 sequences from Oregon, 7 were common to >20 variants, and the remaining 6 were shared by 1–3 other variants (Appendix Tables 2–8, Figure 1).

Because wastewater SARS-CoV-2 RNA is derived from a mixture of variants, sequence reads spanning a B.1.351 mutation site would be expected to include reads derived from B.1.351 RNA molecules as well as reads derived from other variants. For a potential positive identification of a B.1.351-associated mutation in wastewater RNA sequences, a lower limit of 5% of sequence reads (with a minimum of 6 total

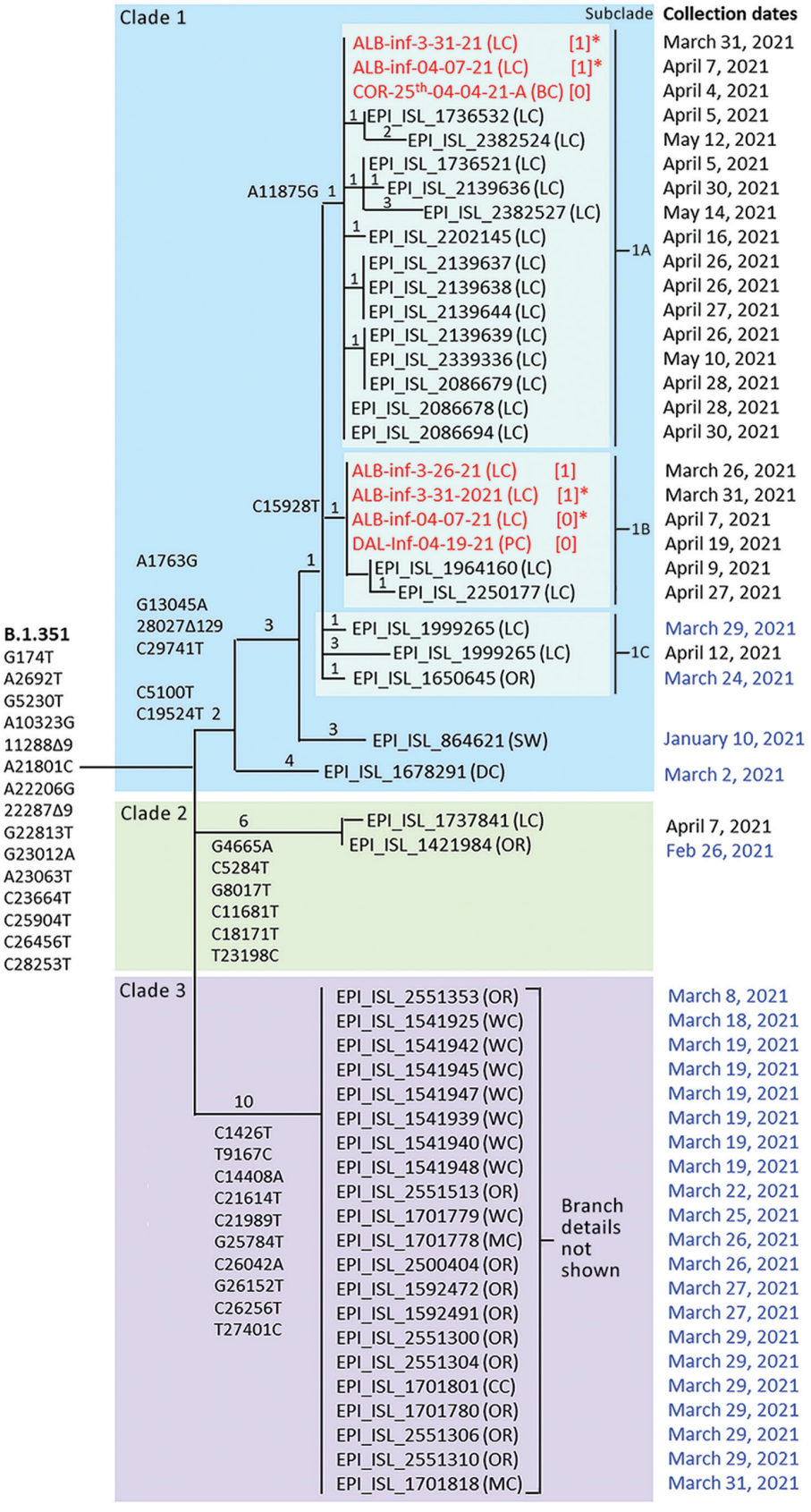


Figure 1. Maximum-parsimony tree demonstrating phylogenetic relationships among SARS-CoV-2 variant B.1.351 clinical specimens and wastewater samples in Linn County, Oregon, USA, and surrounding jurisdictions, March–May 2021. GISAID accession numbers (https://www.gisaid.org) are shown for 19 of 20 B.1.351 specimens identified in Linn County through May 15, 2021, and for 24 additional B.1.351 specimens identified in Oregon through March 31, 2021 (dates in blue). Also included are 2 sequences from outside Oregon (Switzerland and Washington, DC, USA) most closely related to clade 1. Wastewater samples are in red. Exact parsimony trees are shown for clade 1 and 2 sequences, whereas clade 3 sequences are simply listed. Mutations defining B.1.351 and each of the 3 clades, plus subclades 1a and 1b, are shown. Private mutations defining the subbranches of clades 1 and 2 are listed in Appendix Table 9 (https://wwwnc.cdc.gov/EID/article/28/6/21-1821-App1.pdf). Numbers on tree branches indicate the numbers of mutations associated with each branch. Numbers in brackets indicate clade 1 consensus mutations not detected, probably because of poor read coverage. Asterisks indicate samples that appear in both subclades 1a and 1b and are inferred to be a mixture of at ≥2 B.1.351 subtypes. Wastewater sequences ALB-Inf-04-21-21-A and COR-26th-04-04-21-A are not shown because several tracts of those sequences were too uncertain to enable accurate placement on the tree. OR, Oregon; BC, Benton County; CC, Clackamas County; LC, Linn County; MC, Multnomah County; WC, Washington County; DC, Washington, DC; SW, Switzerland.

reads) spanning the mutation site was required. In addition, ≥2 different sites carrying B.1.351 mutations must have been present. The rate of sequence errors produced by the sequencing procedure (a possible source of false-positive reads) was <0.2%.

Ultimately, we used 2 criteria to identify wastewater sample matches to B.1.351: the number of unique mutations present and the normalized proportion of all sequence reads. A minimum of 5 of 9 unique mutations was required for positive identification and we assigned a confidence score as follows: 8–9 matches indicated confident detection, 6–7 matches indicated probable detection, and 5 matches indicated tentative detection. In addition, the normalized proportion of all reads carrying any of the 9 mutations was required to be ≥10% of all reads spanning the 9 mutation sites (Appendix). We reconstructed the putative genome sequences of all variants inferred to be present in wastewater by mapping detected variant-specific mutations onto the reference sequence. We visualized and compared clinical specimen genomes from GISAID and putative viral isolate genomes by using Nextclade version 1.5.2 and used the output to manually construct the maximum-parsimony tree (*29*) (Figure 1). We submitted all wastewater sequences to the National Center for Biotechnology Information Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) (*30*).

Phylogenetic Analysis

We constructed phylogenetic trees by using a maximum-parsimony approach (*29*). Because of the simple structure of the Linn County B.1.351 population, we were able to manually construct a single unambiguous tree by using the mutations specific for each clade and subclade together with the mutations private to each clinical sequence (Appendix Table 9). Mutations private to B.1.351 sequences in wastewater could not be reliably identified because of the presence of other variants, so we did not include them in phylogenetic analysis.

The manual approach enabled us to include sequences that had mutation information missing because of poor sequencing quality. We first constructed the tree by using sequences missing no mutations. Then, we added sequences with missing data to the tree on the basis of available mutation data; this approach imputed the presence of the undetected mutations on the basis of the presence of the available mutations. We did not include in the trees sequences that could not be unambiguously placed on the tree due to missing data.

Results

On March 26 and March 31, 2021, routine wastewater surveillance from the city of Albany, Oregon (Linn County), identified 2 samples that contained SARS-CoV-2 RNA exhibiting mutations specific to the B.1.351 lineage in Oregon (Table; Appendix Figure 1). The wastewater sample from March 26 (ALB-Inf-03–26–21-A) exhibited 7 of 9 specific mutations, whereas the wastewater sample from March 31 (ALB-Inf-03–31–21-A) exhibited all 9. At the time of these initial wastewater detections, no cases of B.1.351 in Linn County or adjacent counties had been identified, and only 25 specimens had been identified as B.1.351 through individual-level whole-genome sequencing statewide. On April 23, 2021, the first case of B.1.351 in Linn County was reported to the local public health authority (specimen collection date April 7). During the following month, 15 additional cases were reported. In Linn County, 20 total cases were identified through sequencing of individual clinical specimens collected through May 15 (Figure 2).

Additional community-level evidence in support of the initial detection of B.1.351 in the wastewater of Albany came from wastewater surveillance in 2 nearby cities as well as subsequent wastewater specimens from Albany. On April 4 and April 19, 2021, routine wastewater surveillance and sequencing of samples from the cities of Corvallis (Benton County; samples COR-25th-04-04-21-A and COR-26th-04-04-21-A) and Dallas (Polk County; sample DAL-Inf-04-19-21-A) identified probable (6/9) to confident (9/9) matches to the unique set of B.1.351 mutations referenced previously (Table; Appendix Figure 1), consistent with local circulation of the B.1.351 variant. Subsequent wastewater surveillance in Albany on April 7 (ALB-Inf-04-07-21-A) and April 21 (ALB-Inf-04-21-21-A) contained confident (8/9) and tentative (5/9) matches to the set of nine unique mutations. In sum, 7 wastewater samples matched ≥5 of the 9 mutations unique to the B.1.351 lineage (Table; Appendix Figure 1). In some cases, the lack of confident matches resulted from poor sequence coverage (<6 reads), whereas in other cases, no match was detected despite moderate sequence coverage (Appendix Figure 1).

