TITLE

Comparative evaluation of clinical and wastewater genomic surveillance for SARS-CoV-2: implications for integrated infectious disease monitoring

Tarah Lynch*[1,3], Andrew Lindsay[2], Matthew Croxen[2,4], Abdul Qureshi[2], Melodie Barnett[2], Jennavieve Bu[1], Theodore Chiu[2], Ashwin Deo[2], Paul Dieu[1], Xiaoli Dong[1], Christina Ferrato[1], Stefan Gavriliuc[2], Fitsum Getachew[1], Kara Gill[1], Linnette Imaraj[4], Roshan Khadka[1], Petya Koleva[1], Linda Lee[1], Vincent Li[2], Colin Lloyd[5], Emily McCullough[2], Stephanie Murphy[5], Mohammed Abu Naser Mohon[1], Osahon Obasuyi[2], Kanti Pabbaraju[1], Navkiran Randhawa[2], Silas Rotich[2], Sandy Shokoples[2], Sali Sidig[2], Todd Skitsko[2], Johanna Skogen[1], Hilary Trevor[2], Anita Wong[1], Christina Yu[2], Xiaoli Pang[4], Judy Qiu[2,4], Natalie Marshall[2,4], Mathew Diggle[2,4,6,7], Nathan Zelyas[2,4], Maria Falsetti[2], Graham Tipples[2,4,6,7]

[1]Alberta Precision Laboratories, Alberta Health Services, Calgary, Alberta, Canada

[2]Alberta Precision Laboratories, Alberta Health Services, Edmonton, Alberta, Canada

[3]Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, Alberta, Canada

[4] Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, Alberta, Canada

[5]National Microbiology Laboratory, Public Health Agency of Canada, Edmonton, Alberta, Canada

[6]Department of Medical Microbiology and Immunology, University of Alberta, Edmonton, Alberta, Canada

[7]Li Ka Shing Institute of Virology, University of Alberta, Edmonton, Alberta, Canada

*** Correspondence:**
Corresponding Author
tarah.lynch@albertaprecisionlabs.ca

ABSTRACT

Introduction The COVID-19 pandemic demonstrated the need for comprehensive, cost-effective surveillance systems integrating multiple data streams. This study directly compares SARS-CoV-2 genomic data from wastewater-based surveillance (WBS) and clinical diagnostic testing (CDT) to evaluate lineage diversity, detection timing, and persistence patterns that could inform integrated surveillance strategies.

Methods We analyzed SARS-CoV-2 genomic data from Alberta, Canada (July 2022 – March 2025), encompassing 13 municipal wastewater treatment plants covering 80% of the provincial population and clinical samples from provincial diagnostic testing. Clinical samples (n=28,610) and wastewater samples (n=1,685) were sequenced using a tiled amplicon approach. We compared lineage richness over time, lead time for first detection using collection dates and explored four additional wastewater metrics: abundance at first detection, peak abundance, time to peak abundance, and total time detected. The comparison grouped lineages into those found only in WBS and those that were seen in both WBS and CDT.

Results Of the 2,586 unique lineages identified over the study period, 1,588 (61.1%) appeared exclusively in WBS, 42 (1.6%) only in CDT, and 956 (36.9%) in both systems. WBS consistently demonstrated higher monthly lineage richness (95-660 lineages) compared to CDT (23-160 lineages). While WBS detected lineages an average of almost 12 days earlier than CDT, the most frequent pattern showed CDT detection first by 8 days, indicating substantial variability. Lineages detected in both systems showed significantly higher initial relative abundance, peak relative abundance, longer persistence, and delayed time to peak compared to WBS-only lineages (all $p<0.0001$).

Conclusions WBS and CDT provide complementary surveillance capabilities with distinct strengths. Rather than relying on "*first detection*" as an early warning metric, integrated surveillance should prioritize concordance patterns and abundance metrics that indicate lineages with sustainable transmission potential. These findings support developing surveillance frameworks that strategically combine population-level WBS monitoring with case-linked CDT data for more effective public health response.

KEYWORDS

SARS-CoV-2, COVID-19, public health genomics, clinical diagnostic testing, wastewater-based surveillance, integrated surveillance system, infectious diseases

KEY MESSAGES

What is already known on this topic

- WBS can detect SARS-CoV-2 and other pathogens at the population level and has been described as an early warning system, while CDT provides case-linked genomic data but may be biased towards symptomatic, high-risk or healthcare-seeking individuals.

- Both surveillance methods were used during the COVID-19 pandemic, but it remains uncertain how to best integrate WBS and CDT genomics and which WBS metrics are most informative.

