The Journal of Open Source Software

stylo: a lightweight nanopore assembly pipeline optimized for enteric bacteria

Arzoo Patel [1,2*], **Mohit Thakur** [1*], **Justin Kim** [1,2], **Peyton Smith** [1], **Lee S. Katz** [1], **Curtis Kapsak** [1,3], **and Jessica Chen** [1]

1 Enteric Diseases Laboratory Branch, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America **2** ASRT Inc., Contractor for National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America **3** Theiagen Genomics, Highlands Ranch, Colorado, United States of America * These authors contributed equally.

DOI: 10.21105/joss.09695

Software

- Review
- Repository
- Archive

Editor: Claudia Solis-Lemus

Reviewers:

- @mberacochea
- @telatin

Submitted: 29 May 2025
Published: 05 June 2026

Summary

Oxford Nanopore Technologies (ONT) sequencing is a promising technology with many potential applications in food safety. We have developed stylo, a lightweight nf-core style assembly workflow for ONT long-reads, specifically optimized for enteric bacteria. The pipeline downsamples, assembles, and performs post-processing and quality control by combining nanoq (E Steinig & L Coin, 2022), Rasusa (M B Hall, 2022), Flye (M Kolmogorov et al., 2019), Dnaapler (Bouras et al., 2024), Dorado (Technologies, 2022), and QUAST (Mikheenko et al., 2018). All of stylo's dependencies are containerized and the pipeline is available on GitHub.

Statement of Need

There is a continuous need for foodborne outbreak detection in public health. To determine the scope or severity of a foodborne outbreak, short-read whole genome sequencing has often been used to generate isolate assemblies of enteric bacteria often which supports rapid and accurate outbreak detection (E M Ribot & K B Hise, 2016; R E Timme et al., 2017). However, as nanopore long-read sequencing becomes more cost-effective and accurate, the need increases for streamlined assembly pipelines to support high-throughput surveillance processing of ONT sequenced isolates (H H Mostafa, 2024; N D Sanderson et al., 2024). With the increased adoption of modern high-performance computing and cloud servers, pipelines built to leverage containerization and custom configurations allow for easy deployment on those servers. To address these needs, we have created stylo, a lightweight nf-core style nanopore assembly pipeline optimized for enteric bacteria (P A Ewels et al., 2020; P Di Tommaso et al., 2017).

Since stylo is lightweight, it can be run in automated disease surveillance settings, such as PulseNet, which facilitates the rapid detection of illness clusters and reduces the likelihood of outbreaks becoming large and widespread (B Tolar et al., 2019; P Gerner-Smidt et al., 2006). Stylo was designed and tested around enteric pathogens and default genome sizes were obtained through PulseNet Standard Operating Procedure ("PulseNet International SOPs," 2025), however, it does not have any strict limitation to enteric bacteria. This allows users to customize the lookup table to include any bacteria that they work with. There exist extensive nanopore assembly workflows such as Donut Falls (E Young & K Florek, 2025), which includes hybrid assembly options and multiple rounds of polishing. However, stylo is a streamlined nanopore workflow, intended for high throughput research groups, surveillance networks, or practical assemblies for general users.

Workflow Overview

1. Input: stylo requires a comma separated value file with columns for sample, fastq, genus, and species. Fastq files must comprise of long-reads generated on an ONT instrument. Genus and species are used to automatically determine genome size via a lookup table built into the pipeline.
2. Filtering and Downsampling: The pipeline filters out reads that are less than a user provided minimum length using nanoq. The resulting fastq is then subsampled to a user provided coverage via Rasusa.
3. Assembly: Flye is run on the subsampled fastq using the "–nano-hq" mode by default, expecting high-quality ONT reads. This parameter can be changed by the user.
4. Post-processing and Quality Control: The pipeline uses "dnaapler all" to reorient contigs to begin at a specific genes. The pipeline then uses "dorado polish" using "–bacteria" to correct assembly sequences. Finally, the assembly quality is assessed via QUAST with the following speedup options: "–no-check", "–no-plots", "–no-html", "–no-icarus", "–no-snps", "–no-sv".
5. Output: The pipeline outputs files for each step. Some key files are the assembly by Flye, the final corrected assembly by Dorado, and the quality control summary by QUAST. Additionally, MultiQC is run to generate an html report summarizing nanoq and QUAST outputs.