Individual-level sequencing results were available for 19 of the 20 B.1.351 specimens identified in Linn County through May 15, 2021, in GISAID. Phylogenetic analysis of all 25 Oregon B.1.351 sequences available in GISAID through March 31, 2021, revealed 3 distinct B.1.351 clades in Oregon (Figure 1). Of the 19 Linn County specimens, 18 resided within a single clade (clade 1) defined by 6 unique mutations (A1763G, C5100T, G13045A, C19524T,

28027Δ129, and C29741T), whereas 1 resided within a second clade defined by a distinct set of 6 mutations (clade 2). Two additional mutations divided clade 1 into 3 subclades: subclade 1a (14 sequences, defined by A11875G), subclade 1b (2 sequences, defined by C15928T), and subclade 1c (2 sequences, defined by neither mutation).

To assess the reliability with which B.1.351 was inferred to be present in the wastewater samples, and to genetically relate the wastewater samples to the individual clinical specimens, we searched the wastewater sequences from Albany and the nearby cities of Corvallis and Dallas for matches to the additional mutations identified in the individual specimens. For the 6 mutations defining clade 1, 6 of the 7 wastewater sequences matched ≥3 mutations and 3 matched all 6 mutation sites (Table; Appendix Figure 1). Three wastewater samples matched clade 1a (defined by mutation A11875G), and 4 samples matched clade 1b (defined by mutation C15928T). Two wastewater samples from Albany (ALB-Inf-03-30-21-A and ALB-Inf-04-07-21-A) matched both mutations, suggesting that those samples contained a mixture of SARS-CoV-2 RNA from both subclades. None of the wastewater samples exhibited mutations characteristic of clade 2, which included only 1 individual clinical specimen. The matches of the wastewater sequences to the additional mutations specific to clade 1 found

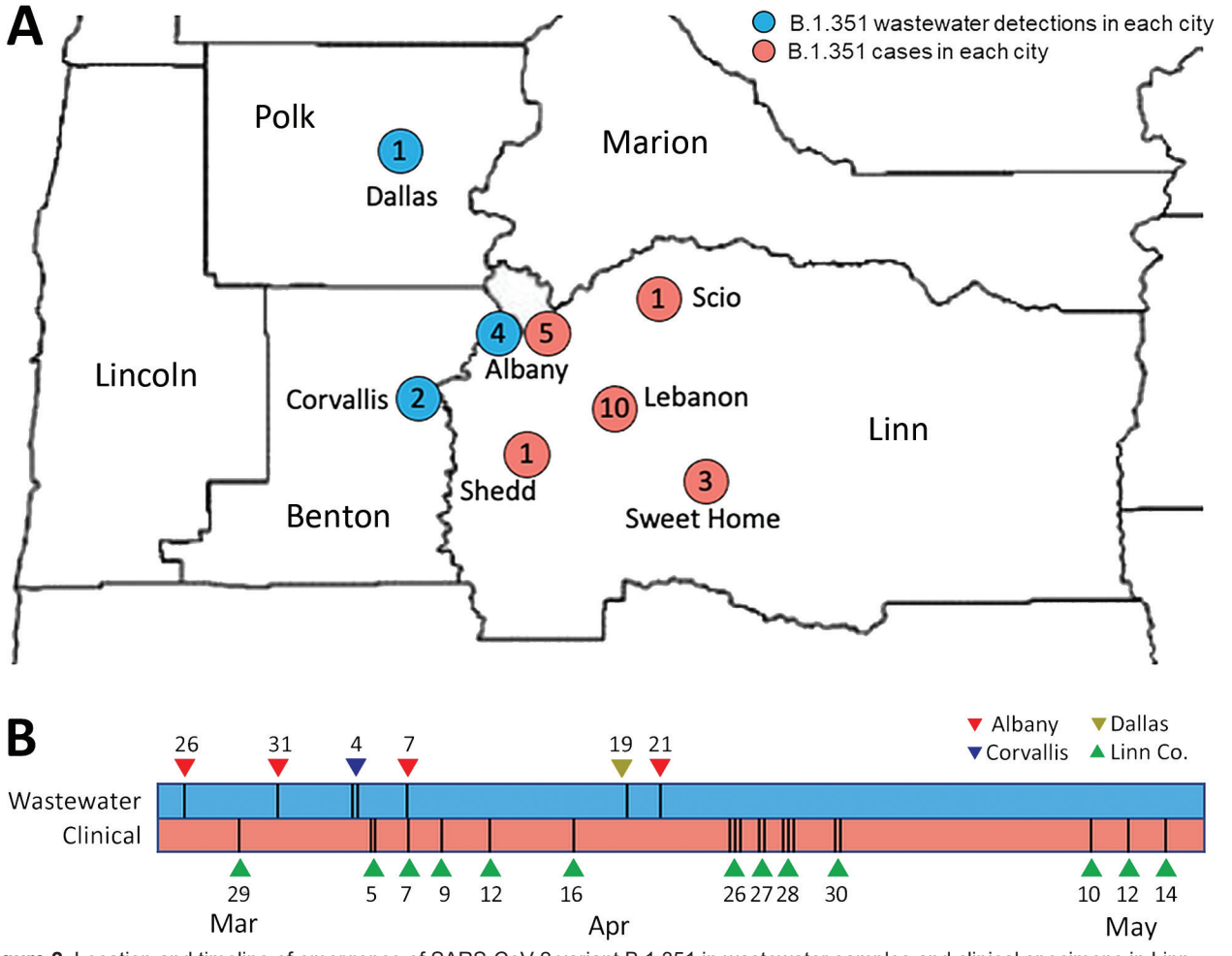


Figure 2. Location and timeline of emergence of SARS-CoV-2 variant B.1.351 in wastewater samples and clinical specimens in Linn County, Oregon, USA, and surrounding jurisdictions, March–May 2021. A) Blue dots represent the sites and numbers of wastewater samples with detections of the B.1.351 variant in Linn County and surrounding jurisdictions. Red dots represent the location and number of individual cases of B.1.351 in Linn County. Initial wastewater samples with evidence of the B.1.351 variant of concern were collected from Albany, Oregon, during March 26–31, 2021, and the first case of B.1.351 infection in Linn County was reported on April 23, 2021 (specimen collection date of April 7, 2021); 18 additional cases were identified through May 15, 2021, including cases with earlier specimen collection dates. B) Timeline of wastewater samples and clinical specimens positive for B.1.351 in Linn County and surrounding jurisdictions. Vertical bars indicate the number of samples or specimens collected on each date. City locations are not given to limit identifiability of individual case-patients.

in most Linn County cases substantially increase the confidence that B.1.351 RNA sequences were correctly identified in the wastewater samples.

We further assessed the sequences of each of the 7 wastewater samples for the presence of 6 additional mutations characteristic of B.1.351 but shared by other variants, including B.1.1.7, B.1.526, and P.1 (Appendix Table 8). To determine if interfering variants were present, we screened the sequences of each wastewater sample for mutations unique to those variants. We did not detect interfering variants in either of the 2 initial wastewater samples (ALB-Inf-03-26-21-A and ALB-03-30-21-A), and both samples exhibited all 6 of the additional shared mutations (Appendix Figure 1). These results provide further evidence for the true identification of B.1.351 in the initial 2 wastewater samples collected in Albany. The remaining 5 wastewater samples from Albany (ALB-Inf-04-07-21-A and ALB-Inf-04-21-21-A), Corvallis (COR-25th-04-04-21-A and COR-26th-04-04-21-A), and Dallas (DAL-Inf-04-19-21-A) demonstrated mutations consistent with B.1.1.7 and, in 1 case, P.1 (Appendix Tables 2–7, Figure 1).

Discussion

In late March 2021, routine sequencing of SARS-CoV-2 wastewater surveillance detected the emergence of the B.1.351 (Beta) variant of concern in rural Linn County, Oregon, before its identification in individual cases. Currently, wastewater surveillance is used to track SARS-CoV-2 transmission trends in several jurisdictions and, in times of minimal transmission, may serve as an early warning system for disease resurgence (*31*). Wastewater surveillance offers local public health authorities and communities actionable data that is independent of symptomatic infection, healthcare access, and testing acceptance and may help in developing vaccination strategy (*32*). Leveraging this surveillance to support genomic surveillance for SARS-CoV-2 offers cost-effective community-level surveillance that may detect not only prevalent circulating variants but emerging variants of concern as well.

Accurate interpretation of wastewater sequencing results faces several challenges. These challenges include the heterogeneous nature of wastewater samples, the fragmentation of viral RNA in wastewater, the need to match wastewater sequences to panels of mutations characteristic of known variants, the variable levels of variant RNA in wastewater samples, the uneven sequence coverage of the viral genome in wastewater sequences, and the sharing of mutations (e.g., N501Y and E484K) across multiple variants. We used the well-established approach of MLST (*13*,*18*) in a novel application to infer the presence of RNA from SARS-CoV-2 variants in community wastewater samples from a statewide wastewater surveillance program.

MLST has been used to detect other pathogens in complex environmental samples, including wastewater. In addition, MLST has been used to analyze fragmented genetic molecules through the rigorous identification of matches to a curated set of mutations (i.e., a mutation panel) (*12*,*13*). Confidence in a detection is based on the proportion of matches to the mutation panel. Amplicon-based sequencing with the Swift Amplicon SARS-CoV-2 Panel is well-suited to MLST, providing excellent coverage of the entire SARS-CoV-2 genome, omitting only 25 bp at each end. With 341 overlapping amplicons of 150 bp on average, this method is robust to most mutations that could disrupt the binding of a primer (i.e., cause primer dropout) (N.L. Washington et al., unpub. data, https://doi.org/10.1101/2020.12.24.20248814).

To establish mutation panels suitable for screening for individual variants, we began with the canonical mutations defining each variant, derived either from the literature (*24*–*27*,*29*,*32*; I. Ferreira, unpub. data; X. Deng et al., unpub. data; M.K. Annavajhala et al., unpub. data) or from the Centers for Disease Control and Prevention (*33*). These sets of mutations were validated through the creation of an expanded panel of mutations identified in statewide individual sequencing data submitted to GISAID. Finally, all mutations shared with known variants were filtered out. This validation step would not be available to an emerging variant for which no local or regional individual-level sequencing data were available, highlighting the complementary properties of individual-level and community-level surveillance.

To address variable levels of variant RNA in the wastewater samples, we conducted in-depth sequencing, producing ≈10–30 million sequence reads per sample, to obtain sufficient sequence data to detect variants comprising as little as 10% of the RNA, even from samples with the lowest levels of viral RNA ($\log_{10}$ gene copies/L of 4.0). To address uneven sequence coverage of the viral genome in wastewater sequences, ranging from <10 to >1,000 reads per amplicon within a single sample, the sequence coverage was normalized to 100 reads per site for mutation sites with coverage of >100 reads. For sites with coverage of <100 reads per site, the actual read numbers were used to assign a proportionately smaller weight to those more poorly sequenced mutation sites.

