AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

SARS-CoV-2 wastewater variant surveillance: pandemic response leveraging FDA's GenomeTrakr network

Ruth Timme, Jacquelina Woods, Jessica Jones, Kevin Calci, Rachel Rodriguez, Candace Barnes, Elizabeth Hyde-Leard, Mark Craven, Haifeng Chen, Cameron Boerner, Christopher Grim, Amanda Windsor, Padmini Ramachandran, Tim Muruvanda, Hugh Rand, Bereket Tesfaldet, Jasmine Amirzadegan, Tunc Kayikcioglu, Tamara Walsky, Marc Allard, Maria Balkey, C Bias, Eric Brown, Kathryn Judy, Tina Lusk Pfefer, Sandra Tallent, Maria Hoffmann, The GenomeTrakr Laboratory Consortium, and James Pettengill

Corresponding Author(s): Ruth Timme, US Food and Drug Administration

Review Timeline:

Submission Date:	January 4, 2024
Editorial Decision:	March 7, 2024
Revision Received:	March 25, 2024
Accepted:	April 5, 2024

Editor: Christopher Marshall

Reviewer(s): The reviewers have opted to remain anonymous.

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

DOI: https://doi.org/10.1128/msystems.01415-23

Re: mSystems01415-23 (SARS-CoV-2 wastewater variant surveillance: pandemic response leveraging FDA's GenomeTrakr network)

Dear Dr. Ruth E Timme:

Thank you for the privilege of reviewing your work. The reviewers recommend mostly minor modifications prior to publication. Below you will find instructions from the mSystems editorial office and the reviewer comments.

Please return the manuscript within 60 days; if you cannot complete the modification within this time period, please contact me. If you do not wish to modify the manuscript and prefer to submit it to another journal, notify me immediately so that the manuscript may be formally withdrawn from consideration by mSystems.

Revision Guidelines
To submit your modified manuscript, log into the submission site at https://msystems.msubmit.net/cgi-bin/main.plex. Go to Author Tasks and click the appropriate manuscript title to begin. The information you entered when you first submitted the paper will be displayed; update this as necessary. Note the following requirements:

• Upload point-by-point responses to the issues raised by the reviewers in a file named "Response to Reviewers," NOT IN YOUR COVER LETTER
• Upload a compare copy of the manuscript (without figures) as a "Marked-Up Manuscript" file
• Upload a clean .DOC/.DOCX version of the revised manuscript and remove the previous version
• Each figure must be uploaded as a separate, editable, high-resolution file (TIFF or EPS preferred), and any multipanel figures must be assembled into one file
• Any supplemental material intended for posting by ASM should be uploaded separate from the main manuscript; you can combine all supplemental material into one file (preferred) or split it into a maximum of 10 files, with all associated legends included

For complete guidelines on revision requirements, see our Submission and Review Process webpage. Submission of a paper that does not conform to guidelines may delay acceptance of your manuscript.

Data availability: ASM policy requires that data be available to the public upon online posting of the article, so please verify all links to sequence records, if present, and make sure that each number retrieves the full record of the data. If a new accession number is not linked or a link is broken, provide mSystems production staff with the correct URL for the record. If the accession numbers for new data are not publicly accessible before the expected online posting of the article, publication may be delayed; please contact production staff (mSystems@asmusa.org) immediately with the expected release date.

Publication Fees: For information on publication fees and which article types are subject to charges, visit ourwebsite. If your manuscript is accepted for publication and any fees apply, you will be contacted separately about payment during the production process; please follow the instructions in that e-mail. Arrangements for payment must be made before your article is published.

ASM Membership: Corresponding authors may join or renew ASM membership to obtain discounts on publication fees. Need to upgrade your membership level? Please contact Customer Service at Service@asmusa.org.

The ASM Journals program strives for constant improvement in our submission and publication process. Please tell us how we can improve your experience by taking this quick Author Survey.

Thank you for submitting your paper to mSystems.

