Science of the Total Environment 948 (2024) 174515

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

A gold standard dataset and evaluation of methods for lineage abundance estimation from wastewater

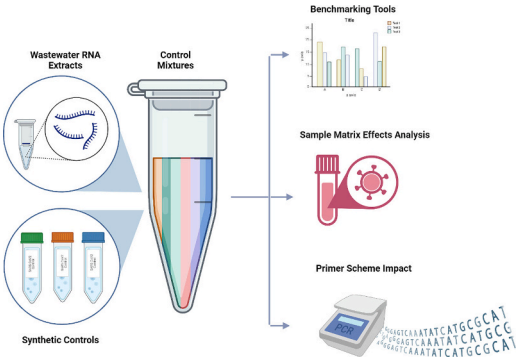
Jannatul Ferdous, Samuel Kunkleman, William Taylor, April Harris, Cynthia J. Gibas, Jessica A. Schlueter *

Department of Bioinformatics and Genomics, UNC Charlotte, 9201 University City Blvd, Charlotte, NC 28223, USA

HIGHLIGHTS

- Generation of a gold standard dataset
- Comparative evaluation of relative abundance estimation software
- Evaluation of deconvolution methods used in CFSAN's C-WAP pipeline

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Jurgen Mahlknecht

Keywords:
SARS-CoV-2
Wastewater
Deconvolution
Benchmark
Control dataset
Deconvoluting tools
Relative abundance

ABSTRACT

During the SARS-CoV-2 pandemic, genome-based wastewater surveillance sequencing has been a powerful tool for public health to monitor circulating and emerging viral variants. As a medium, wastewater is very complex because of its mixed matrix nature, which makes the deconvolution of wastewater samples more difficult. Here we introduce a gold standard dataset constructed from synthetic viral control mixtures of known composition, spiked into a wastewater RNA matrix and sequenced on the Oxford Nanopore Technologies platform. We compare the performance of eight of the most commonly used deconvolution tools in identifying SARS-CoV-2 variants present in these mixtures. The software evaluated was primarily chosen for its relevance to the CDC wastewater surveillance reporting protocol, which until recently employed a pipeline that incorporates results from four deconvolution methods: Freyja, kallisto, Kraken 2/Bracken, and LCS. We also tested Lollipop, a

Abbreviations: SARS, Severe Acute Respiratory Syndrome; COVID-19, Coronavirus Disease 2019; WBE, wastewater-based epidemiology; WB, water background; NWRB, SARS-CoV-2 negative wastewater RNA extract background; PWRB, SARS-CoV-2 positive wastewater RNA extract background; NWSS, National Wastewater Surveillance System; CFSAN, Center for Food Safety and Applied Nutrition; C-WAP, CFSAN Wastewater Analysis Pipeline; ONT, Oxford Nanopore Technologies; NFW, nuclease-free water; RNA, ribonucleic acid; SNV, single nucleotide variant; NCBI, National Center for Biotechnology Information; PCR, polymerase chain reaction; ddPCR, droplet digital PCR; Pangolin, Phylogenetic Assignment of Named Global Outbreak Lineages; VOC, variant of concern; DCIPHER, Data Collation and Integration for Public Health Event Responses; S3C, Swiss SARS-CoV-2 Sequencing Consortium; SIB, Swiss Institute of Bioinformatics; ANOVA, Analysis of variance; HSD, Honestly Significant Difference; Tukey's HSD, Tukey's post hoc Honestly Significant Difference Test.

* Corresponding author.
E-mail address: jschluet@charlotte.edu (J.A. Schlueter).

https://doi.org/10.1016/j.scitotenv.2024.174515
Received 13 February 2024; Received in revised form 20 June 2024; Accepted 3 July 2024
Available online 5 July 2024

deconvolution method used by the Swiss SARS-CoV-2 Sequencing Consortium, and three additional methods not used in the C-WAP pipeline: lineagespot, Alcov, and VaQuERo. We found that the commonly used software Freyja outperformed the other CDC pipeline tools in correct identification of lineages present in the control mixtures, and that the VaQuERo method was similarly accurate, with minor differences in the ability of the two methods to avoid false negatives and suppress false positives. Our results also provide insight into the effect of the tiling primer scheme and wastewater RNA extract matrix on viral sequencing and data deconvolution outcomes.

1. Introduction

SARS-CoV-2 emerged in China in December of 2019 and led to the COVID-19 pandemic (Li et al., 2021). From a public health perspective, as SARS-CoV-2 has continued to mutate, tracking circulating and emerging variants of SARS-CoV-2 has been an essential part of the pandemic response (Aleem et al., 2024). Sequencing-based wastewater surveillance has become a sentinel for monitoring and identifying these new variants and tracking the shifts in variants across populations (Karthikeyan et al., 2022). During outbreaks, COVID-19 clinical samples have been plentiful and provided an adequate basis for identifying the emergence and spread of new variants. However, as governments have changed or ended their commitments to COVID-19 surveillance, PCR testing has been replaced by at-home testing and the few clinical samples available only offer a sampling of individuals who either choose to be tested or are hospitalized.

Wastewater-based epidemiology (WBE) is not a new concept; it has been used since the early 20th century for monitoring other outbreak-causing pathogens (Life, 2021; Ivanova et al., 2019). When early reports in 2020 showed that SARS-CoV-2 was detectable in wastewater and was a leading indicator in advance of spikes in confirmed cases (Peccia et al., 2020), this sparked widespread interest in WBE as a tool for monitoring COVID-19 outbreaks. Wastewater surveillance was implemented on a large scale worldwide, in locations ranging in size from large metropolitan sewersheds (Gregory et al., 2021; Rego et al., 2021) to individual college campus dormitories (Gibas et al., 2021; Solo-Gabriele et al., 2023). The relatively low implementation cost of genome-based wastewater surveillance makes it ideal for areas that lack resources for clinical sample-based sequencing surveillance (Amman et al., 2022) and a useful addition to any areas where clinical sequencing is limited. In response to the COVID-19 pandemic, the CDC launched the National Wastewater Surveillance System (NWSS) in September 2020 (CDC, 2023) and the U.S. Food & Drug Administration (Center for Food Safety, 2023) set up a sequencing project with the collaboration of national and university labs to track and monitor the incidence of SARS-CoV-2. The Center for Food Safety and Applied Nutrition (CFSAN) of the FDA (2013) made their in-house wastewater analysis pipeline, C-WAP, available to collaborating labs that submit wastewater sequencing to the NWSS (C-WAP, n.d.-a). As clinical testing has decreased since 2022, this approach gives epidemiologists and public health officials a means to track the proportion of variants circulating in the community, and can potentially identify new and novel variants as they emerge (Smyth et al., 2021). Recently, the C-WAP pipeline has been archived and is not under active development or maintenance; its successor Aquascope can be used instead and relies primarily on Freyja for deconvolution (C-WAP, n.d.-a).

The most widely used approach for sequencing SARS-CoV-2 viral RNA was developed by the ARTIC Network (Quick, 2020). While the ARTIC protocol was initially developed for clinical samples, it is the foundation for sequencing SARS-CoV-2 from wastewater. ARTIC uses two overlapping pools of primer pairs (Karthikeyan et al., 2022) to tile the whole genome, producing 300–500 bp amplicons which are then pooled and sequenced. This approach is applicable to both intact and fragmented viral RNA, as amplicons will be created whenever the targeted area of sequence is included in an intact segment of RNA, without requiring the presence of the entire genome to be amplified successfully. The full process of wastewater sequencing includes sample concentration, RNA extraction, and target amplification out of the total extracted RNA from each sample (Ferdous et al., 2021). Along with the ARTIC Network's primers, several other amplicon primer sets exist to sequence SARS-CoV-2 and generate full coverage consensus sequences (Ramachandran, 2022; VarSkip, n.d.; Freed et al., 2020; Child, 2022). A second approach employs target enrichment, amplifying only the spike protein regions where many strain-defining mutations occur (Shafer et al., 2022). The subsequent concentration and extraction steps are the same regardless of the amplification approach.

Assembly of amplified viral genomic RNA is achieved by realigning amplicons to a reference genome. The bioinformatic analysis of SARS-CoV-2 sequencing from wastewater is significantly more complicated than it is for clinical samples. With a clinical sample, it is reasonable to assume that an individual is going to be infected with only one variant. The bioinformatics approach to a clinical sample is to align all the amplicons to the reference and then identify the variant calls using standard approaches like minimap and BCFtools (Li, 2018; Danecek et al., 2021).

Wastewater samples are composite collections from the population served by a single building, sewershed, or wastewater treatment facility. Therefore the expectation is that the sample itself is a representation of the variants circulating in that community, and the nature and complexity of the mixture may vary with the population size in the area collected. When wastewater samples are analyzed, rather than simply aligning the sequences collected to a reference, they must be deconvoluted and assigned to specific variants. There are several bioinformatic pipelines available for the analysis of wastewater sequencing data.

As with other environmental samples, wastewater sequencing comes with a variety of challenges. For example, wastewater sample processing can result in biases in downstream quantitation or sequencing processes due to sample quality and sample chemistry (Freyja, n.d.-b). The choice of extraction methods may result in fragmentation of the input RNA (Freyja, n.d.-b), leading to shorter sequence reads of lower quality, and the chemical makeup of extracted samples may include small molecules that inhibit the amplification steps in detection (Gibas et al., 2021; Katayama et al., 2002; Lin et al., 2021) and sequencing. The chemical environment of wastewater leads to viral RNA degradation and fragmentation (Bivins et al., 2020). Low viral concentration requiring an initial sample concentration step is the norm in wastewater samples, and the concentration method during processing can also impact the coverage and quality of sequence data (Karthikeyan et al., 2022). Wastewater also contains both low-frequency and high-frequency variants simultaneously, which can make the detection and relative abundance estimation of the lineages more difficult (Amman et al., 2022), especially since low-abundance components may still be of interest as new variants enter a monitored area. The current bioinformatics tools can often identify the dominant lineage but fail to detect low-frequency or new variants correctly. For these reasons, deconvolution of SARS-CoV-2 variants from mixed viral samples still presents a challenge (Karthikeyan et al., 2022). In this study we tested eight deconvolution methods: Freyja, kallisto, Kraken 2/Bracken, LCS, Lollipop, lineagespot, Alcov, and VaQuERo.