What this study adds

- This study provides the first direct comparison of SARS-CoV-2 genomic data from WBS and CDT processed by a single laboratory over nearly three years.
- WBS consistently captured greater lineage diversity than CDT, but "first detection" showed substantial variability, with lineages frequently detected in clinical samples before wastewater.
- Lineages identified in both surveillance systems showed distinct signatures: higher peak abundance and longer persistence, suggesting these concordance patterns are more reliable indicators than timing alone.

How this study might affect research, practice or policy

- Integrated surveillance frameworks should prioritize monitoring concordance between WBS and CDT, using abundance and persistence metrics to identify lineages with significant transmission potential.
- The complementary strengths of these systems support their combined use for cost-effective surveillance with WBS monitoring population trends and CDT providing clinical context for targeted interventions.

INTRODUCTION

The COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) accelerated our need for surveillance systems that are cost-efficient, resilient, and adaptable. Clinical diagnostic testing (CDT) is essential for patient management, monitoring outcomes, isolating pathogens (cultured or genomically) for advancing characterization, transmission analysis, and tracking pathogen evolution (1–3). Wastewater-based surveillance (WBS) offers a stable population-level signal even when clinical testing wanes. Combining complementary data streams of CDT and WBS may be the way forward to balance unpredictable or limited resources for public health priorities (4), however the direct comparison of the genomics role has not been fully understood. This study addresses this gap by directly comparing genomic data from CDT and WBS to evaluate diversity, persistence, and early detection signals.

The algorithms for molecular testing evolved through the pandemic and into the post-pandemic era. Quantitative PCR (qPCR) assays were developed for fast detection and relative quantification of SARS-CoV-2 viral load (5). SARS-CoV-2 was subsequently

integrated into laboratory-developed and commercial respiratory multi-pathogen panels for CDT while the single target qPCR assay continued to be used for quantifying community trends in WBS (6). Whole genome sequencing quickly gained value to characterize mutations across the full virus genome for lineage designation. In CDT genomics produced a consensus sequence to be shared nationally and internationally to track viral evolution and inform lineage designation (3,7). In WBS, a similar full genome approach was taken to estimate SARS-CoV-2 lineages and their relative abundance, however consensus sequences were not produced (8).

Genomics has proven a valuable tool in outbreak investigations and pathogen evolution studies. During the COVID-19 pandemic, global sequencing of clinical samples enabled variant designation, pathogenicity studies, and vaccine design through data sharing in public repositories. Evidence suggests that WBS can predict clinical burden increases 5-63 days before clinical indicators emerge and detect variants of concern up to 14 days earlier than CDT (8–11). Yet integration of WBS and CDT for effective public health response remains uncertain. The WBS metrics most informative for public health decision-making or understanding disease dynamics need to be identified.

The “first detection” has often been treated as evidence for use of WBS as an early warning system, but it is not obvious when this metric could signal actionable risk, as this statement is often retrospectively used. Additional variables affecting the use of CDT and/or WBS include financial considerations, genomics capacity, and the specific goals of the surveillance program.

This study had three primary objectives: (1) to compare lineage diversity patterns between WBS and CDT, (2) to evaluate the reliability of “first detection” as an early warning metric for WBS, and (3) to identify WBS metrics that correlate with lineages detected in CDT, potentially indicating greater transmission or clinical significance.

MATERIALS AND METHODS

Sample population

This study included CDT and WBS across the province of Alberta, Canada from July 1, 2022 – March 31, 2025, when the Provincial Public Health Laboratory (ProvLab) was conducting both for SARS-CoV-2. This timeframe covers the height of the Omicron wave, through the closure of community assessment centers (March 31, 2023) and into the new paradigm of rapid antigen kits for self-testing at home and more relaxed quarantine guidelines. During the pandemic, molecular testing for SARS-CoV-2 was added to commercial respiratory pathogen panels, thus testing for symptomatic and high-risk individuals continued. A

133 limited amount of community testing for SARS-CoV-2 also occurred through the Sentinel
134 Practitioner Surveillance Network (SPSN) (12).

135 Due to fluctuations in the number of SARS-CoV-2 positive CDT samples throughout the
136 study period, it was not feasible to sequence all samples. Selections were made to
137 prioritize patients admitted to care facilities, outbreaks, and a selection of geographically
138 and age diverse community-based patients across Alberta. This selection skewed the
139 sequenced CDT samples to include more severe infections.

140 **CDT Protocol**

141 SARS-CoV-2 samples were collected by flocked nasopharyngeal swab administered by a
142 health professional and collected in universal or viral transport media. Multiple testing
143 platforms were used for the initial diagnosis (qPCR assay) and resulted to the
144 patient/healthcare provider; samples with Ct<30 was reflexed to the Genomics section at
145 ProvLab for whole genome sequencing using the Freed primers (v5) (13). Illumina libraries
146 were prepared using the DNA Prep Kit (Illumina, San Diego, USA) while the Nanopore
147 libraries were prepared with the Ligation Sequencing Kit LSK110 (Oxford Nanopore
148 Technologies, Oxford, UK). Whole genome sequencing was performed once or twice per
149 week on either the MiniSeq Mid Output 300 Cycle Kit (Illumina) or the MinION R9.4.1 flow
150 cell (Oxford Nanopore Technologies). Frequency of sample processing was dependent on
151 positive qPCR sample volumes.