In this study, we used a panel of 9 mutations identified as specific to B.1.351 to screen for the presence

of this emerging variant of concern. We then used a subsequent panel of 8 additional mutations defining a single clade of B.1.351 sequences identified through statewide individual specimen sequencing to validate the initial set of matches, together with a case-by-case examination of a set of 6 mutations characteristic of B.1.351 but shared with other variants. This 2-step process of screening followed by validation, together with the large number of mutations within the screening and validation panels, rendered the detection of B.1.351 robust to small numbers of mismatches that occurred because of low sequence coverage, low levels of variant RNA, or primer dropout. The ability to compare independent but geographically or temporally related wastewater samples with closely related individual sequences substantially increased confidence in our detection of B.1.351 through wastewater surveillance (Figure 2).

All 7 wastewater sequences and 18 of 19 B.1.351 individual clinical specimen sequences clustered into a single clade (clade 1). The similarity of the sequences and their spatiotemporal proximity suggests a single common origin of the detected viruses. The SARS-CoV-2 sequences most closely related to the sequences in clade 1 were found in Switzerland (Figure 1), suggesting that the Oregon clade 1 cluster in Linn County may have originated from outside the United States. Even though B.1.351 was detected in the wastewater of the nearby cities of Corvallis (Benton County; OSU-25th-04-04-21-A) and Dallas (Polk County; DAL-Inf-04-19-21-A), no cases were identified in Benton or Polk Counties during this period. Thus, the detection of B.1.351 in these 2 counties may have resulted from the transient presence of cases from neighboring counties or may simply reflect insufficient individual-level genomic surveillance to detect B.1.351 in those areas.

Together, the complementary wastewater and clinical data we present clearly support community transmission of the B.1.351 variant in the Linn County region from late March through mid-May 2021. Wastewater sampling detected this emerging variant of concern 12 days before the specimen collection date of the first local case-patient. Wastewater surveillance therefore may be an efficient and reliable means of community-level monitoring for emerging SARS-CoV-2 variants and other human pathogens. Additional studies such as ours must be replicated across rural and urban settings to further understanding of the generalizability and limitations of wastewater surveillance. Scientific consensus regarding methods and minimum thresholds for variant detection in wastewater are urgently needed.

Acknowledgments

This project was conducted in close partnership with the Oregon State University TRACE project (https://trace.oregonstate.edu) and the City of Albany Oregon Public Works Department employees. The authors thank the members of the TRACE leadership team (Benjamin Dalziel, Jeff Bethel, Roy Haggerty, Javier Nieto, Kathryn Higley, Katherine McLaughlin, and Jane Lubchenco) for their support and advice.

Funding for this study came from the National Science Foundation (to T.R. and C.K., award no. 2027679), the David and Lucile Packard Foundation (to the TRACE project), Oregon State University (to the TRACE project, T.R., C.K., and B.M.T.), and the Centers for Disease Control and Prevention (cooperative agreement nos. CK17-1701 and CK-19-1904).

About the Author

Dr. Sutton is medical director of respiratory viral pathogens at Oregon Health Authority. Her research interests include respiratory viral pathogen surveillance, infectious disease epidemiology, and the social determinants of health.

References

1. World Health Organization. WHO coronavirus (COVID-19) dashboard [cited 2022 Mar 29] https://covid19.who.int
2. Centers for Disease Control and Prevention. COVID data tracker weekly review [cited 2022 Mar 29]. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html
3. Walensky RP, Walke HT, Fauci AS. SARS-CoV-2 variants of concern in the United States—challenges and opportunities. JAMA. 2021;325:1037–8. https://doi.org/10.1001/jama.2021.2294
4. Lo SW, Jamrozy D. Genomics and epidemiological surveillance. Nat Rev Microbiol. 2020;18:478. https://doi.org/10.1038/s41579-020-0421-0
5. Robishaw JD, Alter SM, Solano JJ, Shih RD, DeMets DL, Maki DG, et al. Genomic surveillance to combat COVID-19: challenges and opportunities. Lancet Microbe. 2021;2:e481–4. https://doi.org/10.1016/S2666-5247(21)00121-X
6. Rader B, Astley CM, Sy KTL, Sewalk K, Hswen Y, Brownstein JS, et al. Geographic access to United States SARS-CoV-2 testing sites highlights healthcare disparities and may bias transmission estimates. J Travel Med. 2020;27:taaa076.
7. Contreras S, Dehning J, Loidolt M, Zierenberg J, Spitzner FP, Urrea-Quintero JH, et al. The challenges of containing SARS-CoV-2 via test-trace-and-isolate. Nat Commun. 2021;12:378. https://doi.org/10.1038/s41467-020-20699-8
8. Thompson JR, Nancharaiah YV, Gu X, Lee WL, Rajal VB, Haines MB, et al. Making waves: wastewater surveillance of SARS-CoV-2 for population-based health management. Water Res. 2020;184:116181. https://doi.org/10.1016/j.watres.2020.116181
9. Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-Coronavirus-2 RNA in sewage and

correlation with reported COVID-19 prevalence in the early stage of the epidemic in the Netherlands. Environ Sci Technol Lett. 2020;7:511–6. https://doi.org/10.1021/acs.estlett.0c00357
10. Daughton CG. Wastewater surveillance for population-wide Covid-19: The present and future. Sci Total Environ. 2020; 736:139631. https://doi.org/10.1016/j.scitotenv.2020.139631
11. Izquierdo-Lara R, Elsinga G, Heijnen L, Munnink BBO, Schapendonk CME, Nieuwenhuijse D, et al. Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing, the Netherlands and Belgium. Emerg Infect Dis. 2021;27:1405–15. https://doi.org/10.3201/eid2705.204410
12. Durigan M, Abreu AG, Zucchi MI, Franco RMB, de Souza AP. Genetic diversity of *Giardia duodenalis*: multilocus genotyping reveals zoonotic potential between clinical and environmental sources in a metropolitan region of Brazil. PLoS One. 2014;9:e115489. https://doi.org/10.1371/journal.pone.0115489
13. Ma J, Feng Y, Hu Y, Villegas EN, Xiao L. Human infective potential of *Cryptosporidium* spp., *Giardia duodenalis* and *Enterocytozoon bieneusi* in urban wastewater treatment plant effluents. J Water Health. 2016;14:411–23. https://doi.org/10.2166/wh.2016.192
14. Ibarz Pavón AB, Maiden MCJ. Multilocus sequence typing. Methods Mol Biol. 2009;551:129–40. https://doi.org/10.1007/978-1-60327-999-4_11
15. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A. 1998;95:3140–5. https://doi.org/10.1073/pnas.95.6.3140
16. Urwin R, Maiden MC. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol. 2003;11:479–87. https://doi.org/10.1016/j.tim.2003.08.006
17. Wang Z-G, Zheng Z-H, Shang L, Li L-J, Cong LM, Feng MG, et al. Molecular evolution and multilocus sequence typing of 145 strains of SARS-CoV. FEBS Lett. 2005;579:4928–36. https://doi.org/10.1016/j.febslet.2005.07.075
18. Charpentier E, Garnaud C, Wintenberger C, Bailly S, Murat J-B, Rendu J, et al. Added value of next-generation sequencing for multilocus sequence typing analysis of a *Pneumocystis jirovecii* pneumonia outbreak. Emerg Infect Dis. 2017;23:1237–45. https://doi.org/10.3201/eid2308.161295
19. Centers for Disease Control and Prevention. COVID data tracker. 2021 [cited 2021 Jul 18]. https://covid.cdc.gov/covid-data-tracker/#datatracker-home.
20. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 2017;1:33–46. https://doi.org/10.1002/gch2.1018
21. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6. https://doi.org/10.1038/nbt.1754
22. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the Integrative Genomics Viewer. Cancer Res. 2017;77:e31–4. https://doi.org/10.1158/0008-5472.CAN-17-0337
23. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. 2020 [cited 2021 Aug 12]. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563
24. Fujino T, Nomoto H, Kutsuna S, Ujiie M, Suzuki T, Sato R, et al. Novel SARS-CoV-2 variant in travelers from Brazil to Japan. Emerg Infect Dis. 2021;27:1243–5. https://doi.org/10.3201/eid2704.210138
25. Faria NR, Claro IM, Candido D, Franco LM, Andrade PS, Coletti TM, et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. 2021 [cited 2021 Aug 12]. https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586
26. Naveca F, Nascimento V, Souza V, Corado A, Nascimento F, Silva G, et al. Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein. 2021 [cited 2021 Aug 12]. https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585
27. Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E. Emergence of a novel SARS-CoV-2 variant in Southern California. JAMA. 2021;325:1324–6. https://doi.org/10.1001/jama.2021.1612
28. World Health Organization. Weekly epidemiological update on COVID-19, 27 April 2021 [cited 2021 Aug 12]. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—27-april-2021
29. Kannan L, Wheeler WC. Maximum parsimony on phylogenetic networks. Algorithms Mol Biol. 2012;7:9. https://doi.org/10.1186/1748-7188-7-9
30. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2011;39(suppl_1):D38–51. https://doi.org/10.1093/nar/gkq1172
31. Venugopal A, Ganesan H, Sudalaimuthu Raja SS, Govindasamy V, Arunachalam M, Narayanasamy A, et al. Novel wastewater surveillance strategy for early detection of coronavirus disease 2019 hotspots. Curr Opin Environ Sci Health. 2020;17:8–13. https://doi.org/10.1016/j.coesh.2020.05.003
32. Smith T, Cassell G, Bhatnagar A. Wastewater surveillance can have a second act in COVID-19 vaccine distribution. JAMA Health Forum. 2021;2:e201616. https://doi.org/10.1001/jamahealthforum.2020.1616
33. Centers for Disease Control and Prevention. SARS-CoV-2 variant classifications and definitions. 2021 [cited 2021 Aug 12]. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html

Address for correspondence: Melissa Sutton, Public Health Division, Oregon Health Authority, 800 NE Oregon St, Ste 772, Portland, OR 97232, USA; email: melissa.sutton@dhsoha.state.or.us

Article DOI: https://doi.org/10.3201/eid2806.211821

Detection of SARS-CoV-2 B.1.351 (Beta) Variant through Wastewater Surveillance before Case Detection in a Community, Oregon, USA

Appendix

Methods

Sequencing Depth and Quality Assurance

The percentage of sequence reads that aligned to the SARS-CoV-2 reference genome were based principally on the SARS-CoV-2 RNA concentration in wastewater samples. RNA concentrations of $\log_{10}$ gene copies/liter of 4.0 to 5.0 typically yielded 0.1% to 5.0% alignable reads (Appendix Table 1) while RNA concentrations over 7.0 have yielded over 90% alignable reads. In contrast, RNA concentrations below 4.0 produced poor results with the methodology described here, with many Swift amplicons not represented at all among the sequences (data not shown). Best results with MLST, with most mutation sites receiving over 10 sequence reads, were generally obtained when the number of aligned reads was 20,000 or more. To achieve this depth for most samples, a total of 10 to 30 million raw reads per sample were typically required and this was achieved by adjusting the number of sequencing lanes on the Illumina HiSeq3000 to the number of samples in a sequencing batch. Generally, for up to 40 samples, one lane was used; for 40-70 samples, two lanes were used (Appendix Table 1), and for 70-96 samples, three lanes were used.