One of the most commonly used tools for wastewater variant sequencing is Freyja (Karthikeyan et al., 2022). It uses a depth weighted least absolute deviation regression algorithm and reports the relative

lineage abundances in mixed viral samples mapping to a common viral reference from a sequencing dataset (Freyja, n.d.-a). Using bam files, it first calculates the frequency of each mutation and its respective sequencing depth. Then, to solve the regression problem, it uses a barcode matrix of lineage-defining mutations obtained from USHER and the mutation frequency and depth information to weight SNV frequency across each mutation site. These weights allow for prioritization of site-specific information as a function of sequencing depth which is ultimately used to generate a relative abundance of each of the known lineages.

Originally employed for abundance quantification of metagenomic transcripts, kallisto (Bray et al., 2016; Baaijens et al., 2021), has been repurposed to work with SARS-CoV-2 reads. Kallisto constructs an index from the RNA transcriptome with a de Bruijn graph that represents transcript k-mers. Reads are then pseudoaligned with the k-mers, and transcript abundance quantification is performed by likelihood function. For wastewater deconvolution, instead of using RNA-Seq transcripts, kallisto constructs the index of k-mer from multiple SARS-CoV-2 consensus sequences. The reads are pseudoaligned and relative abundance is estimated.

LCS (Lineage deComposition for Sars-cov-2 pooled samples) is a mixture model that determines SARS-CoV-2 variant composition in pooled samples by using a previously defined selection of mutations that characterize SARS-CoV-2 variants from publicly available sources and a matrix of variant signatures (Valieris et al., 2022). The matrix corresponds to the probability of finding an alternate sequence at any polymorphism from any variant. Along with the matrix, it uses the sequencing data containing counts of reads mapped to respective polymorphic loci. Minimap generated alignment files are then used for the estimation of the relative frequencies of the variants with maximum likelihood.

Kraken 2 (Wood et al., 2019) is a metagenomic sequence classification tool that uses alignments for taxonomic assignment. Similar to kallisto, Kraken 2 breaks down the sequences into k-mers and uses each of them to calculate a compact hash code to use as a query for finding the Lowest Common Ancestor (LCA) with a space seed searching scheme. This information is stored in a list and then used to form the classification tree where nodes are weighted based on the number of the k-mers linked with the taxon and root-to-leaf (RTL) path is weighted by addition of all the weights. The query sequence is then classified as the leaf to the maximum RTL path. For wastewater deconvolution, Kraken 2 uses each FASTQ sequence as the query and k-mer match to LCA for lineage identification and finally Bracken uses this lineage classification information to perform the relative abundance of sequence.

Freyja is frequently used as a component of other wastewater analysis pipelines, such as the C-WAP/Aquascope pipeline. Kallisto, LCS and Kraken 2 are components of the C-WAP pipeline as well and obvious choices for inclusion in this analysis. In addition to benchmarking the tools that are part of the C-WAP/Aquascope pipeline, we also evaluated four other commonly used lineage abundance estimation methods, as evidenced by the frequency of citations; LolliPop, Alcov, lineagespot and VaQuERo. LolliPop (Dreifuss et al., 2022) is a part of V-Pipe (Posada-Céspedes et al., 2021) and aids in processing viral sequencing data from NGS for the Swiss SARS-CoV-2 Sequencing Consortium (S3C) as a part of the SIB SARS-CoV-2 surveillance program. LolliPop solves the deconvolution problem by using a least square fitting approach and uses kernel-based smoothing to generate higher confidence relative abundance. It is designed to be integrated with Cojac (Jahn et al., 2022) but can also be used independently. Alcov (Abundance Learning of SARS-CoV-2 Variants) treats lineage abundance estimation as an optimization problem using mutation frequencies (Ellmen et al., 2021). Alcov focuses on nonsynonymous mutations only, and for each lineage-defining amino acid variant, the variant amino acid is back-translated into a nucleotide SNV at the appropriate genomic index. Lineage abundance estimation based on multiple variants is cast as an optimization problem using an ordinary least squares (OLS) approach, considering only mutations for which sufficient read depth is available. Lineagespot is another widely used deconvolution tool that detects variants and assigns lineages by the identification of mutational load by quantifying lineage abundance metrics, computing average allele frequency of all amino acid mutations and generates the mutational load as proportions (Pechlivanis et al., 2022). Finally, we also considered VaQuERo, a method developed at CeMM (Center for Molecular Medicine), Vienna (Amman et al., 2022). VaQuERo uses a SIMPLEX regression to deduce overall variant frequencies from the mutation patterns of individually selected variants. Both LolliPop and VaQuERo can use smoothing approaches to increase confidence in variant abundance estimates when presented with time series data, but the other tools are designed for the analysis of single samples.

As with any software tool designed to estimate an unknown, benchmarking and performance evaluation are essential for determining the most accurate software to identify mixtures of SARS-CoV-2 lineages in wastewater. Kayikcioglu et al., 2023 evaluated performance of 5 different deconvolution tools using simulated sequencing datasets [50]. Given that wastewater is a complex matrix, simulated data sets can be affected by assumptions about variant and error frequency that may not represent the behavior of nucleic acids in the wastewater matrix. These evaluations used the genomic simulator DeepSimulator, but the authors acknowledge that the simulation workflow did not take into account factors such as time, temperature, and the chemical composition of the sample matrix. To avoid the assumptions necessary in read simulation and account for the effect of these factors under standard wastewater processing conditions, we have sequenced synthetic RNA control mixtures. Twist viral controls consisting of a mixture of subgenomic fragments were used as a proxy for the genomic fragmentation known to be observed in wastewater. These controls were spiked into both water and complex wastewater extract backgrounds in known concentrations, and then sequenced on the Oxford Nanopore platform to generate a gold standard dataset for benchmarking and evaluation of bioinformatic deconvolution tools. Our choice of the ONT platform, which is widely advertised as a low-cost and portable sequencing platform was motivated by our own experiences in generating data for state and local public health laboratories. Cost considerations were a key driver of our choice of the ONT platform for use in both clinical and wastewater surveillance. In the course of our work, we found that this less-common choice of sequencing platform rather than the widely used Illumina platform allowed us to produce more sequence data with a smaller budget than comparable labs, but that less was known about the performance and reproducibility of the platform.

We compare predictions and identify strengths and weaknesses of available deconvolution tools, with the goal of informing future method development in wastewater analysis. While not comprehensive in its scope, this study addresses the behavior of fragmented RNA molecules in the post-extraction sequencing workflow and provides a model for construction and analysis of standardized wastewater spike-ins that may be needed to address other sampling contexts and stages in the workflow. Apart from validation and standardization of methods, this dataset has the potential to be used for optimization and capacity building for wastewater-based epidemiology. This is also beneficial as it can be used to evaluate and compare the performance of available sequencing platforms. The dataset is available through NCBI's Sequence Read Archive (SRA) under the BioProject accession PRJNA1031245. The protocol for preparation of control mixtures is available at dx.doi.org/10.17504/protocols.io.261ged2jjv47/v1, and analysis scripts and computational protocols are available at https://github.com/enviro-lab/benchmark-deconvolute.

2. Materials and methods

To conduct this study, 15 different controls were used to prepare 38 different control mixtures that represented equal concentrations as well as different relative abundance. These 38 mixtures were prepared in

three sample matrices: controlled sample mixtures prepared in water in isolation (water background; WB) and controlled sample mixtures spiked into RNA extracts from wastewater samples (Negative wastewater RNA background; NWRB and positive wastewater RNA background; PWRB). RNA extracts from both SARS-CoV-2 negative and SARS-CoV-2 positive samples were used as the matrix for spike-ins for two sets of mixtures. These three sets of mixtures were then sequenced with two different primer sets on an ONT platform. Finally, the sequencing data generated from these analyses were used to evaluate performance of the 8 different deconvolution tools mentioned in Table 1. Supplemental Fig. 1 graphically outlines the overall procedures used in this manuscript.

2.1. Wastewater background samples

Wastewater samples used as background for the constructed control mixtures were collected in April 2023 as a part of the SARS-CoV-2 campus monitoring program at the University of North Carolina Charlotte conducted by the Environmental Monitoring Lab (Gibas et al., 2021).

RNA concentration and extraction were performed using the KingFisher Flex with Nanotrap Microbiome A Particles, Nanotrap Enhancement Reagent, and the MagMAX Microbiome Ultra Nucleic Acid Isolation Kit. Nuclease-free water (NFW) was used as a negative control during extraction and Phosphate buffered Saline (PBS) was concentrated and extracted along with the samples as a processing control. The automated Nanotrap® Microbiome A protocol provided by Ceres Nanosciences (2023) was used without modification.

Extracted RNA was tested using droplet digital PCR (ddPCR) to quantitate SARS-CoV-2 viral RNA. Samples were tested using the CDC N2 primer and probe set with the BioRad one-step RT ddPCR Advanced kit (1-step RT-ddPCR advanced kit for probes, 2024). A positive control and NFW negative control were included in the assay. If out of 10,000 droplets (minimum number of droplets required) at least 3 (minimum positive droplets) were positive, the sample was then considered as positive. All the results were analyzed with the QuantaSoft™ Software, Regulatory Edition #1864011 (QuantaSoft™ software, 2024). Positive samples were used as positive wastewater background mixtures and negative samples were used for the negative wastewater background mixtures. The average concentration of SARS-CoV-2 in the positive wastewater background samples was 2.23 copies/μl.

2.2. Assay-ready synthetic mixed control preparation

Assay-ready synthetic RNA controls representing various SARS-CoV-2 major variants were sourced from Twist BioSciences (San Francisco, CA). These controls were used in known concentrations and in a variety of combinations of variants and mixture complexities to assay the ability of software to deconvolute more complex mixtures. 15 different synthetic controls were used - Control 15 (Alpha-103909), Control 17 (Gamma-104044), Control 23 (Delta-104533), Control 48 (Omicron - BA.1 lineage-105,204), Control 51 (B.1.1.529 + BA.2-England-105346), Control 2 (Wuhan-hu-1 from China-102,024), Control 6 (Wuhan-hu-1 from California-102918), Control 50 (B.1.1.529 + BA.2-Australia-105,345), Control 64 (BA.5-England-106,196), Control 62 (B.2.12.1-Denmark-105865), Control 66 (BA.4-Texas-106198), Control 67 (BA.4-California-106199), Control 63 (B.2.12.1-USA-105857), Control 65 (BA.5-USA-106197) and Control 19 (Iota-104529). From these 15 controls, 38 different mixtures were prepared, as outlined in Supplemental Table 1. Each control was assigned with a letter/symbol for shorthand designation: Control 15 - A, Control 17 - G, Control 23 - D, Control 48-O1, Control 51 —O2, Control 2 - Ø, Control 6 —O, Control 50 - O2, Control 64 —O5, Control 62 - O3, Control 66 —O4, Control 67 —O4, Control 63 - O3, Control 65 - O5, Control 19 - I.