152 Genomic data was analyzed using the ARTIC protocol [https://artic.network/viruses/sars-
153 cov-2/sars-cov-2-bioinformatics-sop.html], with automation scripts by the Connor Lab
154 [https://github.com/connor-lab/ncov2019-artic-nf]. Additional quality assurance metrics
155 were produced using ncov-tools [https://github.com/jts/ncov-tools]. All clinical samples
156 were re-analyzed for this study using Pangolin (v 4.3.1), Pangolin-data (v.1.34), and UShER
157 (v 0.6.3) for lineage assignment. Consensus sequences were submitted to the National
158 Microbiology Laboratory Branch for national surveillance, and are available in GISAID, and
159 the Canadian VirusSeq repositories.

160 **WBS Protocol**

161 Thirteen samples were collected from 13 wastewater treatment plants located in 11
162 municipalities across Alberta, covering approximately 80% of the population. The
163 municipalities included were: Grand Prairie, Fort McMurray, Jasper, Fort Saskatchewan,
164 Edmonton, Red Deer, Banff, Calgary (3 facilities), High River, Medicine Hat, and Lethbridge.

165 Samples were sequenced weekly, full genome amplification of SARS-CoV-2 was done
166 using the ARTIC primer scheme (v 5.3) [https://github.com/quick-lab/SARS-CoV-2] primers
167 and sequenced on the Illumina MiniSeq platform. Sequencing data were analysed with the
168 same artic and ncov-tools pipelines, however instead of consensus sequence generation,
169 BAM files were analyzed with Freyja [https://github.com/andersen-lab/Freyja] using default
170 settings with a minimum relative abundance of 0.1%. All samples were re-analyzed for this
171 study using the latest versions, Freyja (v1.5.1) and UShER (v 0.6.3).

172 **Statistical analysis**

173 Analyses were completed using R (v4.3.0) with the tidyverse package suite (14). Lineage
174 richness was defined as the total number of unique Pangolin lineages per month. First
175 detection lead time was calculated as the difference in collection dates between first
176 appearance in WBS and CDT for each lineage, with negative values indicating earlier
177 clinical detection.

178 Given the non-normal distributions of abundance and persistence datea, comparisons
179 between lineages detected exclusively in WBS versus those found in both systems used the
180 Wilcoxon rank-sum tests. A p-value threshold of 0.01 was used for statistical significance.

181 **Ethics**

182 This study was reviewed and approved by the Conjoint Health Research Ethics Board at the
183 University of Calgary (Ethics ID: REB21-0877_REN4). All consensus sequences from CDT
184 are available through GISAID and the Canadian VirusSeq portal.

185 **Funding**

186 This work was supported by the Public Health Agency of Canada Integral Genomic
187 Innovations, grant number 2223-HQ-000215 and the Public Health Agency of Canada
188 Wastewater Surveillance Fund, grant number 2324-HQ-000049.

189

190 ## RESULTS

191 **CDT testing volumes**

192 Alberta maintained a high proportion of genomic sequencing of SARS-CoV-2 through the
193 pandemic and beyond (Figure 1). The criteria for successful genomic sequencing were
194 based on an empirical threshold qPCR Ct <30 and 85% coverage of the full genome length.
195 The positivity rate of CDT qPCR tests was between 1 – 25% with peaks during high testing

periods and a gradual decrease over the study period. A decrease in testing volumes can be seen from late 2022 to early 2023 when assessment centers closed and at-home rapid testing became available (Figure 1A). The testing volumes then begin to mirror seasonal respiratory patterns as SARS-CoV-2 was added to multi-target respiratory pathogen panels. During the height of the Omicron wave (mid-late 2022), the proportion of genomics were at their lowest (30-40%) as test volumes exceeded capacity. The proportion of successful genomics testing increased to 80% but then plateaus between 50-60% as both qPCR and genomics volumes decrease into the post-pandemic era (Figure 1B).

Lineage diversity patterns across surveillance systems

Over the study period there was a total of 1588 (61.1%) unique lineages found in WBS only, 42 (1.6%) lineages found only in CDT, and 956 (36.9%) identified in both datasets. Monthly lineage richness was calculated for both WBS and CDT to assess temporal variation in diversity (Figure 2). CDT genomic test volumes ranged from 63 – 2,947 per month, compared to 28 – 52 WBS samples. WBS exhibited substantially higher lineage richness throughout the study period (95 – 660 lineages/month) compared with CDT (23 – 160 lineages/month). WBS richness peaked in early 2023, declined through late 2023, and rose again in early 2025, whereas clinical richness remained lower and showed only modest fluctuations. Overall, WBS consistently recovered greater lineage diversity than CDT.

