



ID der Kurzfassung : 280

Development of a Nextflow Pipeline for Structural Variant Detection and Evaluation in German *Mycobacterium tuberculosis* Samples

Inhalt

Broadly utilized in bacterial pathogen surveillance, enabling infection chain and outbreak investigations, are clustering approaches that employ core genome multi-locus sequence typing (cgMLST) to define pairwise genetic distances.

For the *Mycobacterium tuberculosis* complex (MTBC), structural variants (SVs), particularly regions of difference (RDs), which are large-scale insertions and deletions relative to the H37Rv reference genome, have been described as stable phylogenetic markers. However, the presence of such structural variants may interfere with cgMLST-based typing. Specifically, deletions that overlap cgMLST loci can introduce missing values and skew genetic distances between isolates, justifying the exclusion or masking of certain cgMLST loci.

To investigate this issue, we designed Tubrd, a Nextflow-based workflow to evaluate the presence of known RDs and to detect novel structural variants from short-read sequencing data. We analyzed 8,000 MTBC samples collected within the scope of integrated genomic surveillance (IGS).

We anticipate that our results will stratify by lineage, improving the accuracy of molecular epidemiology in TB surveillance. Application of this pipeline to a selection of 100 MTBC isolates from IGS revealed that ~3% of established RDs impact cgMLST profiles, along with 20 novel SVs across the dataset that show a similar effect. These initial findings highlight the importance of assessing structural variation to enable accurate genome-based clustering.

Keywords

Bioinformatics, structural variants, *Mycobacterium tuberculosis*, cgMLST

Registration ID

112

Professional Status of the Speaker

Graduate Student

Junior Scientist Status

Yes, I am a Junior Scientist.

Track Klassifizierung: Genomics & Proteomics

Typ des Beitrags: Poster presentation