2.3. Spike-in mixed control preparation

Control mixtures were spiked into three backgrounds, a water background (WB), and RNA extracts from SARS-CoV-2 positive (PWRB) and SARS-CoV-2 negative wastewater samples (NWRB), to provide a more realistic nucleic acid matrix for subsequent amplification and sequencing steps. The same ratio of synthetic controls used in the control mixtures in the water background was also used with the two RNA backgrounds to allow for comparison between all three backgrounds.

2.4. Sequencing control mixtures

Whole genome sequencing based on tiled amplicon amplification was performed with two different primer sets: 1) ARTIC v4.1 primer, which generates 400 bp long amplicons, and 2) VarSkip short v2a, which generates 550 bp long amplicons. For both of the reactions, the NEBNext® ARTIC SARS-CoV-2 Companion Kit (New England Biolabs) was used. ARTIC v4.1 primers from IDT in a 1:100 dilution were used in place of the ARTIC v3 primers that are contained in the NEBNext kit. For

Table 1
Comparison of reference reconstruction, algorithm, reference set source and input for tested deconvolution Tools.

Tool	Reference Reconstruction	Algorithm	Reference set source(s)	Input	Source
Alcov	Mutation based	Optimization	outbreak.info + covvariants.org	bam	https://www.medrxiv.org/content/10.1101/2021.06.03.21258306v1
Freyja	Mutation based	Depth weighted least absolute regression	cov-lineage	bam, reference fasta	https://www.nature.com/articles/s41586-022-05049-6
kallisto	Reference based	Pseudoalignment of hashed k-mer	GISAID or UShER (UCSC)	fastq, reference fasta	https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-02805-9
LCS	Mutation based	Statistical regression Problem	GISAID or WHO	fastq	https://academic.oup.com/bioinformatics/article/38/7/1809/6519145?login=true
lineagespot	Mutation based	Calculation of allele frequency of unique and AA mutations	outbreak.info	vcf	https://www.nature.com/articles/s41598-022-06625-6
LolliPop	Mutation based	Least square problem	cov-spectrum.org or covariants.org or PHE Genomic's Standardized Variant Definitions	bam	https://www.medrxiv.org/content/10.1101/2022.11.02.22281825v1
VaQuERo	Mutation based	SIMPLEX regression	GISAID and ECDC	bam or vcf	https://www.nature.com/articles/s41587-022-01387-y
Kraken 2	Reference based	K-mer match to LCA taxa	GISAID	fastq	https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1891-0

ARTIC, the protocol was followed as outlined in the NEBNext ARTIC instruction manual (E7660) and for VarSkip the protocol was followed as outlined by Ramachandran et al. (Ramachandran, 2022). For barcoding, the Native Barcode Expansion Kit (Exp-NBD196) from ONT was used. Samples were sequenced on the Oxford Nanopore PromethION using R9 flow cells.

2.5. Sequence trimming and filtering

Samples were analyzed via an in-house covid-analysis pipeline (covid-analysis, n.d.) which trims and filters reads using artic guppyplex basecalling, followed by filtering with Kraken 2, and Porechop to remove reads of human origin and technical sequence (fieldbioinformatics, n.d.). Samples were also analyzed with the C-WAP pipeline (C-WAP, n.d.-b) which was used during the COVID-19 pandemic for analysis of sequencing data from wastewater samples by the CDC national wastewater network. For the ARTIC pipeline (artic guppyplex), the sequence length cutoff range, which reflects the range of expected amplicon sizes, was 305–505 bp on sequences generated using the ARTIC 4.1 primers and for VarSkip, it was 475–675 bp. The BAM, FASTQ, or VCF files generated by the covid-analysis pipeline were then used as input for all subsequent analyses, depending on the input data requirements of each deconvolution tool.

2.6. Deconvolution of mixtures

Freyja (Freyja, n.d.-b), kallisto (Bray et al., 2016), LCS (Valieris et al., 2022) and Kraken 2/Bracken (Wood et al., 2019) were selected due to their use in the C-WAP pipeline and LolliPop (Dreifuss et al., 2022) was selected for its use in V-pipe. In addition to these, lineagespot (Pechlivanis et al., 2022), Alcov (Ellmen et al., 2021), and VaQuERo (Amman et al., 2022) were selected to look at a sampling of approaches to deconvolution that had not been selected for use in C-WAP. The project github repository linked in Supplemental Materials 1 provides analysis scripts used in the project along with a complete description of parameter and option choices used with each tool.

2.7. Standardization of outputs

Each tool compared in this analysis had its own output format. To make our outputs comparable, they have been converted to the same output format used by Freyja. Pangolin lineages provided by each tool were summarized in the manner derived from the way Freyja summarizes lineages, but with a few categories adjusted to best illustrate the proportions of relevant lineages and sublineages; most notably, BA.1, BA.2, BA.4, and BA.5 and their sublineages were grouped separately rather than remaining within a single, broad Omicron category.

2.8. Statistical analysis

Various statistical tests were performed to discern the reliability of our claims, always with a threshold of $p > 0.01$ for significance. We analyzed the variations in outcomes based on the deconvolution tool used, the expected abundance of lineages contained within each sample, the background matrix of a sample, and the primer scheme used during sequencing. Each comparison of this sort was conducted via analysis of variance (ANOVA), or *t*-test when comparing just two categories. ANOVAs with significant values were followed up with Tukey's post hoc Honestly Significant Difference (HSD) test to determine which individual groups varied significantly from one another. Means, standard deviations, and relevant *p*-values are presented and discussed near each statistical analysis.

2.9. NCDHHS sample collection

In Section 3.7 of Results and discussion, we discuss the continuing emergence of variants as a challenge for all types of deconvolution methods based on fixed strain definitions. We base this observation on statewide wastewater sequencing data that we produced for the NC Wastewater Monitoring Network between June 2022 and June 2023. Wastewater samples were collected twice per week from 20 municipal wastewater treatment plants in North Carolina: Beaufort, Msd of Buncombe County, Sugar Creek in Charlotte, Mallard Creek in Charlotte, McDowell Creek in Charlotte, and Charlotte 4 (a subwatershed location on the Greenway), Fayetteville Rockfish, Greensboro North Buffalo, Greenville, Jacksonville, Laurinburg, Marion, Raleigh (Neuse River Resource Recovery Facility), Owasa, Roanoke Rapids, South Durham, Tuckaseigee, City of Wilson, Wilmington North Side City, Wilmington North Side Cunty and Winston Salem by the WWTP's staff. These were then shipped to UNC-Chapel Hill Noble laboratory. These wastewater samples were concentrated using HA filtration as described in Katayama et al. (2002) (Katayama et al., 2002). The samples were then extracted on the KingFisher™ Flex automated magnetic particle analyzer (Thermo Fisher Scientific, Waltham, MA) using the easyMag® NucliSENSE® reagents (bioMerieux, Durham, NC) in the Noble lab. After performing quantification by RT-ddPCR, UNC Chapel Hill shipped an additional concentration filter to UNC Charlotte on dry ice for sequencing using the methods described above.

3. Results and discussion

Deconvolution of SARS-CoV-2 lineage abundance from wastewater extract sequencing data can be accomplished using a variety of algorithms (Table 1). The deconvolution methods assessed in this paper can be grouped depending upon several methodological choices including whether classification is based on recruitment of reads to a reference genome or on a match of detected SNVs to a pattern of defining mutations, the source of the reference genome or reference lineage-defining mutation information used for classification, and the algorithmic approach used. There are also operational differences in each computational workflow, especially as to whether the input required is the raw FASTQ file, a pre-aligned BAM or SAM file, or a VCF file extracted from a read alignment (Table 1). Starting with a fastq vs a bam does not impact the bioinformatics analysis, other than flexibility in downstream analysis and computational efficiency. However, starting analysis from a vcf file means that there is no access to variant frequency information or total number of reads covering the site, and only uses the summarized call of that variant. Definition of variants by each method is with respect to the original Wuhan strain of SARS-CoV-2, except for kallisto, which uses recruitment of reads to a collection of reference strain genomes. Bioinformatics issues specific to the read recruitment workflow are discussed separately in Section 3.7.

3.1. Sequencing outcomes across background type and primer sets

To assess the impact of the wastewater extract matrix itself on sequencing outcomes, we compared sequencing results generated for control mixtures spiked into water background (WB), SARS-CoV-2 negative wastewater RNA extract background (NWRB) and SARS-CoV-2 positive wastewater RNA extract background (PWRB).

We examine the impact of the tiling primer scheme using two of these: the ARTIC v4.1 and VarSkip short v2a primer sets, both of which continue to be widely used. The tiling primer scheme determines the expected amplicon size, as well as which sections of the genome are amplified during the sequencing process, impacting the diversity and coverage of sequences obtained (Lin et al., 2021). To evaluate the potential effects of primer choice on sequencing outcomes, we replicated our experiments using two different tiling primer sets. To assess sequencing performance of the different primer schemes, we compared the mean coverage depth and mean breadth of coverage generated from each sample when aggregating across both the whole genome and for the S gene alone.

3.1.1. Mean depth of coverage, whole genome

For samples sequenced with ARTIC primers, the mean coverage depth values across the whole genome for the three backgrounds differed significantly (WB: mean = 9133.47, std. dev. = 1518.10, NWRB: mean = 5036.28, std. dev. = 595.27, and PWRB: mean = 9579.87, std. dev. = 1914.78), as determined by one-way ANOVA (F = 113.03, p = 1.69e-27). This is visualized in Supplemental Fig. 2a. Similarly, with VarSkip primers, mean coverage across the whole genome also differed significantly (WB: mean = 11,814.91, std. dev. = 1329.63, NWRB: mean = 9485.823, std. dev. = 1355.15, and PWRB: mean = 13,330.34, std. dev. = 995.65), as determined by one-way ANOVA (F = 93.03, p = 1.87e-24).