Sincerely,
Christopher Marshall
Editor
mSystems

Reviewer #1 (Comments for the Author):

The manuscript by Timme et al. tells the story of how the US FDA set up a network of laboratories sequencing SARS-CoV-2 from wastewater. The information combined within provides a summary of all the components necessary for this network and provides a guide for future pandemic response efforts. Some of these protocols and bioinformatics pipelines have been published previously, but what appears to be novel is the quality assessment and quality control scoring system. More information on how this system was developed (how thresholds were decided) may be useful to include. Additionally, with all the samples processed by this network, surely there are some quantitative analysis that might be of interest to readers. One major

analysis that comes to mind is the threshold for gene copies or Cq value below which sequencing wastewater was generally unsuccessful.

L140: "accessed" should be "assessed"
L324: don't understand how this sampling was both systematic and random, suggest deleting "random"
L338: is the "3%" a typo?
Table 2: why does this table include more labs than are shown in Fig S2?
Missing figure legends for supplementary figures. What do the colors mean in Fig S2?

Reviewer #2 (Comments for the Author):

The authors describe SARS-COV-2 wastewater surveillance through the lens of the FDA's GenomeTrakr network. GenomeTrakr is a pathogen genomic surveillance network led by the FDA Center for Food Safety and Applied Nutrition, which collaborates with US Government and State public health agencies for foodborne pathogen surveillance/tracking. This manuscript outlines two main goals: (i) summarizing genomic data from sequencing SARS-CoV-2 in wastewater using the FDA's lab network and (ii) providing best practices for managing population-level Next Generation Sequencing (NGS) data for pathogen surveillance in environmental samples. Overall, this is an important contribution and commendable effort. The manuscript reads well and highlights the key aspects via the included display items. I provide more detailed comments below:

1. My main comment is the state of the GitHub repository. First, the link provided is broken (it leads to a non existent repository: https://github.com/CFSAN-Biostatistics/WW-SC2-745; the line number somehow is getting introduced into the URL). Next, when accessing the actual repository (https://github.com/CFSAN-Biostatistics/WW-SC2-variant-estimations), it points the user to: https://github.com/CFSAN-Biostatistics/C-WAP. Then when accessing that page, the user sees: "*Given the project timeline, C-WAP will no longer be under active development or maintenance come June 30, 2023. Please refer to the C-WAP successor Aquascope for an actively supported workflow. Freyja or Kallisto (two of the tools C-WAP incorporates) may also be of interest. Thank you for joining us on our analytic journey*" So it is unclear which is the recommended best practice pipeline provided in Figure 2. At a minimum, there should be clarification on this in the text.

2. Regarding the included tools in the pipeline, Freyja etc, no parameter settings nor versions were provided. Overall, I found the details on the computational method side to be lacking.

3. Supplementary files, while very useful, were a bit difficult to navigate. Collating them into 1 or 2 files with a table of contents would make them easier to navigate.

4. A quality control filter that defaults to 20% of SARS-CoV-2 genome covered seems like a very low bar to meet. Some justification for this value (as the default) would be helpful.

5. The data package details are pasted in between lines 205 and 206, and not labeled as a display item.

Reviewer #3 (Comments for the Author):

This is a well-conceived, well-written, and well-executed manuscript that will be of interest in the field. I have only minor editorial comments and suggestions as detailed below.

L140 assessed? or accessed?

L178 This sentence seems to match better with Figure 1.

L191 It seems this should reference Figure 2? (At least the PDF that has been labeled Figure 2 in the submission)

L220. It looks like the Figure references in line are off by 1 throughout.

L268 access or assess?

L319. SRA is used prior to this instance without specifying the full meaning. Suggest moving this to that instance.

L325 spelling: staggered

L342 Is the sample size for the Promega TNA kit correct?

L390 delete second "in"

L446 delete second comma.

L477 I think it would be very valuable to add a forward looking paragraph to the discussion. Based on the author's experience and expertise, what are some promising use cases moving forward? How can wastewater surveillance add value for this type of approach/system beyond SARS-CoV-2.

L729 The figures provided in the manuscript file and as separate PDFs do not align with one another. It seems the attached PDFs have an extra figure that is denoted as Figure 1.

18 March 2024

Christopher Marshall, PhD
Editor
mSystems

Dear Dr. Marshall,

Thank you for handling these three thorough reviews of our manuscript. In response to each of the reviewers' comments, I have provided detailed point-by-point responses in this letter (responses in BLUE) and have made the corresponding edits in the manuscript file using track changes.