Negative controls lacking RNA and positive controls containing synthetic SARS-CoV-2 RNA (Twist Biosciences, San Francisco, CA; Controls 12 and 13, parts #103515 and103533) were sequenced with each batch of wastewater samples. The negative controls were used to check for cross-contamination. The positive controls containing synthetic RNA were utilized to assess sequence quality and sensitivity. To minimize the risk of cross-contamination, RNA from nasal swabs, which typically contained high concentrations of SARS-CoV-2 RNA, were

sequenced on different days and analyzed in different Illumina HiSeq3000 lanes than wastewater samples. Sets of samples with failed negative or positive controls were re-sequenced. All data reported here are from samples that passed these quality control tests.

Identification of Mutations Using GATK Software

The GATK software package (version 4.2.0.0, Broad Institute, Cambridge, MA) was used to identify mutations in wastewater sequence reads compared to the reference sequence using the following procedure: variants were called on a per sample basis using GATK's HaplotypeCaller sub-package with the following settings: -stand-call-conf 20, --dont-use-soft-clipped-bases, -mbq 20, --max-reads-per-alignment-start 0, --linked-de-bruijn-graph, --recover-all-dangling-branches, --sample-ploidy 4. gVCF files were merged using GATK's CombineGVCFs sub-package and jointly called with GATK's GenotypeGVCFs sub-package while forcing the output of the standard loci. The final VCF file was converted to a tabular format with GATK's VariantsToTable sub-package.

Establishment of the B.1.351 Mutation Panel for MLST

The B.1.351 mutation panel was initially based on the 18 substitution mutations appearing in greater than 90% of the 190 B.1.351 sequences originally described (23), namely G174T, C241T, C1059T, A2692T, C3037T, G5230T, A10323G, C14408T, A21801C, G22813T, G23012A, A23063T, A23403G, C23664T, C25904T, C26456T, G25563T, C28253T, and C28887T. Subsequent examination of 46 U.S. B.1.351 sequences collected through December 31, 2020 and deposited in GISAID, suggested that four additional mutations had become fixed in the U.S. B.1.351 population, namely A2692T, 11288Δ9, A22206G, and 22283Δ9. These mutations were, therefore, added to the screening panel.

The following six mutations were then removed from the panel due to their co-occurrence in numerous other B.1 lineages: C241T, C1059T, C3037T, C14408T, A23403G, G25563T, and C28887T. Finally, another set of six mutations, A10323G, 11288Δ9, A22206G, G23012A, A23063T, and C23664T, were removed from the panel because they also occurred in other published lineages (B.1.1.7, B.1.1.316, B.1.526, and P.1) or in lineages encountered in sequences from Oregon individuals (B.1.404, B.1.582) (Appendix Table 8). We refer to such lineages as "interfering variants".

In order to detect the possible occurrence of emerging (i.e., previously undefined) interfering variants with mutations overlapping with the B.1.351 panel, we routinely screened for mutations from the B.1.351 panel that were present in the absence of matches to any other mutations in the B.1.351 panel. Sequences exhibiting such mutations were then clustered to determine if there were additional mutations co-occurring with the mutation in question. If clusters of co-occurring mutations were observed, we inferred the possible presence of a novel interfering variant and removed the mutation from the panel. Although this procedure did not detect novel variants that interfered with the detection of B.1.351, the procedure regularly identified novel putative variants interfering with detection of other variants in our broader screening panel.

Estimating the Fraction of RNA Molecules Attributable to B.1.351 in Samples

One criterion used to assess the reliable detection of B.1.351 in a wastewater sample was an estimated percentage of RNA molecules in the sample attributable to B.1.351 greater than 10%, based on information from all the mutations in the screening panel. This quantitation was challenging because the number of sequence reads spanning the nine unique mutation sites could vary over a 250-fold range, due to differing amplification and sequencing efficiencies of the 341 Swift amplicons (Appendix Tables 2–7). We first calculated the fraction of all reads spanning a mutation site that exhibited the mutation and averaged the ratios over all nine mutation sites, however, this approach proved unsatisfactory as mutation sites with low read coverage exhibited highly variable ratios which biased the estimate. We then summed the number of all reads exhibiting any of the nine mutations and divided by the number of reads spanning all the mutation sites, however, this approach also proved unsatisfactory as the estimate was biased by the most deeply sequenced mutation sites. A hybrid method was, therefore, adopted. For mutation sites with more than 100 sequence reads, the numbers of reads exhibiting the mutation were normalized to a total of 100 reads (Appendix Tables 2–7). For mutation sites with fewer than 100 sequence reads, the actual numbers of reads were retained. The normalized numbers of all reads exhibiting any of the nine mutations were then summed and divided by the sum of the normalized numbers of reads spanning all the mutation sites (Appendix Tables 2–7).

When all sites were spanned by more than 100 reads, the ratio produced by the hybrid method was identical to that produced by the first approach of averaging the ratios at the nine sites. Through the hybrid method, when some sites had less than 100 reads, their contribution to

the final estimate was reduced in proportion to their lower read coverage. The hybrid method of estimation reduced bias from both under-sequenced and over-sequenced sites. All samples identified in this paper as including B.1.351 RNA sequences exhibited estimates over the minimum threshold of 0.1 (Appendix Tables 2–7).

Appendix Table 1. Sequencing statistics for wastewater samples

Sample	SARS-CoV-2 RNA*	Samples/Lanes†	Passing reads‡	Aligned reads§	% Aligned§
ALB-Inf-3-26-2021-A	4.50	69/2 HiSeq	14,545,578	19,576	0.13%
ALB-Inf-3-31-2021-A	4.54	69/2 HiSeq	12,916,288	23,610	0.18%
ALB-Inf-4-7-2021-A	4.53	58/2 HiSeq	17,755,758	82,631	0.46%
ALB-Inf-4-14-2021-A¶	4.77	49/2 HiSeq	26,941,183	1,319,703	4.90%
ALB-Inf-4-21-2021-A	4.64	57/2 HiSeq	12,504,849	47,703	0.38%
DAL-Inf-4-19-2021-A	5.04	57/2 HiSeq	18,047,175	558,391	3.09%
COR-25TH-4-4-21-A	4.96	69/2 HiSeq	21,511,528	281,282	1.31%
COR-26TH-4-4-21-A	4.17	69/2 HiSeq	18,426,836	49,452	0.27%
COR-27TH-4-4-21-A¶	4.57	69/2 HiSeq	15,713,982	45,010	0.29%

* SARS-CoV-2 RNA concentrations in the original wastewater sample, as estimated by digital droplet PCR. Units are $\log_{10}$ gene copies/liter.
† Each sample was included in a batch of the indicated number of samples and sequence was generated from the batch using the indicated numbers of lanes of an Illumina HiSeq3000 instrument.
‡ Numbers of reads passing Illumina QC for each sample.
§ Number and percentage of passing reads aligning to the reference sequence for each sample.
¶ These samples were ultimately classified as no detection of B.1.351.

Appendix Table 2. Summary of sequence read analysis of mutations specific to B.1.351 in wastewater samples in Linn County, Oregon and surrounding jurisdictions, March-April 2021

Sample identifier*	Source	Collection Date	Final call†	Read fraction‡	Mutation count§
ALB-Inf-03-26-21-A	Albany wastewater plant	March 26, 2021	Probable	0.423	7/9
ALB-Inf-03-31-21-A	Albany wastewater plant	March 31, 2021	Confident	0.454	9/9
ALB-Inf-04-07-21-A	Albany wastewater plant	April 7, 2021	Confident	0.245	8/9
ALB-Inf-04-14-21-A	Albany wastewater plant	April 14, 2021	Not detected	0	0/9
ALB-Inf-04-21-21-A	Albany wastewater plant	April 21, 2021	Tentative	0.142	5/9
COR-25TH-04-04-21-A	Corvallis 25th St. sewer	April 4, 2021	Confident	0.901	9/9
COR-26TH-04-04-21-A	Corvallis 26th St. sewer	April 4, 2021	Probable	0.566	6/9
COR-27TH-04-04-21-A	Corvallis 27th St. sewer	April 4, 2021	Trace	0.146	3/9
DAL-Inf-04-19-21-A	Dallas wastewater plant	April 19, 2021	Confident	0.552	9/9

* Wastewater samples analyzed in this study including two additional samples, collected from the Albany wastewater plant on April 14, 2021, and from Corvallis 27th St. sewer on April 4, 2021. B.1.351 was not reliably detected in those additional samples, however, they are included here to illustrate the variant calling procedure.
† The final call as to whether B.1.351 was present was based on the proportion of mutant reads and the number of mutations detected that were specific to B.1.351 (out of 9). If the overall proportion of mutant reads was less than 0.1, or the number of individual mutations detected was less than 3/9, then the variant was called as "Not detected". If the overall proportion of mutant reads was ≥ 0.1, the variant was called by the number of individual mutations present: 3-4 as “Trace”, 5 as “Tentative”, 6-7 as “Probable” and 8-9 as “Confident”.
‡ The overall proportion of mutant reads supporting the presence of the variant was calculated as the sum of the normalized read numbers for the 9 B.1.351-specific mutation sites divided by the sum of all the normalized read numbers for the 9 sites.
§ The count was the number of B.1.351-specific mutations meeting detection criteria.