Supplemental Fig. 2b shows the same analysis but with mean coverage depth compared only over the S gene, in which many lineage defining mutations are located. As with the whole genome, the mean coverage over the S gene for the ARTIC primers was significantly different across WB (mean = 11,085.81, std. dev. = 1673.67), NWRB (mean = 5985.21, std. dev. = 539.62), and PWRB (mean = 11,752.88, std. dev. = 1846.24), as determined by one-way ANOVA (F = 174.56, p = 5.32e-35). The mean coverage over just the S gene for VarSkip is similarly significant across WB (mean = 9994.74, std. dev. = 1444.92), NWRB (mean = 8144.81, std. dev. = 1756.88), and PWRB (mean = 12,964.56, std. dev. = 1774.93), as determined by one-way ANOVA (F = 80.96, p = 2.07e-22). In both the whole genome and the S gene, Tukey's HSD test for the ARTIC primers showed significant differences between NWRB and WB and between NWRB and PWRB ($p < 0.01$ for each), but the mean coverage values for WB samples did not vary significantly from PWRB ($p = 0.38$ for whole genome (Supplemental Fig. 2a), $p = 0.12$ for S gene (Supplemental Fig. 2b). For the VarSkip primers, Tukey's HSD showed that all three backgrounds differed with the others ($p < 0.01$ for all). The VarSkip WB and PWRB mixtures had significantly different mean coverage depths from each other differing from sequences produced from the ARTIC primer scheme.

3.1.2. Mean depth of coverage, primer scheme comparison

To understand how the mean coverage depth differs from one primer scheme to another, a t-test was conducted which showed significant differences in mean coverage depth, t(226) = −5.765, p = 2.656e-8, ARTIC (mean = 9607.96, std. dev. = 2970.47), and VarSkip (mean = 11,543.69, std. dev. = 2006.75). The above values are those calculated over the whole genome, but the results when looking only at the S gene corroborate these, t(226) = −7.262, p = 6.102e-12, ARTIC (mean = 7916.54, std. dev. = 2507.84), and VarSkip (mean = 10,368.04, std. dev. = 2588.83).

3.1.3. Mean breadth of coverage

To fully explore the coverage level of all our samples, we also analyzed variance in breadth of coverage by calculating the fraction of the whole genome (Supplemental Fig. 2c) and the S gene alone (Supplemental Fig. 2d) covered at 100× for both the ARTIC and VarSkip datasets. For the whole genome amplified by ARTIC, the 100× ratio was statistically different across WB (mean = 0.942, std. dev. = 0.014), NWRB (mean = 0.925, std. dev. = 0.021), and PWRB (mean = 0.938, std. dev. = 0.018), as determined by one-way ANOVA (F = 9.872, p = 1.057e-4). VarSkip amplified samples show no background as having significantly different mean breadth of coverage over both the whole genome regardless of background across WB (mean = 0.922, std. dev. = 0.016), NWRB (mean = 0.906, std. dev. = 0.049), and PWRB (mean = 0.924, std. dev. = 0.024), as determined by one-way ANOVA (F = 3.72, p = 0.027). For the ARTIC dataset, Tukey's HSD, WB and PWRB lack statistically significant differences ($p = 0.48$), but NWRB differs significantly from the other two backgrounds ($p < 0.01$) for both. While these are statistically significant, they are not as significant as we observed for mean coverage depth.

The variance in breadth of coverage across the S gene for the ARTIC primers agrees with that of the whole genome (ANOVA: F = 35.47, p = 6.86e-13; WB (mean = 0.942, std. dev. = 0.015), NWRB (mean = 0.908, std. dev. = 0.033), and PWRB (mean = 0.948, std. dev. = 0.018)) with similar Tukey's HSD results ($p < 0.01$ for NWRB vs WB and NWRB vs PWRB; $p = 0.52$ for WB vs PWRB). For the VarSkip primer scheme across the S gene alone we do not observe a statistical significance (ANOVA: F = 2.96, $p = 0.055$; WB (mean = 0.874, std. dev. = 0.062), NWRB (mean = 0.835, std. dev. = 0.116), and PWRB (mean = 0.875, std. dev. = 0.070)). These values are not significant enough of a difference for Tukey's HSD in VarSkip. This suggests that the wastewater background itself might have a slight influence on breadth of coverage, but it is not significant enough to be seen in both the negative and positive wastewater backgrounds.

3.1.4. Mean breadth of coverage, primer scheme comparison

To compare the breadth of coverage between the ARTIC and VarSkip primer schemes, we conducted a t-test incorporating all three backgrounds, t(250) = 5.13, p = 5.81, ARTIC (mean = 0.935, std. dev. = 0.019), and VarSkip (mean = 0.917, std. dev. = 0.034) and found that the breadth of coverage is not statistically significant between the primer schemes. Given these results and the preponderance of the ARTIC primer scheme in the literature, we focused our analysis of the deconvolution tools on the ARTIC v4.1 derived dataset.

3.2. There are significant differences in variant identification between deconvolution tools

To assess the performance of the chosen deconvolution methods, we first compared the ability of each of the methods to detect the strains included in the original sample and in the expected proportions. Median pairwise L2 abundance norms, calculated using the approach described in Ye et al. (Ye et al., 2019) provides a summary of the accuracy of abundance estimation across methods, offering a comparison between expected vs estimated abundance and also among the tools themselves (Fig. 1). For most comparisons, the tools were not statistically significantly different as determine by one-way ANOVA (threshold: $p < 0.01$) and Tukey's post hoc HSD Test. However, Kraken 2 (C-WAP) was statistically significantly different from all other predictions. Kallisto, similarly, was different for all predictions except lineagespot. Lineagespot differed significantly from all other predictions except kallisto and LCS. In Supplemental Table 1, we show the expected and predicted composition for each sample mixture. All of the deconvolution methods tested in this study could identify the major lineages that were present in the control mixtures, except for lineagespot and C-WAP's implementation of Kraken 2 (Supplemental Fig. 3). Similar results are found for the VarSkip dataset as show in Supplemental Fig. 4.

While most of the methods identify the variant in the mixed control dataset at the lineage level, there are misidentifications of many variants at the sub-lineage level. Most of these tools use a mutation-based reference set to define lineages and sublineages. As SARS-CoV-2 is a pathogen with a high mutation rate, the presence or absence of a single significant lineage-defining mutation can cause misidentification of sublineages. For example, BA.4 and BA.5 share >50 similar mutations (Tallei et al., 2023) and BA.1 and BA.2 share 29 similar mutations (Kumar et al., 2022). At this degree of similarity, a single missing lineage-defining SNP has the potential to result in a misidentification. This is compounded by the fact that lineages are being defined and classified based upon amplicons and may not be able to be easily assigned to one variant.

Misidentification was also seen in older variants such as Iota, because some of the tools did not incorporate mutations for variants in their lineage dictionary if they were no longer considered as variants of concern or interest, such as Iota. Similarly, for lineagespot, the database is truncated following delta and it does not have a reference for omicron lineages; as a result, lineagespot completely misses calling any of the Omicron lineages assigning those as "other" as seen in Supplemental Fig. 3 and in the large L2 values in Fig. 1. Most of the deconvolution tools

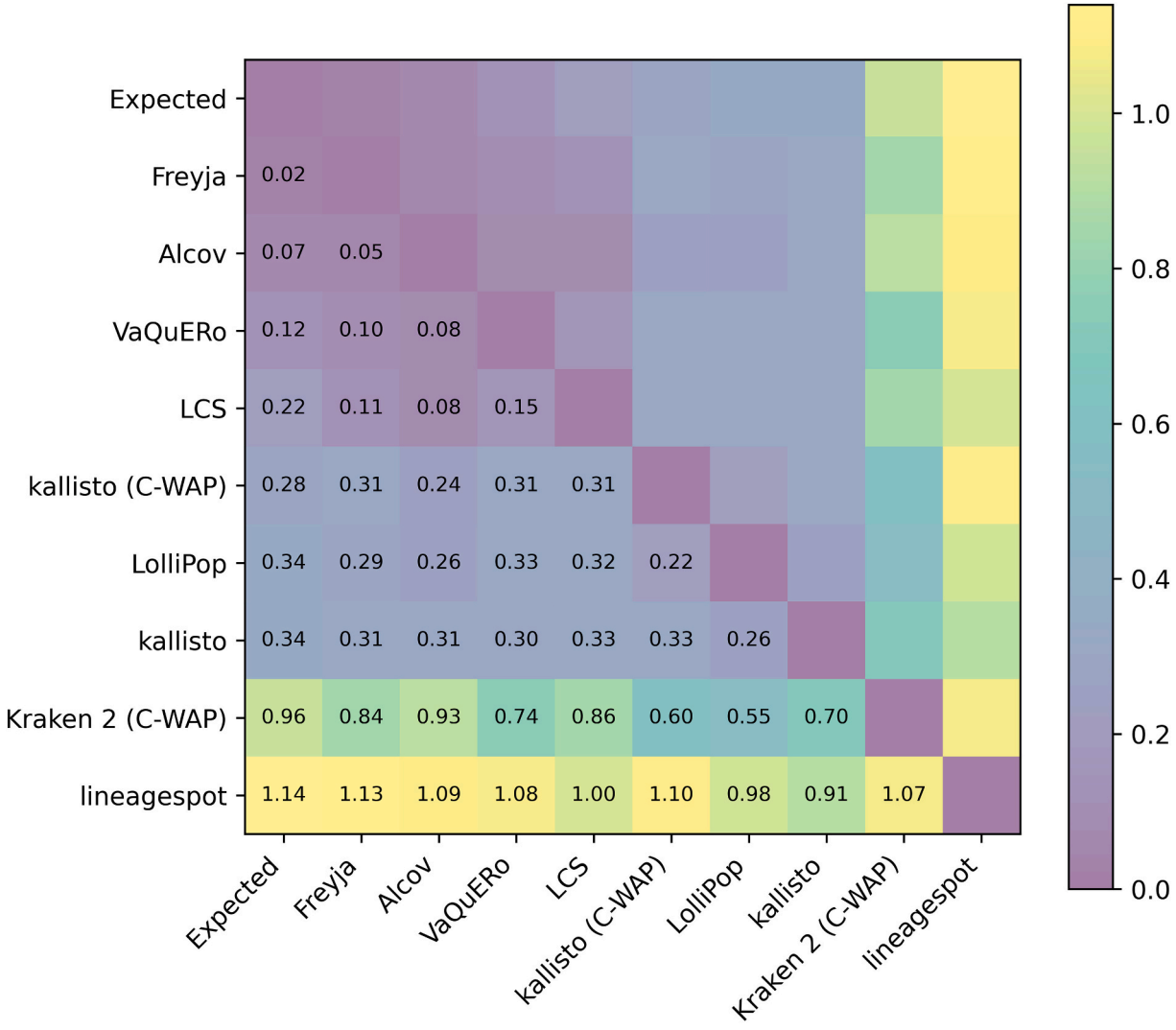


Fig. 1. Median pairwise L2 abundance norms between deconvolution tools. Each value shows the L2 distance between the clusters of estimated and expected abundance profiles for each of the tools, and estimated abundance profiles between tools as well. Lower L2 distance represents closer similarity between the X and Y axis clusters. This shows two different distance values for Kallisto. Kallisto (C-WAP) represents the estimated output generated from using the database that C-WAP pipeline uses and Kallisto represents the estimated output generated from using the in-house database. This same figure but for the VarSkip datasets can be seen in Supplemental Fig. 4.

also did not explicitly identify the original Wuhan-hu-1 strain when it was a component of the mixture, other than Freyja. As most of these tools define lineages based on mutations relative to the Wuhan-hu-1 strain, and there is no mutation present in Wuhan-hu-1 relative to itself, none of them label this lineage correctly when it is present in a wastewater mixture, perhaps because its occurrence is a circumstance that is now unlikely other than in archival samples or constructed standard mixtures like those created for this study. VaQuERo, if it is unable to detect any lineage or sublineage based on mutations, will identify that sample as Wuhan-hu-1 by default.