Reviewer #1 (Comments for the Author):

The manuscript by Timme et al. tells the story of how the US FDA set up a network of laboratories sequencing SARS-CoV-2 from wastewater. The information combined within provides a summary of all the components necessary for this network and provides a guide for future pandemic response efforts. Some of these protocols and bioinformatics pipelines have been published previously, but what appears to be novel is the quality assessment and quality control scoring system. More information on how this system was developed (how thresholds were decided) may be useful to include.

This is a good comment and question by the reviewer. Because this project was rolled out in response to the pandemic, normal FDA method development + formal validation procedures were not performed prior to sample collection. Instead, we started collecting data at the same time we were developing the method. Our QC thresholds were set based on the first couple months of sampling, and we chose thresholds that seemed reasonable at the time given the data we were collecting. We've added the following language to better communicate this process:

"**These thresholds, deemed appropriate based on early data collection, served as a preliminary benchmark. However, we acknowledge the need for a rigorous validation process to fine-tune these thresholds to suit specific applications, recognizing that different use cases—such as general population variant tracking versus the validation of a new diagnostic kit—may require different QC thresholds.**"

Additionally, with all the samples processed by this network, surely there are some quantitative analysis that might be of interest to readers. One major analysis that comes to mind is the threshold for gene copies or Cq value below which sequencing wastewater was generally unsuccessful.

We explored but didn't find a level of target RNA that was useful for this project. We did, however, identify a post PCR QC (QC Step #2, Figure 3) that we found to be most critical to evaluate prior to the expensive library prep and sequencing step.

L140: "accessed" should be "assessed"

U.S. Food & Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993
www.fda.gov

Thanks, fixed this typo.

L324: don't understand how this sampling was both systematic and random, suggest deleting "random"
Agreed, I removed the word “random” here.

L338: is the "3%" a typo?
Yes, I corrected this typo to read, (40.1483 (30%).

Table 2: why does this table include more labs than are shown in Fig S2?
18 of the 20 funded labs collected, tested, and attempted to sequence the samples. California and West Virginia encountered barriers that prevented them even setting up the sample collection, so no samples collected resulted in those labs not being included in the S2 gantt chart.

What do the colors mean in Fig S2?

Thanks for pointing this out – the missing legend has been added back to Fig S2 (Orange bars represent sample-only data, blue bars = sample + sequence data).

Missing figure legends added to the supplementary figures:

Supplemental Figure 1: Wastewater protocol pilot exercise sequencing results. This figure presents the variant results of a pilot study where nine laboratories sequenced the same four wastewater samples. These samples are represented across four columns and are labeled as WPP-sample_B.01, C.01, SA-1.01, and SA-2.01, detailed in Table 1. Each rectangle includes the relative abundance of SARS-CoV-2 variants and sub-lineages recovered for that sample. A summary of the laboratory methodologies and quality control (QC) metrics for each entry are included in Supplemental Table 2.

Supplemental Figure 2. Sample Collection and Sequencing Over Time by Laboratory. Each vertical bar represents a sample submitted to NCBI’s BioSample. Bar size is proportional to the number of samples collected during a week. Sequencing was performed on all samples containing detectable levels of SARS-CoV-2 RNA, and then submitted to NCBI's Sequence Read Archive (SRA) upon satisfying quality control (QC) criteria described in this manuscript.

Reviewer #2 (Comments for the Author):

The authors describe SARS-COV-2 wastewater surveillance through the lens of the FDA's GenomeTrakr network. GenomeTrakr is a pathogen genomic surveillance network led by the FDA Center for Food Safety and Applied Nutrition, which collaborates with US Government and State public health agencies for foodborne pathogen surveillance/tracking. This manuscript outlines two main goals: (i) summarizing genomic data from sequencing SARS-CoV-2 in wastewater using the FDA's lab network and (ii) providing best practices for managing population-level Next Generation Sequencing (NGS) data for pathogen surveillance in environmental samples. Overall, this is an important contribution and commendable effort. The manuscript reads well and highlights the key aspects via the included display items. I provide more detailed comments below:

1. My main comment is the state of the GitHub repository. First, the link provided is broken (it leads to a non existent repository: https://github.com/CFSAN-Biostatistics/WW-SC2-745; the line number somehow is getting introduced into the URL). Next, when accessing the actual repository

(https://github.com/CFSAN-Biostatistics/WW-SC2-variant-estimations), it points the user to: https://github.com/CFSAN-Biostatistics/C-WAP. Then when accessing that page, the user sees: "*Given the project timeline, C-WAP will no longer be under active development or maintenance come June 30, 2023. Please refer to the C-WAP successor Aquascope for an actively supported workflow. Freyja or Kallisto (two of the tools C-WAP incorporates) may also be of interest. Thank you for joining us on our analytic journey*" So it is unclear which is the recommended best practice pipeline provided in Figure 2. At a minimum, there should be clarification on this in the text.