Appendix Table 3. Detailed sequence read analysis of mutations specific to B.1.351 in wastewater samples in Linn County, Oregon and surrounding jurisdictions, March-April 2021

		Mutations‡								
Wastewater samples*	Analysis†	G174T	A2692T	G5230T	A21801C	22283Δ9	G22813T	C25904T	C26456T	C28253T
ALB-Inf-03-26-21-A	All_reads-actual	784	37	86	27	66	1	8	20	27
	Unsupport_reads-actual	516	4	57	9	42	1	8	4	24
	Support_reads-actual	268	33	29	18	24	0	0	16	3
	All_reads-normalized	100	37	86	27	66	1	8	20	27
	Unsupport_reads-normalized	65.8	4.0	57	9	42	1	8	4	24
	Support_reads-normalized	34.2	33.0	29	18	24	0	0	16	3
	Proportion	0.342	0.892	0.337	0.667	0.364	0	0	0.800	0.111
	Mutation call	Present	Present	Present	Present	Present	Insuff	Absent	Present	Present
ALB-Inf-03-31-21-A	All_reads-actual	617	38	189	115	62	16	27	44	116
	Unsupport_reads-actual	303	12	138	34	42	6	5	14	104
	Support_reads-actual	314	26	51	81	20	10	22	30	12
	All_reads-normalized	100	38	100	100	62	16	27	44	100
	Unsupport_reads-normalized	49.1	12.0	73.0	29.6	42	6	5	14	89.7
	Support_reads-normalized	50.9	26.0	27.0	70.4	20	10	22	30	10.3
	Proportion	0.509	0.684	0.270	0.704	0.323	0.625	0.815	0.682	0.103
	Mutation call	Present	Present	Present	Present	Present	Present	Present	Present	Present
ALB-Inf-04-07-21-A	All_reads-actual	4472	247	562	263	162	45	40	109	294
	Unsupport_reads-actual	3370	148	431	196	113	27	38	106	219
	Support_reads-actual	1102	99	131	67	49	18	2	3	75
	All_reads-normalized	100	100	100	100	100	45	40	100	100
	Unsupport_reads-normalized	75.4	59.9	76.7	74.5	69.8	27	38	97.2	74.5
	Support_reads-normalized	24.6	40.1	23.3	25.5	30.2	18	2	2.8	25.5
	Proportion	0.246	0.401	0.233	0.255	0.302	0.400	0.050	0.028	0.255
	Mutation call	Present	Present	Present	Present	Present	Present	Present	Absent	Present
ALB-Inf-04-14-21-A	All_reads-actual	23135	2732	1788	2682	2807	945	1699	241	649
	Unsupport_reads-actual	23135	2732	1788	2682	2807	945	1699	241	649
	Support_reads-actual	0	0	0	0	0	0	0	0	0
	All_reads-normalized	100	100	100	100	100	100	100	100	100
	Unsupport_reads-normalized	100	100	100	100	100	100	100	100	100
	Support_reads-normalized	0	0	0	0	0	0	0	0	0
	Proportion	0	0	0	0	0	0	0	0	0
	Mutation call	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
ALB-Inf-04-21-21-A	All_reads-actual	1208	197	51	120	241	2	11	42	90
	Unsupport_reads-actual	1123	186	51	101	217	2	11	42	44
	Support_reads-actual	85	11	0	19	24	0	0	0	46
	All_reads-normalized	100	100	51	100	100	2	11	42	90
	Unsupport_reads-normalized	93.0	94.4	51	84.2	90.0	2	11	42	44
	Support_reads-normalized	7.0	5.6	0	15.8	10.0	0	0	0	46
	Proportion	0.070	0.056	0	0.158	0.100	0	0	0	0.511
	Mutation call	Present	Present	Absent	Present	Present	Insuff	Absent	Absent	Present
COR-25TH-04-04-21-A	All_reads-actual	18138	829	1924	1046	737	146	204	339	493
	Unsupport_reads-actual	1585	70	151	10	145	12	38	16	59
	Support_reads-actual	16553	759	1773	1036	592	134	166	323	434
	All_reads-normalized	100	100	100	100	100	100	100	100	100
	Unsupport_reads-normalized	8.7	8.4	7.8	1.0	19.7	8.2	18.6	4.7	12.0
	Support_reads-normalized	91.3	91.6	92.2	99.0	80.3	91.8	81.4	95.3	88.0
	Proportion	0.913	0.916	0.922	0.990	0.803	0.918	0.814	0.953	0.880

		Mutations‡								
Wastewater samples*	Analysis†	G174T	A2692T	G5230T	A21801C	22283Δ9	G22813T	C25904T	C26456T	C28253T
	Mutation call	Present	Present	Present	Present	Present	Present	Present	Present	Present
COR-26TH-04-04-21-A	All_reads-actual	707	40	288	186	56	0	57	20	191
	Unsupport_reads-actual	63	9	29	186	56	0	36	16	24
	Support_reads-actual	644	31	259	0	0	0	21	4	167
	All_reads-normalized	100	40	100	100	56	0	57	20	100
	Unsupport_reads-normalized	8.9	9.0	10.1	100	56	0	36	16	12.6
	Support_reads-normalized	91.1	31.0	89.9	0	0	0	21	4	87.4
	Proportion	0.911	0.775	0.899	0	0	0	0.368	0.200	0.874
	Mutation call	Present	Present	Present	Absent	Absent	Insuff	Present	Present	Present
COR-27TH-04-04-21-A	All_reads-actual	133	12	52	66	57	4	6	0	58
	Unsupport_reads-actual	133	7	52	49	30	1	6	0	58
	Support_reads-actual	0	5	0	17	27	3	0	0	0
	All_reads-normalized	100	12	52	66	57	4	6	0	58
	Unsupport_reads-normalized	100	7	52	49	30	1	6	0	58
	Support_reads-normalized	0	5	0	17	27	3	0	0	0
	Proportion	0	0.4166667	0	0.258	0.474	0.750	0	0	0
	Mutation call	Absent	Present	Absent	Present	Present	Insuff	Absent	Insuff	Absent
DAL-Inf-04-19-21-A	All_reads-actual	22203	1650	2192	2738	1761	402	1318	924	2370
	Unsupport_reads-actual	11120	768	924	1729	461	174	742	277	1084
	Support_reads-actual	11083	882	1268	1009	1300	228	576	647	1286
	All_reads-normalized	100	100	100	100	100	100	100	100	100
	Unsupport_reads-normalized	50.1	46.5	42.2	63.1	26.2	43.3	56.3	30.0	45.7
	Support_reads-normalized	49.9	53.5	57.8	36.9	73.8	56.7	43.7	70.0	54.3
	Proportion	0.499	0.535	0.578	0.369	0.738	0.567	0.437	0.700	0.543
	Mutation call	Present	Present	Present	Present	Present	Present	Present	Present	Present

* See Appendix Table 2

† Actual numbers of sequence reads spanning the mutation site (All_reads), the numbers of reads not supporting the presence of the mutation (Unsupport, the reference sequence), and the number of reads supporting the presence of the mutation (Support, mutant sequence) are shown. In addition, the normalized numbers of reads are shown. For normalization, read numbers greater than 100 were reduced proportionately to a total of 100 reads. For read numbers of 100 or less, the actual read number was retained. "Proportion" is the proportion of all reads that carry the mutation. Mutations were called present if the proportion was 0.05 or greater and the number of reads carrying the mutation was 2 or more. Otherwise, the mutation was called Absent. If the total number of reads was 5 or less, the call was "insufficient data" (Insuff).

‡ Mutations found only in B.1.351. See Figure 1 and Appendix Figure 1.

Appendix Table 4. Detailed sequence read analysis of mutations specific to B.1.351 clade 1 in wastewater samples in Linn County, Oregon and surrounding jurisdictions, March-April 2021

		Mutations‡							
Wastewater samples*	Analysis†	A1763G	C5100T	G13045A	C19524T	28026Δ129	C29741T	A11875G§	C15928T§
ALB-Inf-03-26-21-A	All_reads-actual	2	46	30	46	44	10	48	134
	Unsupport_reads-actual	2	30	27	43	44	10	48	86
	Support_reads-actual	0	16	3	3	0	0	0	48
	All_reads-normalized	2	46	30	46	44	10	48	100
	Unsupport_reads-normalized	2	30	27	43	44	10	48	64.2
	Support_reads-normalized	0	16	3	3	0	0	0	35.8
	Proportion	0	0.348	0.100	0.065	0	0	0	0.358
	Mutation call	Insuff	Present	Present	Present	Absent	Absent	Absent	Present
ALB-Inf-03-31-21-A	All_reads-actual	2	13	15	48	48	77	41	216
	Unsupport_reads-actual	2	8	4	15	48	15	14	187
	Support_reads-actual	0	5	11	33	0	62	27	29
	All_reads-normalized	2	13	15	48	48	77	41	100
	Unsupport_reads-normalized	2	8	4	15	48	15	14	86.6
	Support_reads-normalized	0	5	11	33	0	62	27	13.4
	Proportion	0	0.385	0.733	0.688	0	0.805	0.659	0.134
	Mutation call	Insuff	Present	Present	Present	Absent	Present	Present	Present
ALB-Inf-04-07-21-A	All_reads-actual	23	72	169	127	463	161	239	511
	Unsupport_reads-actual	17	67	118	111	463	147	211	450
	Support_reads-actual	6	5	51	16	0	14	28	61
	All_reads-normalized	23	72	100	100	100	100	100	100
	Unsupport_reads-normalized	17	67	69.8	87.4	100	91.3	88.3	88.1
	Support_reads-normalized	6	5	30.2	12.6	0	8.7	11.7	11.9
	Proportion	0.261	0.069	0.302	0.126	0	0.087	0.117	0.119
	Mutation call	Present	Present	Present	Present	Absent	Present	Present	Present
ALB-Inf-04-14-21-A	All_reads-actual	364	1788	1423	1459	12667	769	3720	2285
	Unsupport_reads-actual	364	1788	1423	1459	12667	769	3720	2285
	Support_reads-actual	0	0	0	0	0	0	0	0
	All_reads-normalized	100	100	100	100	100	100	100	100
	Unsupport_reads-normalized	100	100	100	100	100	100	100	100
	Support_reads-normalized	0	0	0	0	0	0	0	0
	Proportion	0	0	0	0	0	0	0	0
	Mutation call	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
ALB-Inf-04-21-21-A	All_reads-actual	4	61	107	65	444	81	117	243
	Unsupport_reads-actual	2	48	95	64	444	0	115	243
	Support_reads-actual	2	13	12	1	0	0	2	0
	All_reads-normalized	4	61	100	65	100	81	98.3	100
	Unsupport_reads-normalized	2	48	88.8	64	100	0	1.7	100
	Support_reads-normalized	2	13	11.2	1	0	0	0	0
	Proportion	0.500	0.213	0.112	0.015	0	0	0.017	0
	Mutation call	Insuff	Present	Present	Absent	Absent	Absent	Absent	Absent
COR-25TH-04-04-21-A	All_reads-actual	59	333	753	355	45	413	42	281
	Unsupport_reads-actual	0	23	28	0	45	31	14	281
	Support_reads-actual	59	310	725	355	0	382	28	0
	All_reads-normalized	59	100	100	100	45	100	42	100
	Unsupport_reads-normalized	0	6.9	3.7	0	45	7.5	14	100
	Support_reads-normalized	59	93.1	96.3	100	0	92.5	28	0
	Proportion	1.000	0.931	0.963	1.000	0	0.925	0.667	0