3.3. The wastewater background had only subtle effects on abundance estimation

The relative abundance of lineages changes very little regardless of whether it is spiked into negative or positive wastewater. The variances between samples from different backgrounds were insignificant ($p > 0.01$ for all) when testing the effect of sample matrix on the O/E ratio by ANOVA for each of the deconvolution tools (Fig. 2). The insignificant differences in the estimated abundance between backgrounds can be attributed to the extra complexity of a metagenomic sample and the effect of addition of lineages of SARS-CoV-2 from the positive wastewater background to the synthetic control mixtures, however, this does not negatively impact the performance of the tools. This also suggests that the deconvolution tools are not highly sensitive to the small coverage differences or sample background for relative abundance estimation. Similar results are seen in the VarSkip dataset (Supplemental Fig. 5).

For the small differences that were observed, ANOVAs were performed for each lineage to see how those differences varied depending on the water or wastewater background of a sample. These results are shown in Table 2. The lineages that preceded Omicron tended to show statistically significant differences between samples from wastewater backgrounds and water backgrounds, suggesting that sample background affects the lineage assignment of older variants more significantly than the newer ones. The addition of wastewater, which has a complex composition of inhibitors and carry through compounds, can inhibit the PCR and sequencing reactions and affect the sensitivity of variant detection, although this does not explain the insignificant differences in more recent lineages. More likely, the absence of defining mutations in the reference lists of deconvolution tools, for older/non-variant-of-interest lineages (as we have seen in Section 3.3) can cause

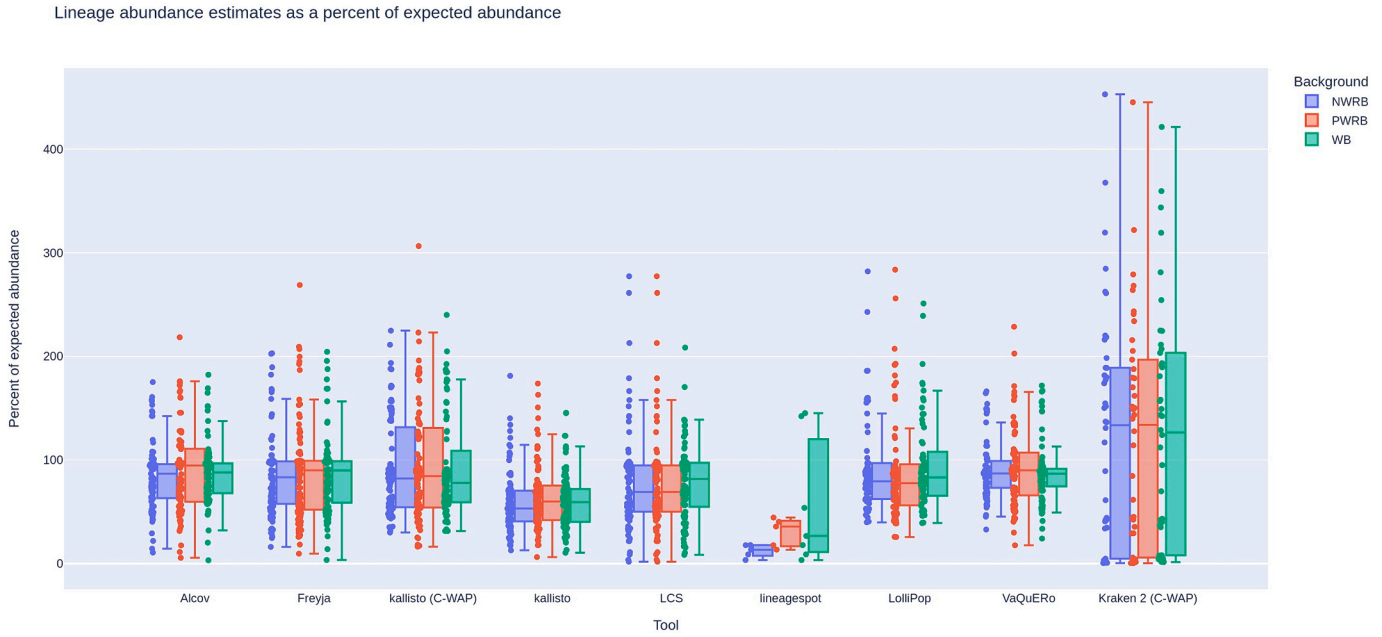


Fig. 2. Relative abundance of strains detected by each deconvolution method (with ARTIC primers) as a percentage of the expected abundance (O/E ratio = observed relative abundance/expected relative abundance) for spike-ins to nuclease free water (green), SARS-CoV-2 negative wastewater background (blue), and SARS-Cov-2 positive wastewater background (red). Data points show the actual O/E ratios of observed to expected for each observed strain in a mixture, while the box plots summarize the distribution of O/E ratio values across all strains and control mixtures. Box plots are delimited at the first and third quartiles. The related results with VarSkip primers can be seen in Supplemental Fig. 5.

Table 2
ANOVA statistic between sample matrix backgrounds (water or wastewater) and Tukey's HSD results when grouped by lineage.

Lineage	ANOVA		Tukey's HSD p-values		
	F-statistic	p-value	WB vs NWRB	WB vs PWRB	NWRB vs PWRB
Wuhan-hu-1	0.047881818	0.953278547	–	–	–
Alpha	14.2040499	1.79E-06	0	0	0.995
Gamma	9.738276664	0.000102748	0.003	0	0.701
Delta	2.996111335	0.053471462	–	–	–
Iota	23.53172099	1.23E-09	0.003	0	0.002
BA.1.X	0.025036687	0.975277084	–	–	–
BA.2.X	1.281574051	0.279354374	–	–	–
BG.X	2.418754658	0.091769427	–	–	–
BA.4.X	2.470374107	0.086528375	–	–	–
BA.5.X	13.12012881	3.74E-06	0.044	0	0.02

misidentification and misestimation.

3.4. Lineage abundance variation between deconvolution methods

To understand how the lineage abundance estimated by each method differed from the expected abundance, the estimated relative abundance in each control dataset with different backgrounds were analyzed with respect to the percent abundance of lineages (Fig. 3 and Supplemental Fig. 6a, b, c). Fig. 3 shows the observed to expected (O/E) ratio for each lineage with abundance estimates from Freyja aggregated across all three backgrounds. This O/E ratio varies significantly between linages with some underestimated (Wuhan-hu-1 and BA.1.X) and some significantly overestimated (BA.4.X). This variation carries over when considering each background separately. Supplemental Fig. 6 shows that certain tools underestimated lineage abundance more often with certain lineages. Several of the tools underestimated the abundance of the older variants, especially LCS with Delta and Iota. Among the pre-Omicron variants only, Alpha was slightly overestimated in some of the mixtures by Freyja, LCS and LolliPop. Gamma was the only variant that had similar relative abundance estimated by all the tools. Conversely, the abundance of Omicron and its sublineages are frequently overestimated by most of the deconvolution tools, especially BA.4.X in all the tools tested. Kraken 2 significantly overestimates most of the Omicron lineages. Kallisto's performance varies greatly depending on the reference database chosen; this is discussed further in Section 3.7.

When considering the effect of the wastewater background (Supplemental Fig. 6b and c), the patterns observed in the water background are for the most part mirrored in the NWRB and PWRB backgrounds continuing to support our conclusion that the matrix does not significantly affect the linage abundance outcome. This is somewhat surprising given that the positive samples were collected from April 2023 when most circulating lineages are Omicron, and we would expect an increase in Omicron concentrations in the samples.

Part of our experimental design was to vary both the number of linages in a mixture as well as the expected abundance such that some mixtures contained up to 10 lineages and the concentrations as copies/uL varied in a single sample (Supplemental Table 1). This intent of this was to more closely mimic a collected wastewater sample that would be a mix of different variants as well as different concentrations of those variants. In Supplemental Fig. 7, the distribution of the O/E ratio by the number of linages in a mixture is plotted. It clearly shows that samples with a single lineage are, for the most part, called accurately. However, as the complexity of a mixture increases, the O/E ratios at 4 lineages and greater do vary and is likely dependent on the strain being called, as discussed above. No statistically significant differences were detected as determined by one-way ANOVA. Supplemental Fig. 8 explores this same O/E ratio based upon the relative expected abundance of a control in the mixture. Again, as expected, when the mixture is composed of a single control (abundance of 1.0), there is a tight O/E ratio at 1. This is similar for expect abundances of 0.625 and 0.3125. What is interesting is that as the expected abundance decreases and a mixture becomes more complex or a sample is added in a lower relative abundance, the O/E ratio begins to vary and on average is calculating the observed abundance as lower than expected. The one-way ANOVA statistics sit at our threshold of $p =$

0.01 suggesting that differences are at most barely significant.

3.5. Freyja yielded the most accurate lineage compositions and fewest false negatives

One of the main objectives of this study was to identify which deconvolution tool most accurately deconvolutes samples of known composition, in terms of overall accuracy and false positive or false negative identifications. In short, a false positive occurs when the tool incorrectly identifies a negative instance as positive and a false negative occurs when the tool incorrectly identifies a positive instance as negative. Output of all methods tested was compared to the expected relative abundance of strains in the prepared samples. Results of each method were also compared to one another. We found that Freyja outperformed other tools in both accuracy and fraction of false negatives.