Thanks for pointing out the broken link, it's been fixed now in the revised version. I also added the following sentence near line 224 (Methods: Data flow and visualization):

"C-WAP has been repackaged as Aquascope, however, despite the rebranding, the underlying algorithm remains the same, focusing on quality control (QC) metrics and using Freyja to infer relative abundances of SARS-CoV-2 lineages in each sample (35)."

2. Regarding the included tools in the pipeline, Freyja etc, no parameter settings nor versions were provided. Overall, I found the details on the computational method side to be lacking.

We have added details to the pipeline and emphasize the previously published manuscript that describes the method in greater detail and github repository as a source of additional information.

3. Supplementary files, while very useful, were a bit difficult to navigate. Collating them into 1 or 2 files with a table of contents would make them easier to navigate.

I tried to combine the supplemental files but ended up having to separate them again during resubmission. I expect once they get lined correctly to the published manuscript, they'll be easy to locate for the reader.

4. A quality control filter that defaults to 20% of SARS-CoV-2 genome covered seems like a very low bar to meet. Some justification for this value (as the default) would be helpful.

Thank you for pointing out this potentially conflicting statement. Our QC threshold for % genome covered was 60% or higher, so I see why this would be confusing. I decided to remove this statement altogether. A little background, we added a sliding button to the dashboard so the user could play around with different QC filters. The default setting we chose for this slider was arbitrary (in retrospect, we should have made it higher, since most of our data are of much higher quality).

5. The data package details are pasted in between lines 205 and 206, and not labeled as a display item.

Thanks for pointing this out. We'll work with the editor to ensure this list of attributes are formatted correctly for the journal. They can be converted to a Table if necessary.

Reviewer #3 (Comments for the Author):

This is a well-conceived, well-written, and well-executed manuscript that will be of interest in the field. I have only minor editorial comments and suggestions as detailed below.

L140 assessed? or accessed?

"assessed", thank you for catching this typo!

L178 This sentence seems to match better with Figure 1.

The confusion might stem from the incorrectly labeled figure files – these have been corrected in this revision. However, there's no figure citation in this section (L178) - the QC checkpoint figure is first referenced in the results as "Figure 3".

L191 It seems this should reference Figure 2? (At least the PDF that has been labeled Figure 2 in the submission)

The PDF file names for Figures 1-3 were off by one and have been corrected in this revision. The in-manuscript figures were correctly numbered. Thank you for catching this error.

L220. It looks like the Figure references in line are off by 1 throughout.
The PDF file names for Figures 1-3 were off by one and have been corrected in this revision. The in-manuscript figures were correctly numbered. Thank you for catching this error.

L268 access or assess?
Typo corrected to "assessed".

L319. SRA is used prior to this instance without specifying the full meaning. Suggest moving this to that instance.

Thank you for catching this error – we correctly defined the abbreviation in L200 at first use and removed the definition here.

L325 spelling: staggered
Spelling corrected to "staggered".

L342 Is the sample size for the Promega TNA kit correct?
The sample size has been corrected to "Promega large volume TNA capture kit (n=455)".

L390 delete second "in"
Correction made.

L446 delete second comma.
Edit made.

L477 I think it would be very valuable to add a forward looking paragraph to the discussion. Based on the author's experience and expertise, what are some promising use cases moving forward? How can wastewater surveillance add value for this type of approach/system beyond SARS-CoV-2.

We agree this is a good idea, so we added the following language to L479-483:

"Incorporating a signal provided through wastewater sampling to the existing US surveillance strategies for enteric pathogens would provide a more complete picture of where pathogens are and

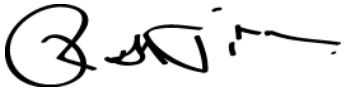
are not circulating across the country, enabling more precise scoping of foodborne outbreaks. The potential for this expansion is currently being explored within the framework of the US National Wastewater Surveillance System."

