AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Spectrum

VPipe: An Automated Bioinformatics Platform for Assembly and Management of Viral Next-Generation Sequencing Data

Darlene Wagner, Rachel Marine, Edward Ramos, Terry Fei Fan Ng, Christina Castro, Margaret Okomo-Adhiambo, Krysten Harvey, Gregory Doho, Reagan Kelly, Yatish Jain, Roman Tatusov, Hideky Silva, Paul Rota, Agha Khan, and M. Steven Oberste

Corresponding Author(s): Margaret Okomo-Adhiambo, Centers for Disease Control and Prevention

Review Timeline:

Submission Date:	December 9, 2021
Editorial Decision:	December 20, 2021
Revision Received:	January 6, 2022
Accepted:	January 11, 2022

Editor: Wei-Hua Chen

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/spectrum.02564-21

December 20, 2021

Dr. Margaret Okomo-Adhiambo
Centers for Disease Control and Prevention
Influenza Division
1600 Clifton Rd
Atlanta, GA 30329

Re: Spectrum02564-21 (VPipe: An Automated Bioinformatics Platform for Assembly and Management of Viral Next-Generation Sequencing Data)

Dear Dr. Margaret Okomo-Adhiambo:

Thank you for submitting your manuscript to Microbiology Spectrum. Your manuscript have been reviewed by two reviewers. Although their comments are in general positive, both complained about the user interface (among other things).

In light of their comments, I am happy to invite you to submit a revised version of your manuscript. Below please find their comments. When submitting the revised version of your paper, please provide (1) point-by-point responses to the issues raised by the reviewers as file type "Response to Reviewers," not in your cover letter, and (2) a PDF file that indicates the changes from the original submission (by highlighting or underlining the changes) as file type "Marked Up Manuscript - For Review Only". Please use this link to submit your revised manuscript - we strongly recommend that you submit your paper within the next 60 days or reach out to me. Detailed instructions on submitting your revised paper are below.

Link Not Available

Thank you for the privilege of reviewing your work. Below you will find instructions from the Microbiology Spectrum editorial office and comments generated during the review.

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.

Sincerely,

Wei-Hua Chen

Editor, Microbiology Spectrum

Journals Department
American Society for Microbiology
1752 N St., NW
Washington, DC 20036
E-mail: spectrum@asmusa.org

Reviewer comments:

Reviewer #2 (Comments for the Author):

In this article, Darlene D. Wagner et al developed an Automated Bioinformatics Platform for Assembly and Management of Viral (VPipe), which can automate genome assembly for a wide range of viruses, including high-consequence pathogens such as SARS-CoV-2. The author showed that VPipe generated assemblies with greater contiguity compared to other NGS pipelines and possesses a set of data management and query functions not commonly found in other freely available NGS pipelines. Although demonstrated a certain practicality, my main concern is that the VPipe does not show enough novelty and innovation.
1. It seems VPipe did not develop it own algorithm, instead, refer to existing sequence alignment software and algorithms?
2. The assembly of the VPipe does not exhibit particularly obvious advantages compared to drVM, 102 EDGE, VirMAP, and Genome Detective, except for contig length and contiguity.
3. Lacking user interactive interface, this will bring a bad user experience.

Reviewer #1: (No comments were given to the author)

Staff Comments:

Preparing Revision Guidelines
To submit your modified manuscript, log onto the eJP submission site at https://spectrum.msubmit.net/cgi-bin/main.plex. Go to Author Tasks and click the appropriate manuscript title to begin the revision process. The information that you entered when you first submitted the paper will be displayed. Please update the information as necessary. Here are a few examples of required updates that authors must address:

• 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.
• Each figure must be uploaded as a separate file, and any multipanel figures must be assembled into one file.
• Manuscript: A .DOC version of the revised manuscript
• Figures: Editable, high-resolution, individual figure files are required at revision, TIFF or EPS files are preferred

For complete guidelines on revision requirements, please see the journal Submission and Review Process requirements at https://journals.asm.org/journal/Spectrum/submission-review-process. **Submissions of a paper that does not conform to Microbiology Spectrum guidelines will delay acceptance of your manuscript.** "

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, please notify me of your decision immediately so that the manuscript may be formally withdrawn from consideration by Microbiology Spectrum.