Lineage first detection comparison

We calculated the lead time between first detection in WBS and CDT using collection dates for the 956 lineages detected in both datasets (Figure 3). Collection dates were used because although processing times for WBS and CDT are similar (7-14 days), the differing reporting pathways make reported dates unsuitable for comparison. Lead times ranged from -612 to 767 days, with negative values indicating earlier CDT detection (n=385) and positive values indicating earlier WBS detection (n=559); 11 lineages were detected on the same day. The distribution had a mean lead time of 11.9 days (WBS earlier) but a mode of -8 days (CDT earlier), reflecting substantial variability in which system detected a lineage first.

WBS metrics to identify patterns of lineages found in WBS or both

To assess whether WBS metrics correlated with clinical detectability, we compared lineages observed exclusively in WBS with those found in both WBS and CDT (Figure 4). Among the 1,588 wastewater-only and 956 shared lineages, all four metrics: abundance at first detection (Fig. 4A), maximum abundance (Fig. 4B), time to peak abundance (Fig. 4C), and total time detected (Fig. 4D) – all differed significantly (Wilcoxon, $p < 0.0001$). Lineages detected in both WBS and CDT appeared at higher initial abundance, reached higher

231 abundance peak, took longer to reach maximum abundance, and persisted longer in
232 wastewater. In contrast, wastewater-only lineages were generally lower in abundance,
233 peaked sooner, and were detected for shorter durations.

234 DISCUSSION

235 This study provides the first detailed comparison of SARS-CoV-2 genomic data from WBS
236 and CDT processed through a single laboratory over an extended timeframe spanning
237 multiple pandemic phases. Our findings demonstrate that these surveillance approaches
238 capture different dimensions of viral circulation and should be integrated strategically
239 rather than viewed as redundant or competing systems.

240 WBS and CDT methods have unique benefits and limitations for infectious disease
241 surveillance (summarized in Table 1). WBS is non-invasive, provides community-level
242 coverage, and includes all severity levels of infection regardless of healthcare access. The
243 purpose of CDT is individual-level results to track symptoms and outcomes required for
244 patient management. While both approaches proved valuable during the pandemic,
245 optimal integration strategies for public health action remain unclear.

246 Our analysis demonstrates that WBS consistently captured greater lineage diversity than
247 CDT throughout the study period. Despite CDT's bias towards higher-risk groups and more
248 severe infections, the significant overlap between systems (36.9% of lineages detected in
249 both) suggests CDT remains representative of predominant circulating virus. This finding
250 gains support from our observation that lineages detected in both systems showed higher
251 relative abundance and greater persistence in wastewater, potentially reflecting increased
252 viral shedding associated with symptomatic or severe illness, though viral shedding
253 kinetics are multifactorial (15).

254 A major finding challenges the widely held assumption that WBS serves as an early warning
255 system for emerging lineages. While WBS detected more lineages earlier on average (mean
256 = 11.9 days), the most frequent pattern showed CDT detection occurring first (mode = -8
257 days) with the distribution stretching much longer. This substantial variability suggests that
258 "first detection" alone is not a reliable metric for WBS-based early warning systems.
259 Instead, concordance between WBS and CDT appears more informative, as share lineages
260 demonstrated higher abundance and persistence, characteristics that may indicate greater
261 transmission potential and clinical significance.

262 These findings support developing surveillance strategies that leverage the complementary
263 strengths of both approaches. WBS excels at monitoring population-level trends and can
264 inform community-wide interventions such as public health messaging, vaccination
265 campaigns, and preventative measures (16,17). CDT provides case-linked data crucial for

clinical interventions and detailed pathogen characterization. Integrating these approaches enables public health programs to balance broad population monitoring with targeted clinical responses in a cost-effective manner.

Several limitations warrant consideration. Different primer schemes between WBS (ARTIC v5.3) and CDT (Freed v5) may have introduced bias affecting genome coverage and lineage designation, though all met established quality standards. The study did not examine clinical factors such as vaccination status or symptom severity, which may influence lineage diversity and circulation patterns. Third, collection dates across WBS sites were often recorded as a single day each week and therefore may impact the lead time calculations by +/- up to 3 days compared to CDT samples where collection dates are recorded to the minute.