L729 The figures provided in the manuscript file and as separate PDFs do not align with one another. It seems the attached PDFs have an extra figure that is denoted as Figure 1.

We thank the reviewer for catching this Figure numbering error. The figures, file names, and citations have been corrected throughout the manuscript.

Thank you once again for your time as editor and to the three reviewers who provided thorough and thoughtful reviews. I am confident that these revisions have strengthened the overall quality of our manuscript.

With warm regards,



Ruth E. Timme, PhD
GenomeTrakr Program lead
Center for Food Safety and Applied Nutrition
US Food and Drug Administration

Re: mSystems01415-23R1 (SARS-CoV-2 wastewater variant surveillance: pandemic response leveraging FDA's GenomeTrakr network)

Dear Dr. Ruth E Timme:

Congratulations, your manuscript has been accepted and I am forwarding it to the ASM production staff for publication. Your paper will first be checked to make sure all elements meet the technical requirements. ASM staff will contact you if anything needs to be revised before copyediting and production can begin. Otherwise, you will be notified when your proofs are ready to be viewed.

Data Availability: ASM policy requires that data be available to the public upon online posting of the article, so please verify all links to sequence records, if present, and make sure that each number retrieves the full record of the data. If a new accession number is not linked or a link is broken, provide production staff with the correct URL for the record. If the accession numbers for new data are not publicly accessible before the expected online posting of the article, publication may be delayed; please contact ASM production staff immediately with the expected release date.

Publication Fees: For information on publication fees and which article types have charges, please visit ourwebsite. We have partnered with Copyright Clearance Center (CCC) to collect author charges. If fees apply to your paper, you will receive a message from no-reply@copyright.com with further instructions. For questions related to paying charges through RightsLink, please contact CCC at ASM_Support@copyright.com or toll free at +1-877-622-5543. CCC makes every attempt to respond to all emails within 24 hours.

ASM Membership: Corresponding authors may join or renew ASM membership to obtain discounts on publication fees. Need to upgrade your membership level? Please contact Customer Service at Service@asmusa.org.

PubMed Central: ASM deposits all mSystems articles in PubMed Central and international PubMed Central-like repositories immediately after publication. Thus, your article is automatically in compliance with the NIH access mandate. If your work was supported by a funding agency that has public access requirements like those of the NIH (e.g., the Wellcome Trust), you may post your article in a similar public access site, but we ask that you specify that the release date be no earlier than the date of publication on the mSystems website.

Embargo Policy: A press release may be issued as soon as the manuscript is posted on themSystems Latest Articles webpage. The corresponding author will receive an email with the subject line "ASM Journals Author Services Notification" when the article is available online.

Cover Image Submissions: If you would like to submit a potential Cover Image, please email a file and a short legend to msystems@asmusa.org. Please note that we can only consider images that (i) the authors created or own and (ii) have not been previously published. By submitting, you agree that the image can be used under the same terms as the published article. Image File requirements: TIF/EPS, 7.5 inches wide by 8.25 inches tall (at least 2,250 pixels wide by 2,475 pixels tall), minimum 300 dpi resolution (600 dpi preferred), RGB, and no figure elements, e.g., arrows or panel labels. The legend should be a short description of the image, 1-2 sentences recommended.

Author Video:: For mSystems research articles, you are welcome to submit a short author video for your recently accepted paper. Videos are normally 1 minute long and are a great opportunity for junior authors to get greater exposure. Importantly, this video will not hold up the publication of your paper and you can submit it at any time.

Details of the video are:
· Minimum resolution of 1280 x 720
· .mov or .mp4 video format
· Provide video in the highest quality possible but do not exceed 1080p
· Provide a still/profile picture that is 640 (w) x 720 (h) max
· Provide the script that was used

We recognize that the video files can become quite large, so to avoid quality loss ASM suggests sending the video file via https://www.wetransfer.com/. When you have a final version of the video and the still ready to share, please send it to mSystems staff at mSystems@asmusa.org.

Thank you for submitting your paper to mSystems.

Sincerely,

Christopher Marshall
Editor
mSystems

Reviewer #1 (Comments for the Author):

The manuscript has been sufficiently improved for publication, and my concerns have been addressed.