If your manuscript is accepted for publication, you will be contacted separately about payment when the proofs are issued; please follow the instructions in that e-mail. Arrangements for payment must be made before your article is published. For a complete list of **Publication Fees**, including supplemental material costs, please visit our website.

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.

Thank you for submitting your paper to Microbiology Spectrum.

Reviewer #2

General Comment:

In this article, Darlene D. Wagner et al developed an Automated Bioinformatics Platform for Assembly and Management of Viral (VPipe), which can automate genome assembly for a wide range of viruses, including high-consequence pathogens such as SARS-CoV-2. The author showed that VPipe generated assemblies with greater contiguity compared to other NGS pipelines and possesses a set of data management and query functions not commonly found in other freely available NGS pipelines. Although demonstrated a certain practicality, my main concern is that the VPipe does not show enough novelty and innovation.

Response:

Thank you for your concern. The intention of this publication is to demonstrate a pipeline comprised of common tools applied to a unique domain (i.e., small public health laboratories which lack bioinformatics staffing). Rather than implementing novel software components, VPipe utilizes software which performs consistently well for viral data as shown through internal testing and previous studies (Sutton et al., 2019, Lin et al., 2017, García-López et al., 2015, Ji et al., 2018) (revised manuscript, lines 211-212). The setup of VPipe is unique in its off-the-shelf approach to availability for viral genomics analysis. Laboratory access to VPipe is obtained by contacting system administrators at VPipe@cdc.gov and no subsequent software download/installation is needed. Moreover, VPipe does not require user interaction at the data input stage of the pipeline, further accommodating users who lack bioinformatics training/experience. For high-consequence pathogens such as SARS-CoV-2, VPipe automatically performs SNPs analysis in addition to genome assembly. Granted, EDGE, Genome Detective, VirMAP, and VirAMP (Wan et al., 2015), may provide feasible analysis solutions for mid-to-large-sized laboratories which possess bioinformatics analysis capabilities. As referenced on lines 100 through 103 in the revised manuscript, VPipe is highlighted as the only tool that provides free access to public health users, data security, and long-term retention of analysis results.

Specific Comment 1:

It seems VPipe did not develop it own algorithm, instead, refer to existing sequence alignment software and algorithms?

Response:

VPipe utilizes open-source software rather than custom algorithms among its constituent components; Its overall implementation is optimized for small public-health laboratories with limited or no bioinformatics capabilities (see response to general comment). Unlike comprehensive tools such as EDGE (Li et al., 2107), VPipe does not require software downloads or containerization, improving usability for non-bioinformaticians. Finally, VPipe implements both assembly and SNPs discovery by default for SARS-CoV-2. VPipe is distinguished from pipelines which offer assembly only, such as Bio-Docklets (Kim et al., 2017), and InteMAP (Lai et al., 2015) or SNPs only, such as DNAscan (Iacoangeli et al., 2019).

Specific Comment 2:

The assembly of the VPipe does not exhibit particularly obvious advantages compared to drVM, EDGE, VirMAP, and Genome Detective, except for contig length and contiguity.

Response:

VPipe has multiple advantages over drVM, EDGE, VirMAP, and Genome Detective for its intended user base of small public health laboratories. Based upon the results shown in Table 1, VPipe assemblies were comparable to drVM, EDGE, VirMAP, and Genome Detective. For larger numbers of runs from a single project as shown in Figure 2, however, VPipe assemblies show statistically significantly higher contiguity than Genome Detective and EDGE (without read quality settings that mimic VPipe). When EDGE parameters are adjusted to imitate VPipe quality control parameters, EDGE matches VPipe assembly performance. VPipe also offers features not found or freely-available in the other tools. Genome Detective, drVM, and VirMAP do not offer long-term storage of assemblies or other analysis results. Although EDGE enables long-term storage of analysis results, it requires local installation using Docker and lacks the remote server storage advantages of VPipe. The many customizable parameters make EDGE a good choice for experienced genome biologists/bioinformaticists, but VPipe is optimized to generate reproducible results for inexperienced users.