		Mutations‡							
Wastewater samples*	Analysis†	A1763G	C5100T	G13045A	C19524T	28026Δ129	C29741T	A11875G§	C15928T§
	Mutation call	Present	Present	Present	Present	Absent	Present	Present	Absent
COR-26TH-04-04-21-A	All_reads-actual	22	52	44	154	228	1	77	46
	Unsupport_reads-actual	4	9	29	135	228	1	77	46
	Support_reads-actual	18	43	15	19	0	0	0	0
	All_reads-normalized	22	52	44	100	100	1	77	46
	Unsupport_reads-normalized	4	9	29	87.7	100	1	77	46
	Support_reads-normalized	18	43	15	12.3	0	0	0	0
	Proportion	0.818	0.827	0.341	0.123	0	0	0	0
	Mutation call	Present	Present	Present	Present	Absent	Insuff	Absent	Absent
COR-27TH-04-04-21-A	All_reads-actual	1	29	31	49	85	10	40	158
	Unsupport_reads-actual	1	29	20	19	85	10	40	158
	Support_reads-actual	0	0	11	30	0	0	0	0
	All_reads-normalized	1	29	31	49	85	10	40	100
	Unsupport_reads-normalized	1	29	20	19	85	10	40	100
	Support_reads-normalized	0	0	11	30	0	0	0	0
	Proportion	0	0	0.355	0.612	0	0	0	0
	Mutation call	Insuff	Absent	Present	Present	Absent	Absent	Absent	Absent
DAL-Inf-04-19-21-A	All_reads-actual	194	1051	1010	1555	3487	1632	1428	4004
	Unsupport_reads-actual	69	289	395	501	3487	347	1428	797
	Support_reads-actual	125	762	615	1054	0	1285	0	3207
	All_reads-normalized	100	100	100	100	100	100	100	100
	Unsupport_reads-normalized	35.6	27.5	39.1	32.2	100	21.3	100	19.9
	Support_reads-normalized	64.4	72.5	60.9	67.8	0	78.7	0	80.1
	Proportion	0.644	0.725	0.609	0.678	0	0.787	0	0.801
	Mutation call	Present	Present	Present	Present	Absent	Present	Absent	Present

* See Appendix Table 2

† Actual numbers of sequence reads spanning the mutation site (All_reads), the numbers of reads not supporting the presence of the mutation (Unsupport, the reference sequence), and the number of reads supporting the presence of the mutation (Support, mutant sequence) are shown. In addition, the normalized numbers of reads are shown. For normalization, read numbers greater than 100 were reduced proportionately to a total of 100 reads. For read numbers of 100 or less, the actual read number was retained. "Proportion" is the proportion of all reads that carry the mutation. Mutations were called present if the proportion was 0.05 or greater and the number of reads carrying the mutation was 2 or more. Otherwise, the mutation was called Absent. If the total number of reads was 5 or less, the call was "insufficient data" (Insuff).

‡ Mutations found only in clade 1 of B.1.351 as defined in this paper. See Figure 1 and Appendix Figure 1.

§ The mutations A11875G C15928T define sub-clades 1a and 1b respectively, within clade 1. Clade 1 sequences lacking both mutations are defined as clade 1c. See Figure 1.

Appendix Table 5. Detailed sequence read analysis of B.1.351 mutations shared with 1-3 other variants in wastewater samples in Linn County, Oregon and surrounding jurisdictions, March-April 2021

		Mutations‡					
Wastewater samples*	Analysis†	A10323G	11288Δ9	A22206G	G23012A	A23063T	C23664T
ALB-Inf-03-26-21-A	All_reads-actual	38	43	25	7	7	76
	Unsupport_reads-actual	25	22	9	5	5	29
	Support_reads-actual	13	21	16	2	2	47
	All_reads-normalized	38	43	25	7	7	76
	Unsupport_reads-normalized	25	22	9	5	5	29
	Support_reads-normalized	13	21	16	2	2	47
	Proportion	0.342	0.488	0.640	0.286	0.286	0.618
	Mutation call	Present	Present	Present	Present	Present	Present
ALB-Inf-03-31-21-A	All_reads-actual	33	71	38	6	6	136
	Unsupport_reads-actual	10	13	26	1	1	106
	Support_reads-actual	23	58	12	5	5	30
	All_reads-normalized	33	71	38	6	6	100
	Unsupport_reads-normalized	10	13	26	1	1	77.9
	Support_reads-normalized	23	58	12	5	5	22.1
	Proportion	0.697	0.817	0.316	0.833	0.833	0.221
	Mutation call	Present	Present	Present	Present	Present	Present
ALB-Inf-04-07-21-A	All_reads-actual	98	176	75	76	73	402
	Unsupport_reads-actual	79	108	42	19	15	267
	Support_reads-actual	19	68	33	57	58	135
	All_reads-normalized	98	100	75	76	73	100
	Unsupport_reads-normalized	79	61.4	42	19	15	66.4
	Support_reads-normalized	19	38.6	33	57	58	33.6
	Proportion	0.194	0.386	0.440	0.750	0.795	0.336
	Mutation call	Present	Present	Present	Present	Present	Present
ALB-Inf-04-14-21-A	All_reads-actual	3599	4035	401	945	977	10041
	Unsupport_reads-actual	1440	0	401	945	1	10041
	Support_reads-actual	2159	4035	0	0	976	0
	All_reads-normalized	100	100	100	100	100	100
	Unsupport_reads-normalized	40	0	100	100	0	100
	Support_reads-normalized	60.0	100	0	0	99.9	0
	Proportion	0.600	1.000	0	0	0.999	0
	Mutation call	Present	Present	Absent	Absent	Present	Absent
ALB-Inf-04-21-21-A	All_reads-actual	65	74	60	1	1	499
	Unsupport_reads-actual	53	20	44	1	1	499
	Support_reads-actual	12	54	16	0	0	0
	All_reads-normalized	65	74	60	1	1	100
	Unsupport_reads-normalized	53	20	44	1	1	100
	Support_reads-normalized	12	54	16	0	0	0
	Proportion	0.185	0.730	0.267	0	0	0
	Mutation call	Present	Present	Present	Insuff	Insuff	Absent
COR-25TH-04-04-21-A	All_reads-actual	440	966	293	219	218	1125
	Unsupport_reads-actual	6	3	0	9	0	78
	Support_reads-actual	434	963	293	210	218	1047
	All_reads-normalized	100	100	100	100	100	100
	Unsupport_reads-normalized	1.4	0.3	0	4.1	0	6.9
	Support_reads-normalized	98.6	99.7	100	95.9	100	93.1
	Proportion	0.986	0.997	1.000	0.959	1.000	0.931

Wastewater samples*	Analysis†	Mutations‡					
		A10323G	11288Δ9	A22206G	G23012A	A23063T	C23664T
	Mutation call	Present	Present	Present	Present	Present	Present
COR-26TH-04-04-21-A	All_reads-actual	108	59	56	1	1	43
	Unsupport_reads-actual	55	28	56	1	1	43
	Support_reads-actual	53	31	0	0	0	0
	All_reads-normalized	100	59	56	1	1	43
	Unsupport_reads-normalized	50.9	28	56	1	1	43
	Support_reads-normalized	49.1	31	0	0	0	0
	Proportion	0.491	0.525	0	0	0	0
	Mutation call	Present	Present	Absent	Insuff	Insuff	Absent
COR-27TH-04-04-21-A	All_reads-actual	9	38	18	21	21	78
	Unsupport_reads-actual	6	24	8	21	3	31
	Support_reads-actual	3	14	10	0	18	47
	All_reads-normalized	9	38	18	21	21	78
	Unsupport_reads-normalized	6	24	8	21	3	31
	Support_reads-normalized	3	14	10	0	18	47
	Proportion	0.333	0.368	0.556	0	0.857	0.603
	Mutation call	Present	Present	Present	Absent	Present	Present
DAL-Inf-04-19-21-A	All_reads-actual	1095	1485	649	1215	1183	1895
	Unsupport_reads-actual	206	145	140	124	109	1122
	Support_reads-actual	889	1340	509	1091	1074	773
	All_reads-normalized	100	100	100	100	100	100
	Unsupport_reads-normalized	18.8	9.8	21.6	10.2	9.2	59.2
	Support_reads-normalized	81.2	90.2	78.4	89.8	90.8	40.8
	Proportion	0.812	0.902	0.784	0.898	0.908	0.408
	Mutation call	Present	Present	Present	Present	Present	Present

* See Appendix Table 2

† Actual numbers of sequence reads spanning the mutation site (All_reads), the numbers of reads not supporting the presence of the mutation (Unsupport, the reference sequence), and the number of reads supporting the presence of the mutation (Support, mutant sequence) are shown. In addition, the normalized numbers of reads are shown. For normalization, read numbers greater than 100 were reduced proportionately to a total of 100 reads. For read numbers of 100 or less, the actual read number was retained. "Proportion" is the proportion of all reads that carry the mutation. Mutations were called present if the proportion was 0.05 or greater and the number of reads carrying the mutation was 2 or more. Otherwise, the mutation was called Absent. If the total number of reads was 5 or less, the call was "insufficient data" (Insuff).