The median pairwise L2 abundance norms calculated in Fig. 1, are generated from the distances between expected control vector and deconvolution tool datasets, and show that Freyja has the lowest distance (0.02) from the expected vector, indicating that Freyja has the highest accuracy in identifying the relative abundance of lineages. Supplemental Fig. 4 shows Freyja's detection of lineage composition in comparison to other deconvolution tools when the VarSkip primer set is used. Similarly, the lowest distance is also Freyja to expected with 0.03. When considering false positives, Freyja consistently identified the lineages with minimal false positives and zero false negatives (Figs. 4 and 5). With highest accuracy, no false negatives and few false positives, Freyja most accurately replicates the lineage identification and abundance of expected control mixtures. Fig. 4 compares the detected and undetected lineages and for all tools using the ARTIC primer dataset, and the VarSkip results are in Supplemental Fig. 9. The expanded lineage detection composition as outlined in Fig. 5 for the other tools tested can be found in Supplemental Fig. 10.

When comparing the effectiveness of each tool's lineage detection (Fig. 4), VaQuERo most closely follows Freyja in performance. The rest of the tools detected multiple false positives, lineages which were not present in the control mixture dataset and sometimes also failed to detect the spiked-in lineages correctly. As false positive identification presents a major challenge in wastewater lineage deconvolution, the relatively low false positive calls in Freyja and VaQuERo carry significant weight in tool choice. This can be seen in Fig. 5 and Supplemental Fig. 10 where false positives appear below the X-axis. Within Freyja, omicron lineages are classified into either sublineages BA.1, BA.2, BA.4 or BA.5 or a more generalized category of "Omicron", but not a specific sublineage. With SARS-CoV-2, the variations that separate a VOC from a less concerning strain may consist of just one or two changes. False positive detection of a VOC, especially during times of increasing incidence, may result in unnecessary or confusing announcements from public health authorities. Perceived false alarms may increase public skepticism of public health data and measures. When we consider how each algorithm works, Freyja uses each of the mutations from the alignment files and their respective sequencing depth to perform the depth weighted least absolute deviation regression. Similarly, VaQuERo also uses all unique and non-unique marker mutations to solve the deconvolution problem using SIMPLEX regression approach. Given the similar approaches, it is not surprising that they both perform well.

3.6. Kallisto performance was strongly dependent on the reference database

Kallisto is originally an RNA transcript quantification tool. As the variant abundance estimation is found to be computationally similar to RNA transcript abundance estimation (Baaijens et al., 2021), it has been repurposed to be used as a wastewater deconvolution tool for relative abundance estimation of the SARS-CoV-2 variants. Kallisto takes a set of reference sequences that contains multiple genomic sequences per lineage, with the recommendation to include multiple sequences per lineages to reduce biases relating to within-lineage variation (Baaijens et al., 2021). As a consequence, the composition of the reference set can significantly affect the performance of a reference-based deconvolution tool (Aßmann et al., 2023). We developed two databases to test how the performance of kallisto varies with respect to the database used.

The in-house database preparation is described in the configuration

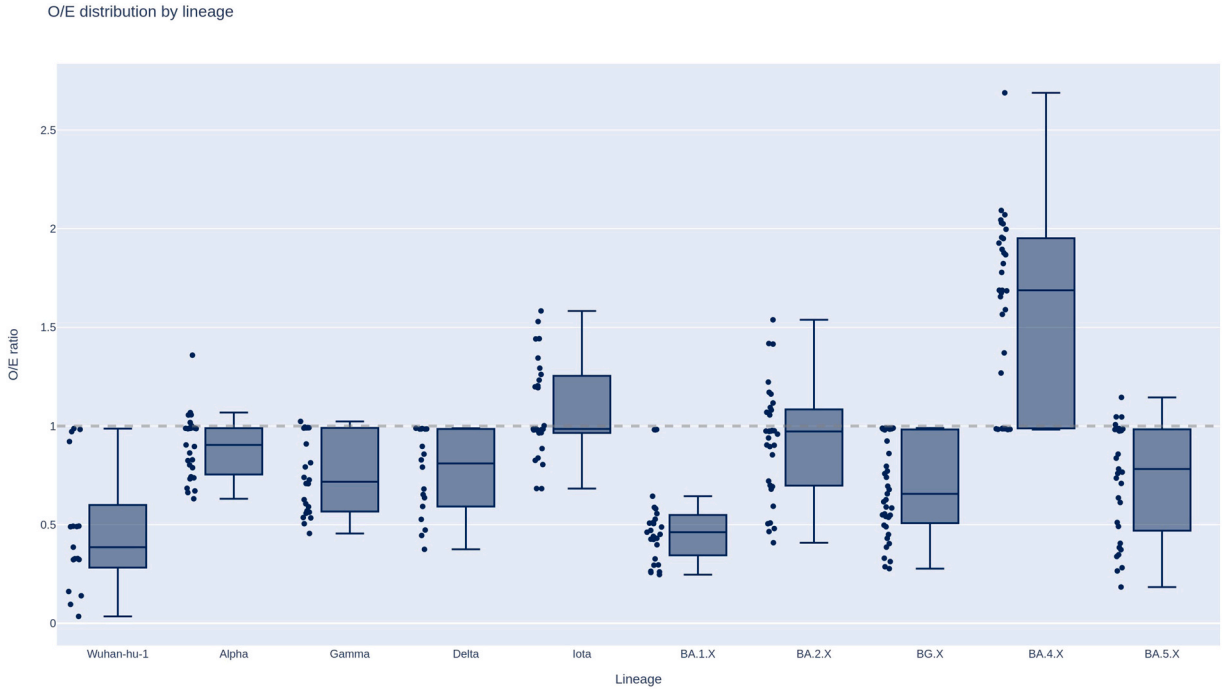


Fig. 3. O/E ratio distribution for each lineage with abundance estimates from Freyja aggregated across all three backgrounds. *p*-value: 1.75e-45 f-value: 40.94. The O/E was significantly different across lineages, as determined by one-way ANOVA ($F = 40.94$, $p = 1.75e\text{-}45 < 0.01$). O/E ratios are broken down similarly in Supplemental Fig. 6 for all nine deconvolution tools and all three backgrounds separately.

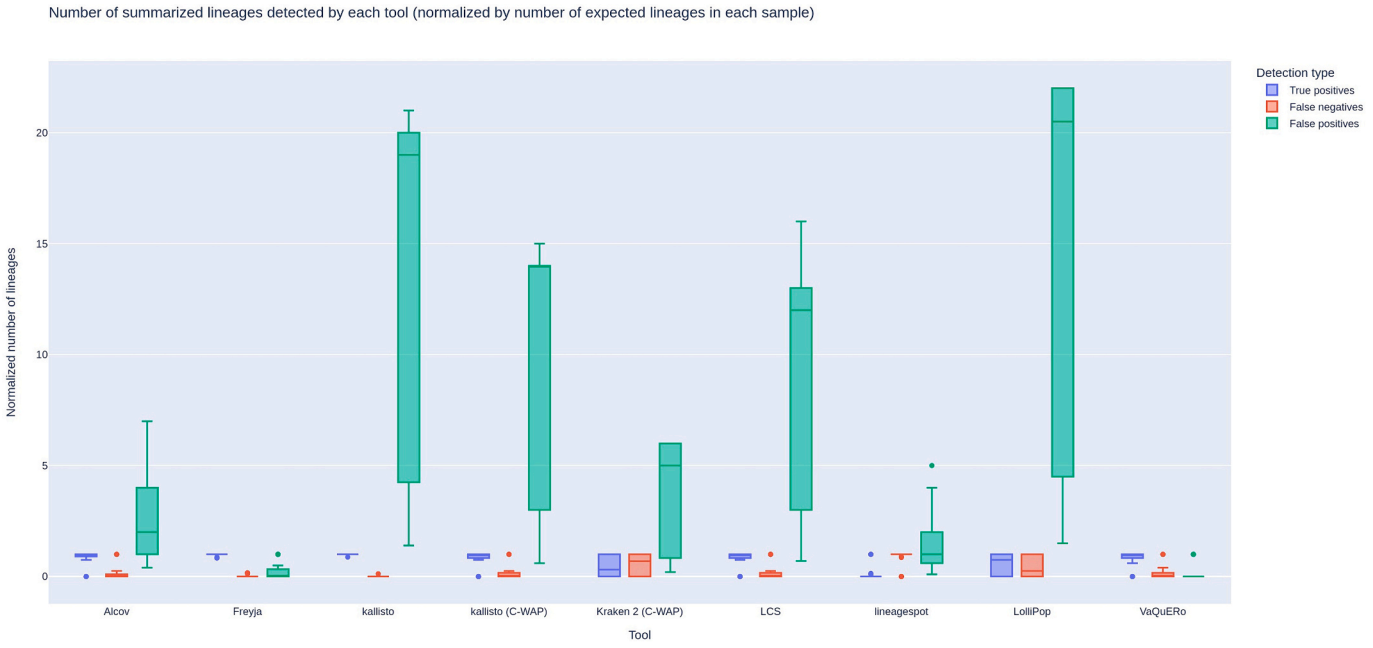


Fig. 4. Distribution of the number of lineages each tool detected in each mixture categorized by true positives and false positives as well as lineages that went undetected despite being present in the mixture (false negatives). All counts were normalized to the number of lineages that should have been present in any given mixture. Only WB mixtures were included in this plot. This plot with VarSkip data can be seen in Supplemental Fig. 9.