Specific Comment 3:

Lacking user interactive interface, this will bring a bad user experience.

Response:

VPipe has a well-maintained graphical user interface and example screen shots are included in the Supplemental Materials. Because it is hosted on a U.S. government system, user access is granted to the VPipe homepage by sending a request to VPipe@cdc.gov.

Comment on Literature:

The editor was helpful in listing ten publications as "PubMed Similar Manuscripts" which potentially share methodology and application scope with "VPipe: An Automated Bioinformatics Platform for Assembly and Management of Viral Next-Generation Sequencing Data". Full references to the ten PubMed Similar Manuscripts are shown below and are sorted in order of relevance, with the top four references cited in the revised manuscript:

Wan, Yinan, et al. (2015) VirAmp: a galaxy-based viral genome assembly pipeline. GigaScience. 4:19.

Lai, Binbin, et al. (2015) InteMAP: Integrated metagenomic assembly pipeline for NGS short reads. BMC Bioinformatics. 16:244

Kim, Baekdoo, et al. (2017) Bio-Docklets: virtualization containers for single-step execution of NGS pipelines. GigaScience, 6, 2017, 1–7

Iacoangeli, A., et al. (2019) DNAscan: personal computer compatible NGS analysis, annotation and visualisation. BMC Bioinformatics. 20:213

Kilianski, Andy, et al. (2015) Pathosphere.org: pathogen detection and characterization through a web-based, open source informatics platform. BMC Bioinformatics. 16:416

Nagasaki, Hideki, et al. (2013) DDBJ read annotation pipeline: a cloud computing-based pipeline for high-throughput analysis of next-generation sequencing data. DNA Research 20:383–390

Alnaji, Fadi G. et al. (2019) Sequencing Framework for the Sensitive Detection and Precise Mapping of Defective Interfering Particle-Associated Deletions across Influenza A and B Viruses. Journal of Virology. 93:e00354-19

Massart, Sebastien, et al. (2019) Virus Detection by High-Throughput Sequencing of Small RNAs: Large-Scale Performance Testing of Sequence Analysis Strategies. Phytopathology. 109:488-497

Berry, Irina M., et al. (2019) Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity. The Journal of Infectious Diseases. 221:S292–S307

Vicedomini, Riccardo, et al. (2013) GAM-NGS: genomic assemblies merger for next generation sequencing. BMC Bioinformatics. 14(Suppl 7):S6

Response to Comment on Literature:

The tools, VirAmp, InteMAP, Bio-Docklets, and DNAscan are freely available, well-maintained, and could potentially see use by public health laboratories specializing in human pathogens. For these reasons, we have added mention of their respective publications, Wan et al. (2015), Lai et al. (2015), Kim et al. (2017), and Iacoangeli et al. (2019), to the Introduction and/or Discussion sections of the revised manuscript.

January 11, 2022

Dr. Margaret Okomo-Adhiambo
Centers for Disease Control and Prevention
Influenza Division
1600 Clifton Rd
Atlanta, GA 30329

Re: Spectrum02564-21R1 (VPipe: An Automated Bioinformatics Platform for Assembly and Management of Viral Next-Generation Sequencing Data)

Dear Dr. Margaret Okomo-Adhiambo:

Thank you for submitting your work to Microbiology Spectrum. Your revision has been reviewed by previous reviewers, who were quite happy with your revision. I am thus happy to notify you that your manuscript has been accepted. I am forwarding it to the ASM Journals Department for publication. You will be notified when your proofs are ready to be viewed.

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.

As an open-access publication, Spectrum receives no financial support from paid subscriptions and depends on authors' prompt payment of publication fees as soon as their articles are accepted. You will be contacted separately about payment when the proofs are issued; please follow the instructions in that e-mail. Arrangements for payment must be made before your article is published. For a complete list of **Publication Fees**, including supplemental material costs, please visit ourwebsite.

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.

Thank you for submitting your paper to Spectrum.

Sincerely,

Wei-Hua Chen
Editor, Microbiology Spectrum

Journals Department
American Society for Microbiology
1752 N St., NW
Washington, DC 20036
E-mail: spectrum@asmusa.org

Supplemental Material 1 of 1: Accept