‡ Mutations found in B.1.351 sequences but also found in 1 to 3 other variants. See Appendix Table 8 for a list of variants that share these mutations.

Appendix Table 6. Detailed sequence read analysis of B.1.351 mutations shared with >20 other variants in wastewater samples in Linn County, Oregon and surrounding jurisdictions, March-April 2021

		Mutations‡						
Wastewater samples*	Analysis†	C241T	C1059T	C3037T	C14408T	A23403G	G25563T	C28887T
ALB-Inf-03-26-21-A	All_reads-actual	800	45	180	29	25	64	12
	Unsupport_reads-actual	4	0	0	0	0	20	5
	Support_reads-actual	796	45	180	29	25	44	7
	All_reads-normalized	100	45	100	29	25	64	12
	Unsupport_reads-normalized	0.5	0	0	0	0	20	5
	Support_reads-normalized	99.5	45	100	29	25	44	7
	Proportion	0.995	1.000	1.000	1.000	1.000	0.688	0.583
	Mutation call	Present	Present	Present	Present	Present	Present	Present
ALB-Inf-03-31-21-A	All_reads-actual	567	87	239	32	13	274	36
	Unsupport_reads-actual	0	0	0	0	0	0	0
	Support_reads-actual	567	87	239	32	13	274	36
	All_reads-normalized	100	87	100	32	13	100	36
	Unsupport_reads-normalized	0	0	0	0	0	0	0
	Support_reads-normalized	100	87	100	32	13	100	36
	Proportion	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Mutation call	Present	Present	Present	Present	Present	Present	Present
ALB-Inf-04-07-21-A	All_reads-actual	3662	379	649	139	78	439	283
	Unsupport_reads-actual	0	18	0	0	0	45	218
	Support_reads-actual	3662	361	649	139	78	394	65
	All_reads-normalized	100	100	100	100	78	100	100
	Unsupport_reads-normalized	0	4.7	0	0	0	10.3	77.0
	Support_reads-normalized	100	95.3	100	100	78	89.7	23.0
	Proportion	1.000	0.953	1.000	1.000	1.000	0.897	0.230
	Mutation call	Present	Present	Present	Present	Present	Present	Present
ALB-Inf-04-14-21-A	All_reads-actual	31849	2956	13349	1841	2525	7242	1990
	Unsupport_reads-actual	117	2688	1	0	3	7242	1990
	Support_reads-actual	31732	268	13348	1841	2522	0	0
	All_reads-normalized	100	100	100	100	100	100	100
	Unsupport_reads-normalized	0	91	0	0	0	100	100
	Support_reads-normalized	99.6	9.1	100.0	100	99.9	0	0
	Proportion	0.996	0.091	1.000	1.000	0.999	0	0
	Mutation call	Present	Present	Present	Present	Present	Absent	Absent
ALB-Inf-04-21-21-A	All_reads-actual	1126	85	362	83	35	626	78
	Unsupport_reads-actual	1	19	0	0	0	549	59
	Support_reads-actual	1125	66	362	83	35	77	19
	All_reads-normalized	100	85	100	83	35	100	78
	Unsupport_reads-normalized	0.1	19	0	0	0	87.7	59
	Support_reads-normalized	99.9	66	100	83	35	12.3	19
	Proportion	0.999	0.776	1.000	1.000	1.000	0.123	0.244
	Mutation call	Present	Present	Present	Present	Present	Present	Present
COR-25TH-04-04-21-A	All_reads-actual	11695	1247	3461	566	347	1321	292
	Unsupport_reads-actual	8	37	3	0	0	31	10
	Support_reads-actual	11687	1210	3458	566	347	1290	282
	All_reads-normalized	100	100	100	100	100	100	100
	Unsupport_reads-normalized	0.1	3.0	0.1	0	0	2.3	3.4
	Support_reads-normalized	99.9	97.0	99.9	100	100	97.7	96.6
	Proportion	0.999	0.970	0.999	1.000	1.000	0.977	0.966

Wastewater samples*	Analysis†	Mutations‡						
		C241T	C1059T	C3037T	C14408T	A23403G	G25563T	C28887T
	Mutation call	Present	Present	Present	Present	Present	Present	Present
COR-26TH-04-04-21-A	All_reads-actual	513	102	407	48	70	556	94
	Unsupport_reads-actual	0	0	0	0	0	91	57
	Support_reads-actual	513	102	407	48	70	465	37
	All_reads-normalized	100	100	100	48	70	100	94
	Unsupport_reads-normalized	0	0	0	0	0	16.4	57
	Support_reads-normalized	100	100	100	48	70	83.6	37
	Proportion	1.000	1.000	1.000	1.000	1.000	0.836	0.394
	Mutation call	Present	Present	Present	Present	Present	Present	Present
COR-27TH-04-04-21-A	All_reads-actual	358	6	23	25	9	33	31
	Unsupport_reads-actual	0	0	0	0	0	0	19
	Support_reads-actual	358	6	23	25	9	33	12
	All_reads-normalized	100	6	23	25	9	33	31
	Unsupport_reads-normalized	0	0	0	0	0	0	19
	Support_reads-normalized	100	6	23	25	9	33	12
	Proportion	1.000	1.000	1.000	1.000	1.000	1.000	0.387
	Mutation call	Present	Present	Present	Present	Present	Present	Present
DAL-Inf-04-19-21-A	All_reads-actual	25790	2420	5834	495	965	6530	1035
	Unsupport_reads-actual	10	568	4	0	3	2801	293
	Support_reads-actual	25780	1852	5830	495	962	3729	742
	All_reads-normalized	100	100	100	100	100	100	100
	Unsupport_reads-normalized	0.0	23.5	0.1	0	0.3	42.9	28.3
	Support_reads-normalized	100.0	76.5	99.9	100	99.7	57.1	71.7
	Proportion	1.000	0.765	0.999	1.000	0.997	0.571	0.717
	Mutation call	Present	Present	Present	Present	Present	Present	Present

* See Appendix Table 2

† Actual numbers of sequence reads spanning the mutation site (All_reads), the numbers of reads not supporting the presence of the mutation (Unsupport, the reference sequence), and the number of reads supporting the presence of the mutation (Support, mutant sequence) are shown. In addition, the normalized numbers of reads are shown. For normalization, read numbers greater than 100 were reduced proportionately to a total of 100 reads. For read numbers of 100 or less, the actual read number was retained. "Proportion" is the proportion of all reads that carry the mutation. Mutations were called present if the proportion was 0.05 or greater and the number of reads carrying the mutation was 2 or more. Otherwise, the mutation was called Absent. If the total number of reads was 5 or less, the call was "insufficient data" (Insuff).

‡ Mutations found in more than 20 variants including B.1.351.

Appendix Table 7. Detailed sequence read analysis of private mutations exhibited by B.1.351 individuals and found in wastewater samples in Linn County, Oregon and surrounding jurisdictions, March-April 2021

		Mutations‡														
Wastewater samples*	Analysis†	C355T	C904T	C2232T	G7393T	G8017T	C13059T	C15026T	G18382T	C18452T	C24579T	G25130T	A28254C	T28853G	G28048T§	A28111G§
ALB-Inf-03-26-21-A	All_reads-actual	67	43	61	5	119	54	15	34	29	10	13	27	16	48	11
	Unsupport_reads-actual	67	43	61	5	119	54	15	34	29	10	13	27	16	48	11
	Support_reads-actual	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	All_reads-normalized	67	43	61	5	100	54	15	34	29	10	13	27	16	48	11
	Unsupport_reads-normalized	67	43	61	5	100	54	15	34	29	10	13	27	16	48	11
	Support_reads-normalized	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Proportion	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Mutation call	Absent	Absent	Absent	Insuff	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
ALB-Inf-03-31-21-A	All_reads-actual	57	40	45	2	74	73	11	75	75	26	28	71	37	108	32
	Unsupport_reads-actual	57	40	45	2	74	66	11	75	75	26	28	71	37	108	32
	Support_reads-actual	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0
	All_reads-normalized	57	40	45	2	74	73	11	75	75	26	28	71	37	100	32
	Unsupport_reads-normalized	57	40	45	2	74	66	11	75	75	26	28	71	37	100	32
	Support_reads-normalized	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0
	Proportion	0	0	0	0	0	0.096	0	0	0	0	0	0	0	0	0
	Mutation call	Absent	Absent	Absent	Insuff	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
ALB-Inf-04-07-21-A	All_reads-actual	682	540	139	12	209	527	24	152	110	31	64	172	119	486	109
	Unsupport_reads-actual	682	540	139	12	208	527	24	152	110	31	40	172	119	210	103
	Support_reads-actual	0	0	0	0	1	0	0	0	0	0	24	0	0	276	6
	All_reads-normalized	100	100	100	12	100	100	24	100	100	31	64	100	100	100	100
	Unsupport_reads-normalized	100	100	100	12	99.5	100	24	100	100	31	40	100	100	43.2	94.5
	Support_reads-normalized	0	0	0	0	0.5	0	0	0	0	0	24	0	0	56.8	5.5
	Proportion	0	0	0	0	0.005	0	0	0	0	0	0.375	0	0	0.568	0.055
	Mutation call	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent	Absent	Present	Present
ALB-Inf-04-14-21-A	All_reads-actual	6052	2796	730	797	6565	10121	2799	2534	4093	800	1000	649	1989	11410	2541
	Unsupport_reads-actual	6052	2796	730	797	6563	10121	2799	2534	4093	800	1000	649	1989	8	2
	Support_reads-actual	0	0	0	0	2	0	0	0	0	0	0	0	0	11402	2539
	All_reads-normalized	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	Unsupport_reads-normalized	100	100	100	100	100	100	100	100	100	100	100	100	100	0.1	0.1
	Support_reads-normalized	0	0	0	0	0	0	0	0	0	0	0	0	0	99.9	99.9
	Proportion	0	0	0	0	0	0	0	0	0	0	0	0	0	0.999	0.999
	Mutation call	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present
ALB-Inf-04-21-21-A	All_reads-actual	77	236	47	11	135	838	55	61	61	12	9	58	55	446	39
	Unsupport_reads-actual	77	236	47	11	135	838	55	61	61	12	5	58	55	23	10
	Support_reads-actual	0	0	0	0	0	0	0	0	0	0	4	0	0	423	29
	All_reads-normalized	77	100	47	11	100	100	55	61	61	12	9	58	55	100	39
	Unsupport_reads-normalized	77	100	47	11	100	100	55	61	61	12	5	58	55	5.2	10
	Support_reads-normalized	0	0	0	0	0	0	0	0	0	0	4	0	0	94.8	29
	Proportion	0	0	0	0	0	0	0	0	0	0	0.444	0	0	0.948	0.744
	Mutation call	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent	Absent	Present	Present
COR-25TH-04-04-21-A	All_reads-actual	1875	878	186	47	1312	1769	271	293	750	115	326	480	269	34	32
	Unsupport_reads-actual	1875	878	186	43	1312	1769	271	293	750	115	326	480	269	1	32
	Support_reads-actual	0	0	0	4	0	0	0	0	0	0	0	0	0	33	0
	All_reads-normalized	100	100	100	47	100	100	100	100	100	100	100	100	100	34	32
	Unsupport_reads-normalized	100	100	100	43	100	100	100	100	100	100	100	100	100	1	32
	Support_reads-normalized	0	0	0	4	0	0	0	0	0	0	0	0	0	33	0
	Proportion	0	0	0	0.085	0.000	0	0	0	0	0	0	0	0	0.971	0