Future studies should integrate more comprehensive clinical variables with WBS data to develop predictive models that better utilize both surveillance streams. Research examining the relationship between WBS data and clinical outcomes could enhance our understanding of which WBS parameters best predict significant transmission events. Modelling could also determine optimal sampling strategies and testing requirements for CDT based on population and pathogen transmission dynamics. Long-term studies across different pathogen systems would help validate these findings beyond SARS-CoV-2.

CONCLUSIONS

This study demonstrates that WBS and CDT provide complementary rather than redundant surveillance capabilities for SARS-CoV-2. Rather than relying on “first detection” as an early warning metric, integrated surveillance frameworks should prioritize concordance patterns and abundance metrics that identify lineages with sustainable transmission potential. WBS consistently captures greater lineage diversity and provides stable population-level monitoring, while CDT offers essential clinical context for targeted interventions.

ACKNOWLEDGEMENTS

We would like to acknowledge additional contributions from Qiang Jiang, Melissa Wilson, Kimberly Tang, Poonam Patel, Seema Kailey, and Sydney Rudko for their collaborations.

REFERENCES

1. Roberts CC. Emerging infectious disease laboratory and diagnostic preparedness to accelerate vaccine development. Hum Vaccines Immunother. 2019 Oct 3;15(10):2258–63.

2. Ladner JT, Sahl JW. Towards a post-pandemic future for global pathogen genome sequencing. PLOS Biol. 2023 Aug 1;21(8):e3002225.

3. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 2021 Dec 16;7(2): veab064.

4. Islam MdA, Hossen F, Rahman MdA, Sultana KF, Hasan MN, Haque MdA, et al. An opinion on Wastewater-Based Epidemiological Monitoring (WBEM) with Clinical Diagnostic Test (CDT) for detecting high-prevalence areas of community COVID-19 infections. Curr Opin Environ Sci Health. 2023 Feb;31:100396.

5. Pabbaraju K, Zelyas N, Wong A, Croxen MA, Lynch T, Buss E, et al. Evolving strategy for an evolving virus: Development of real-time PCR assays for detecting all SARS-CoV-2 variants of concern. J Virol Methods. 2022 Sep;307:114553.

6. Qiu Y, Yu J, Pabbaraju K, Lee BE, Gao T, Ashbolt NJ, et al. Validating and optimizing the method for molecular detection and quantification of SARS-CoV-2 in wastewater. Sci Total Environ. 2022 Mar;812:151434.

7. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020 Jul 15;5(11):1403–7.

8. Karthikeyan S, Levy JI, De Hoff P, Humphrey G, Birmingham A, Jepsen K, et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature. 2022 Sep 1;609(7925):101–8.

9. Li L, Uppal T, Hartley PD, Gorzalski A, Pandori M, Picker MA, et al. Detecting SARS-CoV-2 variants in wastewater and their correlation with circulating variants in the communities. Sci Rep. 2022 Sep 27;12(1):16141.

10. Joung MJ, Mangat C, Mejia E, Nagasawa A, Nichani A, Perez-Iratxeta C, et al. Coupling wastewater-based epidemiological surveillance and modelling of SARS-COV-2/COVID-19: Practical applications at the Public Health Agency of Canada. Can Commun Dis Rep. 2023 May 31;49(5):166–74.

11. Shah S, Gwee SXW, Ng JQX, Lau N, Koh J, Pang J. Wastewater surveillance to infer COVID-19 transmission: A systematic review. Sci Total Environ. 2022 Jan;804:150060.

12. Skowronski DM, Zhan Y, Kaweski SE, Sabaiduc S, Khalid A, Olsha R, et al. 2023/24 mid-season influenza and Omicron XBB.1.5 vaccine effectiveness estimates from the Canadian Sentinel Practitioner Surveillance Network (SPSN). Eurosurveillance [Internet]. 2024 Feb 15 [cited 2025 Oct 2];29(7). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2024.29.7.2400076

13. Freed NE, Vlková M, Faisal MB, Silander OK. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. Biol Methods Protoc. 2020 Jan 1;5(1):bpaa014.