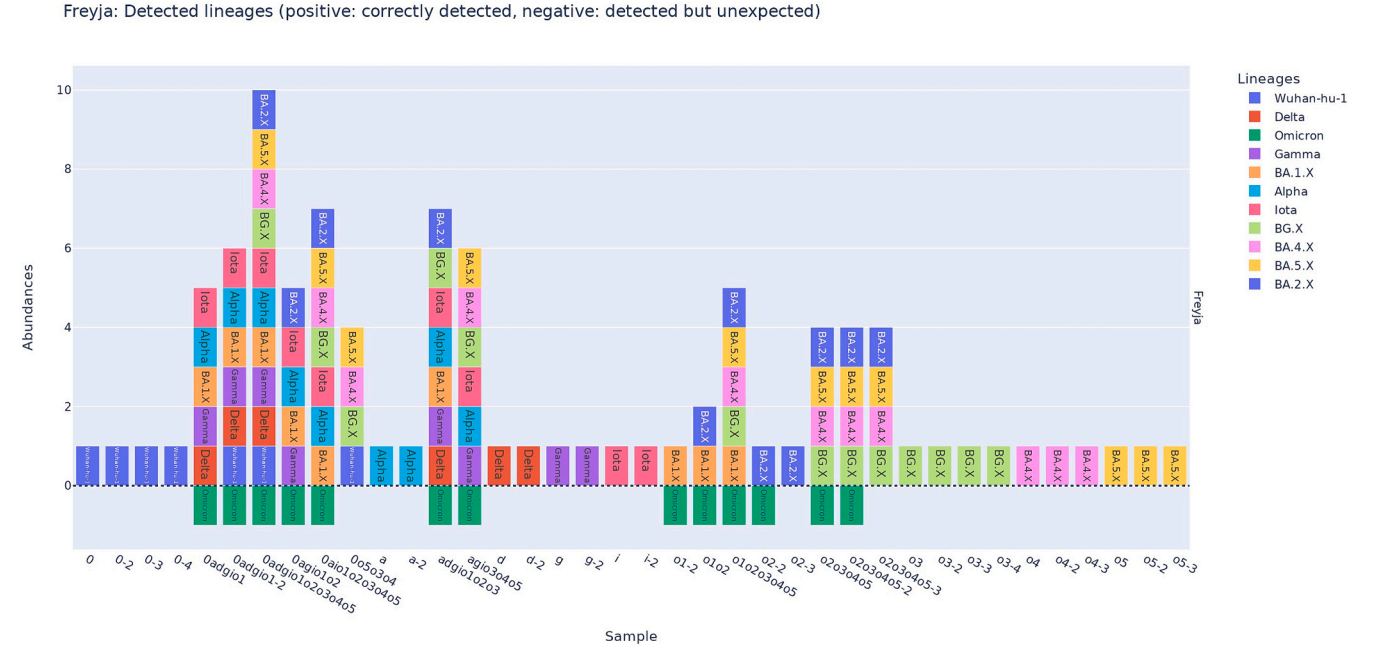


Fig. 5. Detected lineage composition by Freyja (WB mixtures only). Lineages in the positive Y direction are true positives that should have been present and were correctly detected. Lineages in the negative direction are those that should not have been present but the deconvolution tools detected.

parameters (Supplemental Materials 1) and includes 406 sequences with up to four sequences per lineage for 299 different lineages. On the other hand, the C-WAP databases for both kallisto and Kraken 2 used 30 sequences with up to six sequences for each of the following 13 included lineages: Wuhan-hu-1, Alpha, Beta, Gamma, Eta, Epsilon, Delta, Iota, Kappa, BA.1, BA.1.1, BA.2, and BA.3 (Kayikcioglu et al., 2023). Those databases were produced one time, receiving no further updates over time. The in-house database used lineages that were more specific than those in the C-WAP database, including many sublineages of the lineages listed for C-WAP's database. Fig. 1 shows that the accuracy of kallisto varies between these reference databases. Kallisto run with the C-WAP database has the median distance of 0.28 with respect to the expected abundances, and kallisto with the in-house database has a distance of 0.35. This means that the accuracy of kallisto is improved when using the C-WAP database. Similarly, more false positive lineages were identified with the in-house database (Fig. 4). We hypothesize that because the in-house database has significantly more reference sequences per lineage, the algorithm may have had difficulty with unambiguous read assignment and aggregation into lineages. Therefore, the goal of correctly interpreting within-strain variation may be at odds with the

goal of unambiguous read mapping, and the reference database should be chosen with care.

3.7. Interpretation of increasingly complex variant profiles is a future challenge

The reference sequences necessary for deconvoluting wastewater mixtures mostly rely on the availability of SARS-CoV-2 clinical sequencing samples. As PCR-based lab testing has declined in favor of at-home rapid testing, access to clinical specimens has decreased significantly. As fewer clinical sequences are being surveyed, the availability of reference sequences for current circulating variants will similarly decrease. Theoretically, this absence of reference sequence will cause a drift of current sequences away from the reference data, which will at some point start to impact the performance of deconvolution tools in identifying lineages. If this is indeed happening, we should see an increase in the abundance of unidentified or unclassified lineages in wastewater across time.

As part of our research program, we have been sequencing wastewater samples across North Carolina using the ARTIC primer set and Freyja for deconvolution. If we examine data from September 2022 to July 2023 (Fig. 6), we can see that the trendline shows abundance of "unclassified by Freyja" and "below abundance threshold" lineages increasing over time. In diverse metagenomic datasets, an abundance threshold is often applied to filter out very low abundance lineages and simplify interpretation and visualizations. The fraction of circulating SARS-CoV-2 lineages in NC-DHHS wastewater samples that fall below such an abundance threshold is continuously increasing (Fig. 6). As subvariants of major lineages proliferate, without a threshold the visualization and analysis of wastewater sequence signal will lose its clarity, making it more difficult to interpret current trends. Grouping of identified strains into major lineages can also clarify interpretation of complicated mixtures, but it is also a challenge for strain-reference-based deconvolution algorithms to correctly identify subvariants with accuracy. We can see in Section 3.6 that the most accurate output-generating tool Freyja is overcalling generic Omicron signatures in our most complex standard mixtures, where multiple controls are present in varying proportions starting from 625 to 2500 copies/µl. We hypothesize that as variants continue to diversify, under sampling of the most recently emerged variants needs to be taken into account during data analysis to improve algorithm performance. These trends also suggest that methods of wastewater sequence analysis that do not focus on deconvolution of variants to existing reference genomes and variant patterns will be required going forward. The gradually increasing rate of identification failures observed even with Freyja, the most accurate of the deconvolution tools we studied, supports the concern that reference-based identification of variants will be increasingly challenging as time goes by.

4. Conclusions and future work

This study addresses the ability of commonly used deconvolution methods to distinguish the presence and abundance of SARS-CoV-2 variant spike ins when controlled mixtures are sequenced. We find that Freyja, which has been widely adopted throughout the COVID-19 pandemic, produces variant abundance calls with the closest relationship to the expected ratios when tested on controlled variant spike-in mixtures. The error profile produced by Freyja is dominated by false positive lineage identifications rather than false negatives, including the identification of indeterminate omicron lineages that do not correspond to any particular control, which are seen mainly in complex mixtures containing multiple spiked in variants. We also examine the influence of the wastewater background on deconvolution outcomes. We find that the impact of the wastewater matrix on variant deconvolution is insignificant, despite significant differences in sequence coverage resulting from the wastewater matrix. The influence of different tiling amplicon primer schemes on deconvolution outcomes is also negligible. The impact of the concentration and extraction process on viral RNA detection and quantitation has been extensively studied. The impact of that process on wastewater variant sequencing could straightforwardly be investigated by repeating this study beginning from encapsulated viral controls spiked into raw wastewater, but that is a separate question, less about bioinformatics method performance than about sample processing and chemistry. We also found that without regularly updated clinical reference data, the amount of sequence classified as unknown is gradually increasing. The virus continues to diversify, and the practice of classifying wastewater sequence into strain assignments may prove untenable in the long run. Approaches that have the additional capability to focus on the changing abundance of SNVs at key sites and identify emerging mutation clusters, like VaQueRo, are one potential way forward. It would likely be interesting to repeat this study on the commonly-used Illumina sequencing platform as well. However, given that we observe minimal differences in variant identification due to either tiling amplification scheme or exposure of RNA to the wastewater extract background, and the ongoing improvements to basecalling error rates on the Oxford Nanopore platform, the influence of the sequencing platform on variant deconvolution is likely to be relatively subtle as well.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2024.174515.

CRediT authorship contribution statement

Jannatul Ferdous: Writing – original draft, Methodology,

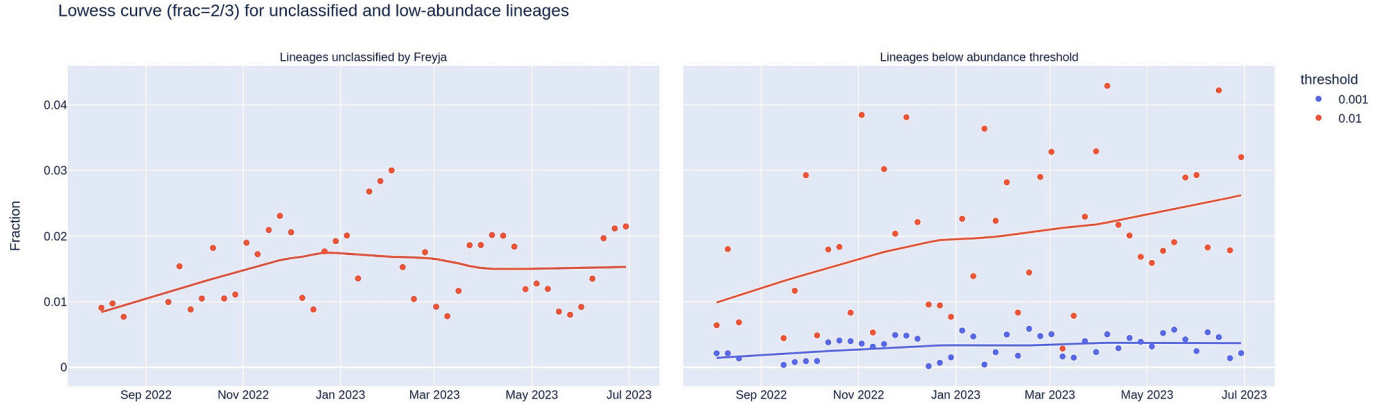


Fig. 6. LOWESS trendline showing how the fraction of unclassified lineages (left) and the fraction of lineages below two threshold abundance values (right) changed over time in wastewater surveillance sequencing data collected by NCDHHS.

Investigation, Formal analysis, Data curation. **Samuel Kunkleman:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **William Taylor:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **April Harris:** Methodology, Investigation, Data curation. **Cynthia J. Gibas:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Jessica A. Schlueter:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data is under BioProject PRJNA1031245, the protocol at doi.org/10.17504/protocols.io.261ged2jjv47/v1, and scripts and methods at https://github.com/enviro-lab/benchmark-deconvolute

Acknowledgements

We gratefully acknowledge all data contributors, i.e. the authors and their originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. The authors of a few tools were most helpful when asked for clarification on how to use their tools or adapt them to the constraints of this study. We would like to thank Fabian Amman and his team from VaQuERo, Art Poon from Gromstole, and Ivan Topolsky from LolliPop for their assistance. We would also like to acknowledge BioRender, which was used to generate the Graphical abstract and Supplemental Fig. 1. Funding for this project was provided by the North Carolina Department of Health and Human Services.

References

1-step RT-ddPCR advanced kit for probes, 2024. Bio-Rad Laboratories [Internet]. cited 7 Feb. Available: https://www.bio-rad.com/en-us/product/-step-rt-ddpcr-advanced-kit-for-probes?ID=NTGCRI15.

Aleem, A., Ab, A.S., Slenker, A.K., 2024. Emerging Variants of SARS-CoV-2 and Novel Therapeutics Against Coronavirus (COVID-19). cited 24 Jan. Available: https://europepmc.org/article/nbk/nbk570580.