		Mutations‡														
Wastewater samples*	Analysis†	C355T	C904T	C2232T	G7393T	G8017T	C13059T	C15026T	G18382T	C18452T	C24579T	G25130T	A28254C	T28853G	G28048T§	A28111G§
	Mutation call	Absent	Absent	Absent	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent
COR-26TH-04-04-21-A	All_reads-actual	79	73	65	1	178	254	17	77	77	10	91	93	94	232	35
	Unsupport_reads-actual	79	73	65	1	178	254	17	77	77	10	91	93	94	70	27
	Support_reads-actual	0	0	0	0	0	0	0	0	0	0	0	0	0	162	8
	All_reads-normalized	79	73	65	1	100	100	17	77	77	10	91	93	94	100	35
	Unsupport_reads-normalized	79	73	65	1	100	100	17	77	77	10	91	93	94	30.2	27
	Support_reads-normalized	0	0	0	0	0	0	0	0	0	0	0	0	0	69.8	8
	Proportion	0	0	0	0	0	0	0	0	0	0	0	0	0	0.698	0.229
	Mutation call	Absent	Absent	Absent	Insuff	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present
COR-27TH-04-04-21-A	All_reads-actual	98	20	46	0	115	249	0	57	31	30	7	58	32	85	38
	Unsupport_reads-actual	98	20	46	0	114	249	0	57	31	30	7	58	32	85	38
	Support_reads-actual	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	All_reads-normalized	98	20	46	0	100	100	0	57	31	30	7	58	32	100	38
	Unsupport_reads-normalized	98	20	46	0	99.1	100	0	57	31	30	7	58	32	100	38
	Support_reads-normalized	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0
	Proportion	0	0	0	0	0.009	0	0	0	0	0	0	0	0	0	0
	Mutation call	Absent	Absent	Absent	Insuff	Absent	Absent	Insuff	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
DAL-Inf-04-19-21-A	All_reads-actual	5517	2605	360	105	20000	3084	883	1324	1191	1379	648	1794	754	3541	638
	Unsupport_reads-actual	5517	2605	360	105	19999	3084	883	1324	1191	1379	648	1794	754	1531	163
	Support_reads-actual	0	0	0	0	1	0	0	0	0	0	0	0	0	2010	475
	All_reads-normalized	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	Unsupport_reads-normalized	100	100	100	100	100	100	100	100	100	100	100	100	100	43.2	25.5
	Support_reads-normalized	0	0	0	0	0	0	0	0	0	0	0	0	0	56.8	74.5
	Proportion	0	0	0	0	0	0	0	0	0	0	0	0	0	0.568	0.745
	Mutation call	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present

* See Appendix Table 2

† Actual numbers of sequence reads spanning the mutation site (All_reads), the numbers of reads not supporting the presence of the mutation (Unsupport, the reference sequence), and the number of reads supporting the presence of the mutation (Support, mutant sequence) are shown. In addition, the normalized numbers of reads are shown. For normalization, read numbers greater than 100 were reduced proportionately to a total of 100 reads. For read numbers of 100 or less, the actual read number was retained. "Proportion" is the proportion of all reads that carry the mutation. Mutations were called present if the proportion was 0.05 or greater and the number of reads carrying the mutation was 2 or more. Otherwise, the mutation was called Absent. If the total number of reads was 5 or less, the call was "insufficient data" (Insuff).

‡ Private mutations, namely mutations found in individual clinical B.1.351 clade 1 sequences that were not shared extensively by numerous other B.1.351 clade 1 sequences. Private mutations can be used for phylogenetic analysis and can occasionally be used to identify samples with a very recent common origin. See Appendix Table 9 for a list of private mutations exhibited by the Oregon B.1.351 individual clinical specimens.

§ These two private mutations are shared with B.1.1.7.

Appendix Table 8. Mutations present in B.1.351 shared with other variants

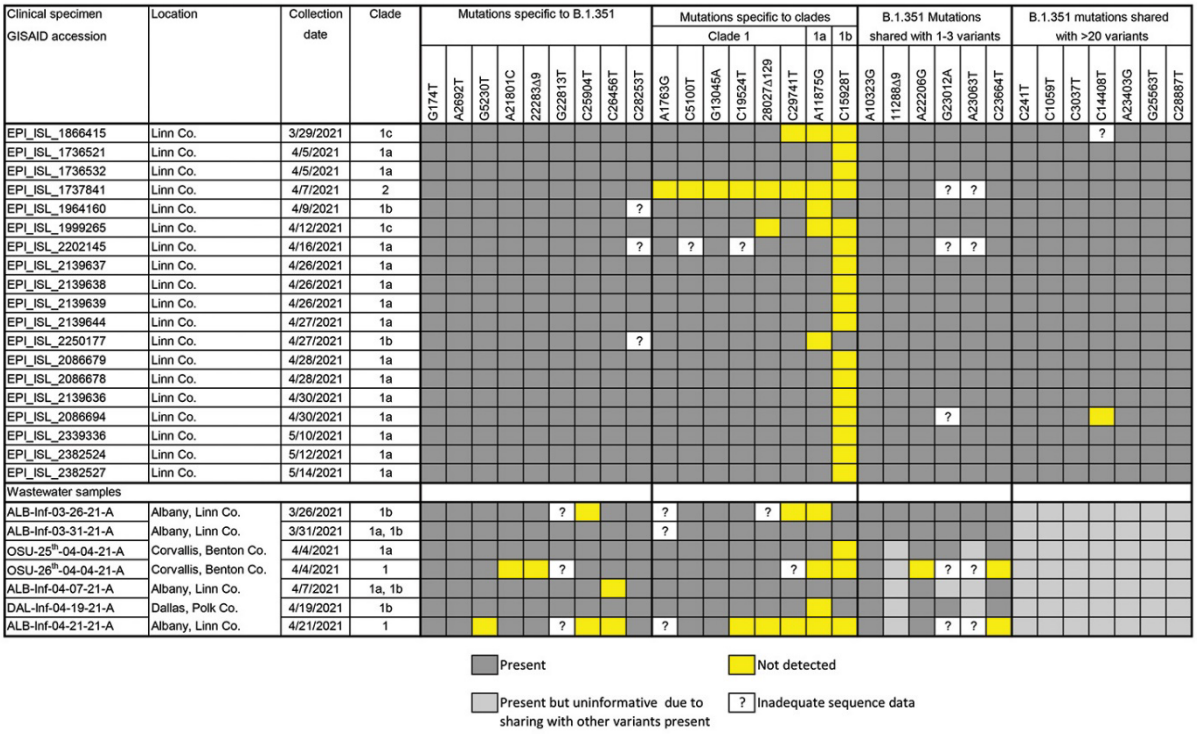
Mutation	Variants sharing the mutation
A10323G	B.1.404, B.1.582
11288Δ9	B.1.1.7, B.1.526, P.1
A22206G	B.1.526
G23012A	B.1.1.316, B.1.526, P.1
A23063T	B.1.1.7, P.1
C23664T	B.1.526

Appendix Table 9. Private mutations exhibited by B.1.351 cases in Linn County, Oregon, March—May 2021

GISAID accession number*	Private mutations†
EPI_ISL_2382524	G25130T, G7393T
EPI_ISL_1736532	G25130T, G18382T
EPI_ISL_1736521	C24579T
EPI_ISL_2139636	C24579T, C2232T
EPI_ISL_2382527	C24579T, C904T, C18452T, T28853G
EPI_ISL_2202145	G8017T
EPI_ISL_2139637	C355T
EPI_ISL_2139638	C355T
EPI_ISL_2139644	C355T
EPI_ISL_2139639	G29254T
EPI_ISL_2339336	G29254T
EPI_ISL_2086679	G29254T
EPI_ISL_2086678	none
EPI_ISL_2086694	none
EPI_ISL_1964160	A28254C
EPI_ISL_2250177	A28254C, C15026T
EPI_ISL_1866415	G29402T
EPI_ISL_1999265	C13059T, G28048T, A28111G
EPI_ISL_1737841	C11514T

* GISAID accession numbers are listed for each specimen. The sequences are listed in the same order, top to bottom, in which they appear in Figure 1 and Appendix Figure 2.

† Private mutations are defined as mutations that occur in small numbers of sequences and that do not define a variant, clade or sub-clade.



Appendix Figure. Mutations detected in B.1.351 clinical specimens and wastewater samples in Linn County, Oregon and surrounding jurisdictions, March-May 2021. Clinical specimens are identified by their GISAID accession numbers. Wastewater samples are identified by their field collection identifier. Cities where individual clinical specimens were collected are not provided to reduce identifiability of cases. Mutations are described relative to the Wuhan-Hu-1 reference sequence. Mutations were considered present in wastewater samples if at least 5% of sequence reads exhibited the mutation and at least two reads exhibited the mutation. Inadequate sequence data was defined as a gap spanning a mutation in individual specimen data or fewer than six reads spanning a mutation in the case of wastewater sample data. Mutations were considered uninformative when other variants with shared mutations were present. Additional mutations found in small numbers of individual clinical specimens (private mutations) are listed in Appendix Table 9.