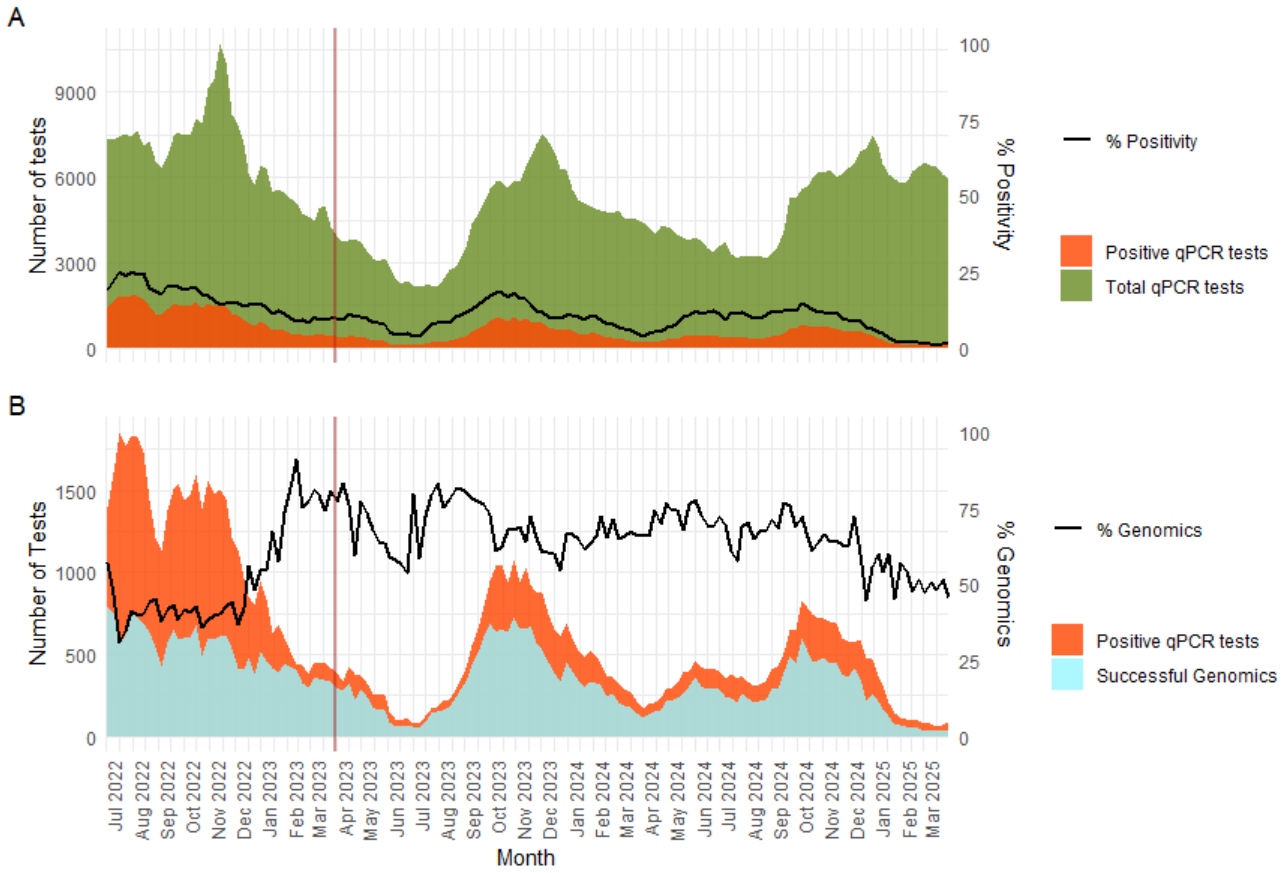
14. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019 Nov 21;4(43):1686.

15. Puhach O, Meyer B, Eckerle I. SARS-CoV-2 viral load and shedding kinetics. Nat Rev Microbiol [Internet]. 2022 Dec 2 [cited 2025 Oct 2]; Available from: https://www.nature.com/articles/s41579-022-00822-w

16. Frieden TR. A Framework for Public Health Action: The Health Impact Pyramid. Am J Public Health. 2010 Apr;100(4):590–5.

17. Havelaar AH, Bräunig J, Christiansen K, Cornu M, Hald T, Mangen M -J. J, et al. Towards an Integrated Approach in Supporting Microbiological Food Safety Decisions. Zoonoses Public Health. 2007 May;54(3–4):103–17.

363 ## FIGURES



364

365 **Figure 1.** Clinical qPCR and genomic testing volumes and positivity rates. Panel A
366 compares the total number of SARS-CoV-2 qPCR tests (green) to the subset with positive
367 results (orange) and % positivity (black line). Panel B is the proportion of successful
368 genomic sequencing (blue) as a proportion of positive qPCR tests (orange) and the % of
369 samples successfully sequenced (black line). The red vertical line denotes closing of
370 community assessment centers on March 31, 2023.