Amman, F., Markt, R., Endler, L., Hupfauf, S., Agerer, B., Schedl, A., et al., 2022. Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. Nat. Biotechnol. 40, 1814–1822.

Aßmann, E., Agrawal, S., Orschler, L., Böttcher, S., Lackner, S., Hölzer, M., 2023. Impact of reference design on estimating SARS-CoV-2 lineage abundances from wastewater sequencing data. bioRxiv. https://doi.org/10.1101/2023.06.02.543047, p. 2023.06.02.543047.

Baaijens, J.A., Zulli, A., Ott, I.M., Petrone, M.E., Alpert, T., Fauver, J.R., et al., 2021. Variant abundance estimation for SARS-CoV-2 in wastewater using RNA-Seq quantification. medRxiv. https://doi.org/10.1101/2021.08.31.21262938.

Bivins, A., Greaves, J., Fischer, R., Yinda, K.C., Ahmed, W., Kitajima, M., et al., 2020. Persistence of SARS-CoV-2 in water and wastewater. Environ. Sci. Technol. Lett. 7, 937–942.

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527.

CDC, 2023. National Wastewater Surveillance System (NWSS). Centers for Disease Control and Prevention [Internet], 20 Jun. [cited 24 Jan 2024]. Available: https://www.cdc.gov/nwss/wastewater-surveillance.html.

Center for Food Safety, 2023. Nutrition A. Wastewater Surveillance for SARS-CoV-2 Variants. U.S. Food and Drug Administration [Internet]. FDA, 4 May. [cited 6 Feb 2024]. Available: https://www.fda.gov/food/whole-genome-sequencing-wgs-program/wastewater-surveillance-sars-cov-2-variants.

Child, T., 2022. H. Wasetewater Sequencing using the EasySeq™ RC-PCR SARS CoV-2 (Nimagen) V3.0 v2. https://doi.org/10.17504/protocols.io.81wgb7bx3vpk/v3.

covid-analysis: SARS-CoV-2 sequencing and strain identification pipeline for nanopore samples. Github; Available: https://github.com/enviro-lab/covid-analysis.

C-WAP: SC2 variant detection and composition pipeline. Github; Available: https://github.com/CFSAN-Biostatistics/C-WAP.

C-WAP: SC2 variant detection and composition pipeline. Github; Available: https://github.com/CFSAN-Biostatistics/C-WAP.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., et al., 2021. Twelve years of SAMtools and BCFtools. Gigascience 10. https://doi.org/10.1093/gigascience/giab008.

Dreifuss, D., Topolsky, I., Icer Baykal, P., Beerenwinkel, N., 2022. Tracking SARS-CoV-2 genomic variants in wastewater sequencing data with *LolliPop*. bioRxiv. https://doi.org/10.1101/2022.11.02.22281825.

Ellmen, I., Lynch, M.D.J., Nash, D., Cheng, J., Nissimov, J.I., Charles, T.C., 2021. Alcov: estimating variant of concern abundance from SARS-CoV-2 wastewater sequencing data. medRxiv. https://doi.org/10.1101/2021.06.03.21258306, p. 2021.06.03.21258306.

Ferdous, J., Weathers, T., Bharati Barua, V., Stiers, E., France, A., Lambirth, C.K., et al., 2021. A SARS-CoV-2 Surveillance Sequencing Protocol Optimized for Oxford Nanopore PromethION v1. https://doi.org/10.17504/protocols.io.butbnwin.

fieldbioinformatics: The ARTIC field bioinformatics pipeline. Github; Available: https://github.com/artic-network/fieldbioinformatics.

Freed, N.E., Vlková, M., Faisal, M.B., Silander, O.K., 2020. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford nanopore rapid barcoding. Biol Methods Protoc. 5, bpaa014.

Freyja: Depth-weighted De-Mixing. Github; Available: https://github.com/andersen-lab/Freyja.

Freyja: Depth-weighted De-Mixing. Github; Available: https://github.com/andersen-lab/Freyja.

Gibas, C., Lambirth, K., Mittal, N., Juel, M.A.I., Barua, V.B., Brazell, L.R., et al., 2021. Implementing building-level SARS-CoV-2 wastewater surveillance on a university campus. Sci. Total Environ. 146749.

Gregory, D.A., Wieberg, C.G., Wenzel, J., Lin, C.-H., Johnson, M.C., 2021. Monitoring SARS-CoV-2 populations in wastewater by amplicon sequencing and using the novel program SAM refiner. Viruses 13. https://doi.org/10.3390/v13081647.

Ivanova, O.E., Yarmolskaya, M.S., Eremeeva, T.P., Babkina, G.M., Baykova, O.Y., Akhmadishina, L.V., et al., 2019. Environmental surveillance for poliovirus and other enteroviruses: long-term experience in Moscow, Russian Federation, 2004–2017. Viruses 11. https://doi.org/10.3390/v11050424.

Jahn, K., Dreifuss, D., Topolsky, I., Kull, A., Ganesanandamoorthy, P., Fernandez-Cassi, X., et al., 2022. Early detection and surveillance of SARS-CoV-2 genomic variants in wastewater using COJAC. Nat. Microbiol. 7, 1151–1160.

Karthikeyan, S., Levy, J.I., De Hoff, P., Humphrey, G., Birmingham, A., Jepsen, K., et al., 2022. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature 609, 101–108.

Katayama, H., Shimasaki, A., Ohgaki, S., 2002. Development of a virus concentration method and its application to detection of enterovirus and Norwalk virus from coastal seawater. Appl. Environ. Microbiol. 68, 1033–1039.

Kayikcioglu, T., Amirzadegan, J., Rand, H., Tesfaldet, B., Timme, R.E., Pettengill, J.B., 2023. Performance of methods for SARS-CoV-2 variant detection and abundance estimation within mixed population samples. PeerJ 11, e14596.

Kumar, S., Karuppanan, K., Subramaniam, G., 2022. Omicron (BA.1) and sub-variants (BA.1.1, BA.2, and BA.3) of SARS-CoV-2 spike infectivity and pathogenicity: a comparative sequence and structural-based computational assessment. J. Med. Virol. 94, 4780–4791.

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100.

Li, J., Lai, S., Gao, G.F., Shi, W., 2021. The emergence, genomic diversity and global spread of SARS-CoV-2. Nature 600, 408–418.

Life, 2021. A brief history of wastewater testing and pathogen detection. In: Life in the Lab [Internet]. Thermo Fisher Scientific, 5 Oct. [cited 24 Jan 2024]. Available: https://www.thermofisher.com/blog/life-in-the-lab/a-brief-history-of-wastewater-testing-and-pathogen-detection/.

Lin, X., Glier, M., Kuchinski, K., Ross-Van Mierlo, T., McVea, D., Tyson, J.R., et al., 2021. Assessing multiplex tiling PCR sequencing approaches for detecting genomic variants of SARS-CoV-2 in municipal wastewater. mSystems 6, e0106821.

Peccia, J., Zulli, A., Brackney, D.E., Grubaugh, N.D., Kaplan, E.H., Casanovas-Massana, A., et al., 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nat. Biotechnol. 38, 1164–1167.

Pechlivanis, N., Tsagiopoulou, M., Maniou, M.C., Togkousidis, A., Mouchtaropoulou, E., Chassalevris, T., et al., 2022. Detecting SARS-CoV-2 lineages and mutational load in municipal wastewater and a use-case in the metropolitan area of Thessaloniki, Greece. Sci. Rep. 12, 2659.

Posada-Céspedes, S., Seifert, D., Topolsky, I., Jablonski, K.P., Metzner, K.J., Beerenwinkel, N., 2021. V-pipe: a computational pipeline for assessing viral genetic diversity from high-throughput data. Bioinformatics 37, 1673–1680.

QuantaSoft™ software, 2024. Regulatory edition #1864011. In: Bio-Rad Laboratories [Internet] cited 7 Feb. Available: https://www.bio-rad.com/en-us/life-science/digital-pcr/qx200-droplet-digital-pcr-system/quantasoft-software-regulatory-edition.

Quick, J., 2020. nCoV-2019 Sequencing Protocol v3 (LoCost) v3. https://doi.org/10.17504/protocols.io.bp2l6n26rgqe/v3.

Ramachandran, P., 2022. Modified NEBNext® VarSkip Short SARS-CoV-2 Enrichment and Library Prep for Oxford Nanopore Technologies- Adapted for Wastewater samples v2. https://doi.org/10.17504/protocols.io.3byl4bwervo5/v2.

Rego, N., Costábile, A., Paz, M., Salazar, C., Perbolianachis, P., Spangenberg, L., et al., 2021. Implementation of a qPCR Assay Coupled With Genomic Surveillance for Real-time Monitoring of SARS-CoV-2 Variants of Concern. bioRxiv. medRxiv. https://doi.org/10.1101/2021.05.20.21256969.

Shafer, M.M., Bobholz, M.J., Vuyk, W.C., Gregory, D., Roguet, A., Haddock Soto, L.A., et al., 2022. Tracing the origin of SARS-CoV-2 Omicron-like spike sequences detected in wastewater. bioRxiv. https://doi.org/10.1101/2022.10.28.22281553.

Smyth, D.S., Trujillo, M., Gregory, D.A., Cheung, K., Gao, A., Graham, M., et al., 2021. Tracking Cryptic SARS-CoV-2 Lineages Detected in NYC Wastewater. bioRxiv. medRxiv. https://doi.org/10.1101/2021.07.26.21261142.

Solo-Gabriele, H.M., Kumar, S., Abelson, S., Penso, J., Contreras, J., Babler, K.M., et al., 2023. Predicting COVID-19 cases using SARS-CoV-2 RNA in air, surface swab and wastewater samples. Sci. Total Environ. 857, 159188.

Tallei, T.E., Alhumaid, S., AlMusa, Z., Fatimawali, Kusumawaty D., Alynbiawi, A., et al., 2023. Update on the omicron sub-variants BA.4 and BA.5. Rev. Med. Virol. 33, e2391.

Valieris, R., Drummond, R.D., Defelicibus, A., Dias-Neto, E., Rosales, R.A., Tojal da Silva, I., 2022. A mixture model for determining SARS-Cov-2 variant composition in pooled samples. Bioinformatics 38, 1809–1815.

VarSkip: VarSkip multiplex PCR designs for SARS-CoV-2 sequencing. Github; Available: https://github.com/nebiolabs/VarSkip.

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20, 257.

Ye, S.H., Siddle, K.J., Park, D.J., Sabeti, P.C., 2019. Benchmarking metagenomics tools for taxonomic classification. Cell 178, 779–794.