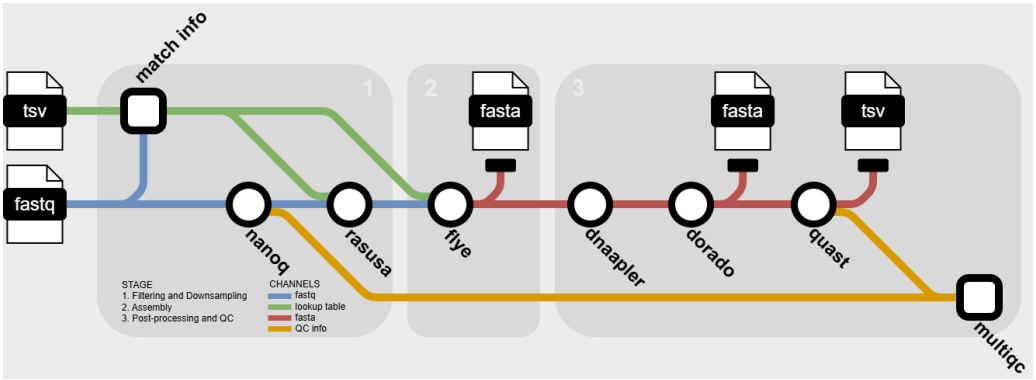

Figure 1: Diagram of stylo steps.

Availability

Stylo is freely available and open-source. It can be downloaded from the GitHub repository available at https://github.com/ncezid-biome/stylo.

Acknowledgements

These authors contributed equally as co-first: Arzoo Patel and Mohit Thakur. We acknowledge helpful discussions from Joe Wirth.

References

B Tolar, L A Joseph, M N Schroeder, S Stroika, E M Ribot, K B Hise, & P Gerner-Smidt. (2019). An overview of PulseNet USA databases. *Foodborne Pathogens and Disease*, *16*(7), 457–462. https://doi.org/10.1089/fpd.2019.2637

Bouras, G., Grigson, S. R., Papudeshi, B., Mallawaarachchi, V., & Roach, M. J. (2024). Dnaapler: A tool to reorient circular microbial genomes. *Journal of Open Source Software*, *9*(93), 5968. https://doi.org/10.21105/joss.05968

E M Ribot, & K B Hise. (2016). Future challenges for tracking foodborne diseases. *EMBO Rep.*, *17*(11), 1499–1505. https://doi.org/10.15252/embr.201643128

E Steinig, & L Coin. (2022). Nanoq: Ultra-fast quality control for nanopore reads. *Journal of Open Source Software*, *7*(69), 2991. https://doi.org/10.21105/joss.02991

E Young, & K Florek. (2025). Donut falls. In *GitHub repository*. GitHub. https://github.com/UPHL-BioNGS/Donut_Falls

H H Mostafa. (2024). An evolution of nanopore next-generation sequencing technology: Implications for medical microbiology and public health. *Journal of Clinical Microbiology*, *62*(5), e00246–24. https://doi.org/10.1128/jcm.00246-24

M B Hall. (2022). Rasusa: Randomly subsample sequencing reads to a specified coverage. *Journal of Open Source Software*, *7*(69), 3941. https://doi.org/10.21105/joss.03941

M Kolmogorov, J Yuan, Y Lin, & P A Pevzner. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*, 540–546. https://doi.org/10.1038/s41587-019-0072-8

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150. https://doi.org/10.1093/bioinformatics/bty266

N D Sanderson, K M V Hopkins, M Colpus, M Parker, S Lipworth, D Crook, & N Stoesser. (2024). Evaluation of the accuracy of bacterial genome reconstruction with oxford nanopore R10.4.1 long-read-only sequencing [Journal Article]. *Microbial Genomics*, *10*(5). https://doi.org/10.1099/mgen.0.001246

P A Ewels, A Peltzer, S Fillinger, H Patel, J Alneberg, A Wilm, M U Garcia, P Di Tommaso, & S Nahnsen. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, *38*, 276–278. https://doi.org/10.1038/s41587-020-0439-x

P Di Tommaso, M Chatzou, E W Floden, P P Barja, E Palumbo, & C Notredame. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*, 316–319. https://doi.org/10.1038/nbt.3820

P Gerner-Smidt, K Hise, J Kincaid, S Hunter, S Rolando, E Hyytiä-Trees, E M Ribot, & B Swaminathan. (2006). PulseNet USA: A five-year update. *Foodborne Pathogens and Disease*, *3*(1), 9–19. https://doi.org/10.1089/fpd.2006.3.9

PulseNet international SOPs. (2025). In *APHL repository*. APHL. https://www.aphl.org/programs/global_health/Pages/PulseNet-International-SOPs.aspx

R E Timme, H Rand, M Shumway, E K Trees, M Simmons, R Agarwala, S Davis, G E Tillman, S Defibaugh-Chavez, H A Carleton, W A Klimke, & L S Katz. (2017). Benchmark datasets for phylogenomic pipeline validation, applications for foodborne pathogen surveillance. *PeerJ*, *5*(e3893), e3893. https://doi.org/10.7717/peerj.3893

Technologies, O. N. (2022). Dorado. In *GitHub repository*. GitHub. https://github.com/nanoporetech/dorado