371

372

373

374

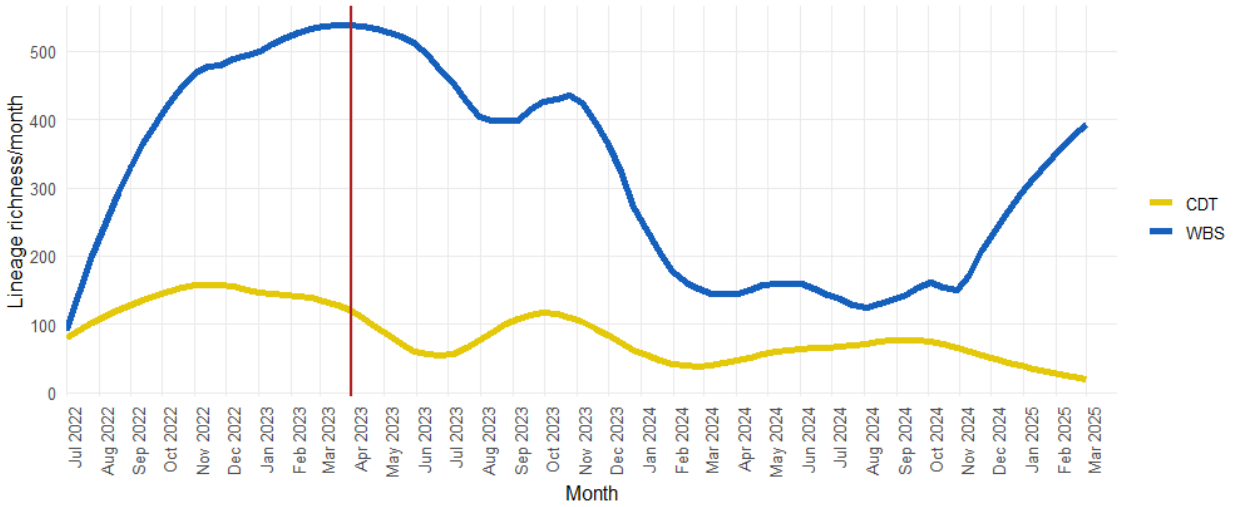


Figure 2. Lineage richness over time for WBS (blue) and CDT (yellow). The red vertical line denotes the closing of the community assessment centers. Lineage richness is calculated as total unique lineages per month.

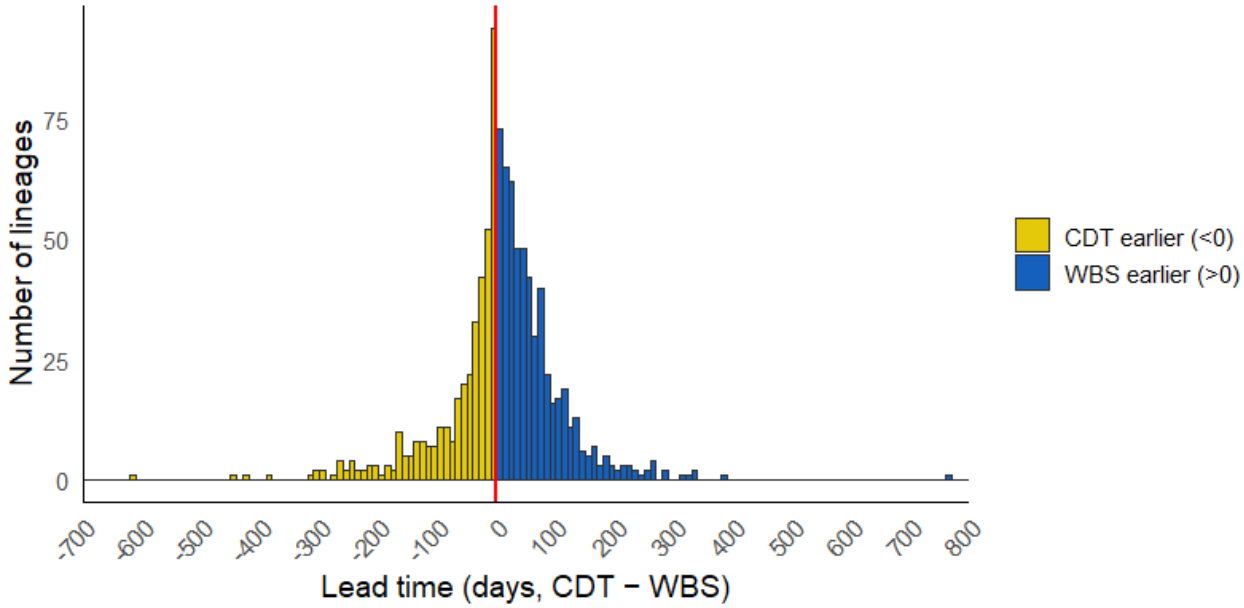


Figure 3. Frequency and timing of first lineage detection in WBS or CDT. Of lineages present in both datasets, lead time is the difference in days between first detection in WBS and CDT using the collection dates. Lead time is displayed on the x-axis: < 0 days are lineages detected in CDT first, 0 = lineages first detected same day, and >0 days are lineages first detected in WBS.

