Presenter: Kathryn Morin

Institution: Centers for Disease Control and Prevention

Department: Division of Sexually Transmitted Disease Prevention

Mailing Address:
ROYBAL BLDG
1600 Clifton Rd
Atlanta, GA 30333

Phone number: 404-718-7715

Email address: xqc3@cdc.gov

neissflow: Streamlining Genomic Epidemiology of Neisseria gonorrhoeae with Nextflow

Kathryn Morin[1], Ethan Hetrick[2], Apurva Shrivastava[1], Eric Tran[3], Jack Cartee[4], Amanda Smith[4], Arvon A. Clemons[4], Katherine Hebrank[1], Kim M. Gernert[4], Matthew W. Schmerer[4], and Sandeep J. Joseph[4]

Affiliations:
1: Oak Ridge Institute for Science and Education, Oak Ridge Associated Universities, 1299 Bethel Valley Rd, Oak Ridge, TN 37830, USA
2: Leidos, Leidos Headquarters, 1750 Presidents St, Reston, VA 20190, USA
3. ASRT Inc., ASRT Inc., 4158 Onslow Pl SE. Smyrna, GA 30080, USA
4. Division of STD Prevention, Centers for Disease Control and Prevention, 1600 Clifton Rd, Atlanta, GA 30333, USA

Background The rising prevalence of antibiotic resistance (AR) in *Neisseria gonorrhoeae* (Ng), the causative agent of gonorrhea, poses a critical public health challenge. Effective lab-based surveillance and public health response to AR Ng depends on timely genetic characterization of Ng strain types and antimicrobial resistant (AMR) genetic markers. With expansion of whole genome sequencing (WGS) capacity in public health labs (PHLs), there is an urgent and specific need for an automated tool that facilitates standardized and efficient WGS analysis for Ng in PHLs.

Methods We developed neissflow, a Nextflow pipeline, by leveraging bioinformatic tools, Python and Bash scripts, and Singularity containers. The pipeline built to nf-core standards, ensures reproducibility and scalability. The pipeline is subdivided into 5 subworkflows with 2 automated quality control (QC) steps. The subworkflows perform intermediate analysis steps, preprocessing of raw reads, species identification, *de novo* assembly, AMR typing, and phylogenetic analysis, to modularize the pipeline. This allows the user to perform only the analysis their sample set requires. Automated QC checks include a read quality and a species and assembly quality check. The phylogenetic analysis subworkflow includes homologous recombination correction prior to maximum likelihood tree generation and outbreak cluster detection. The pipeline produces detailed reports and figures. Neissflow can be run from the command line or through Nextflow Tower, a graphical user interface developed by Sequera, on cloud or on-premises resources.

For preliminary testing of functionality and resource utilization of the pipeline, 469 Ng Illumina WGS were analyzed. Bacterial isolates were collected via CDC surveillance programs (2022-2024) and sequenced via

CDC's regional Antimicrobial Resistance Lab Network. Neissflow bioinformatic analyses were compared to previous in-house analysis pipelines. A small sample set (n=12) was used to test the clade analysis capabilities of neissflow and was thus run through the pipeline in its entirety. The full sample set (n=469) was used to test the resource utilization optimizations of the pipeline and check the accuracy of AMR typing, so the phylogenetic analysis step was skipped.

Results Neissflow is computationally scalable and thus maximizes the utilization of the available computational resources. Using CDC's on-premises cluster, the analysis of 12 samples (204 processes) completed in 35 min. If performed manually instead, these processes would run serially as opposed to parallelly, taking at least 2.75 hr. Using the same resources, the analysis of 469 samples (6534 processes) finished in 8.25 hr. These processes, if performed serially, would take >53.5 hr to complete. The pipeline has 100% concordance for species identification of known Ng samples determined through biochemical methods, and 100% concordance for GyrA S91 mutations compared to previous variant analysis. Neissflow accurately identified clades, verified by previous phylogenetic analysis and single nucleotide polymorphism (SNP) differences.

Conclusions Following internal testing and validation, neissflow could potentially be deployed through CDC's Advanced Molecular Detection (AMD) platform for PHL usage with proper authentications. Neissflow offers a scalable, automated, modular, and quality-managed solution for Ng WGS analysis by PHLs to enhance surveillance of AMR and monitor outbreaks.