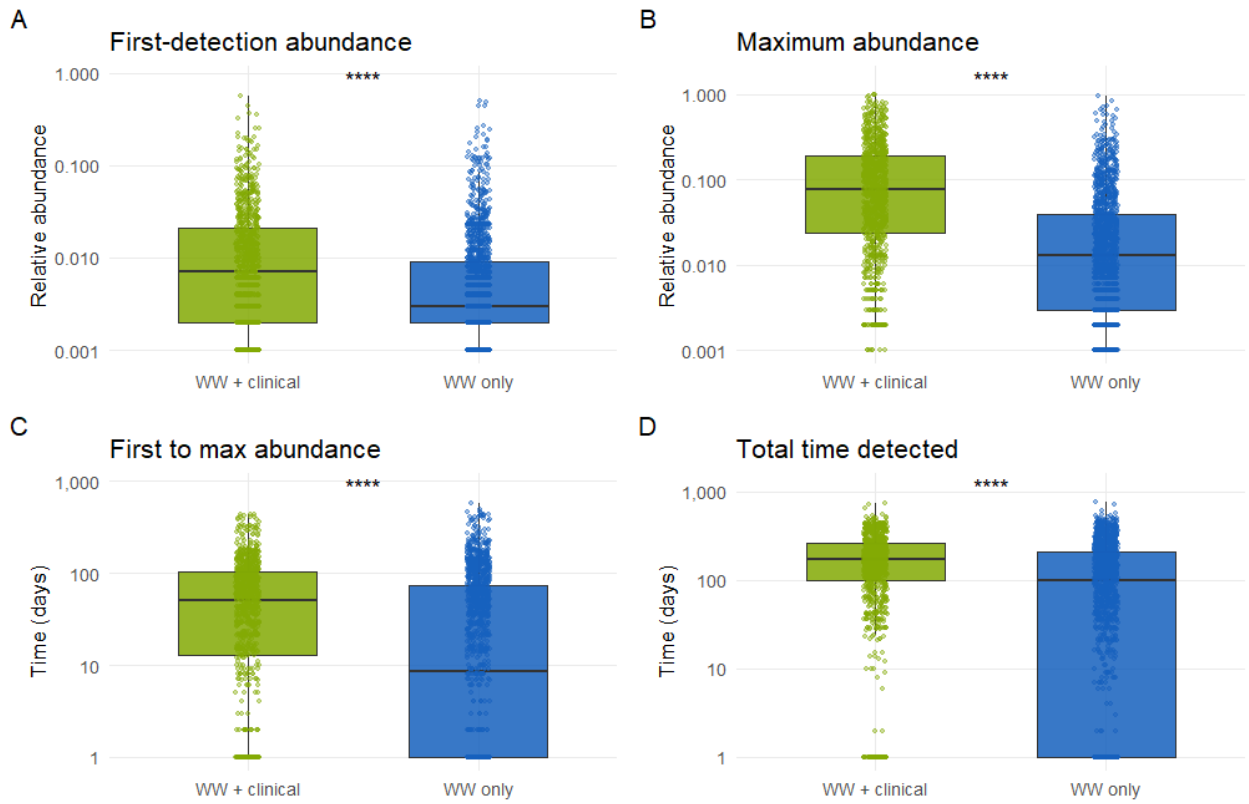


Figure 4. Exploring WBS data to compare lineages exclusively in WBS (WW only, blue) against lineages found in both WBS and CDT (WW + Clinical, green). Total number of lineages in WW only = 1588, WW + Clinical = 967. All comparisons used Wilcoxon rank sum test (**** = $p < 0.0001$). (A)= relative abundance at first detection, (B)= maximum abundance reached. (C) = length of time from first detection to maximum abundance, (D) = total time detected in WBS. All y-axes are displayed on the log 10 scale.

402 Table 1. Overview of CDT and WBS genomics for infectious disease surveillance

Variable	**Wastewater-Based Surveillance (WBS)**	**Clinical Diagnostic Testing (CDT)**
Scope	Population-level coverage, capturing asymptomatic to severe infections.	May be biased to symptomatic, high-risk, and healthcare-seeking individuals.
Early Detection	Can detect variants earlier but clinical relevance may be unknown.	Detection may lag due to restricted testing but infections with high morbidity/mortality identified early (i.e. hospital cases).
Cost	Laboratory cost equivalent per test, monitors larger urban populations.	Laboratory cost equivalent per test, covers fewer individuals.
Lineage Data	Shows broader diversity and longer detection windows.	Provides high-resolution data linked to clinical outcomes.
Pathogen Isolation	Complex polymicrobial samples – culture isolation possible but increased labour cost. Difficult to produce consensus sequences for data sharing (currently not done, not impossible).	Pathogen culture isolation possible. Genome data shared to public repositories for global phylogenetic placement and further research and characterization.
Limitations	No direct link to severity or outcomes, sewer infrastructure and protocol variability. Data may be missed by diaper users.	Sampling bias, lower population coverage, and limited asymptomatic detection.

403