



ID der Kurzfassung : 397

Is This a Gene? The Impact of ORF Filtering on cgMLST-Based Clustering and Distance Thresholds for Genomic Surveillance

Inhalt

Core genome multilocus sequence typing (cgMLST) is a widely applied method for bacterial pathogen surveillance and outbreak investigations, in which pairwise genetic distances are commonly used to calculate relatedness between samples through single-linkage clustering. However, cluster definitions heavily depend on the underlying typing scheme and allele calling strategy.

Using datasets of *Mycobacterium tuberculosis* and *Listeria monocytogenes*, we evaluated the impact of open reading frame (ORF) filtering on cgMLST-based clustering. We directly compared clustering results calculated from complete and ORF-filtered cgMLST profiles generated by multiple established allele calling tools. For this, an element-centric similarity score was utilized.

We determined allelic distance thresholds required to obtain comparable clustering results across different allele calling methods. In summary, ORF-filtering reduced pairwise allelic distances between isolates by increasing cgMLST missingness. Consequently, complete profiles produced clusters with potentially increased genomic resolution compared to ORF-filtered profiles at fixed distance thresholds.

We observed that ORF-filtering substantially affects cgMLST-based distances and outbreak cluster definitions. Re-evaluating clustering thresholds based on each method enables a more consistent, meaningful intersectoral exchange regarding outbreak analyses and routine genomic surveillance.

Keywords

Bioinformatics, cgMLST, Clustering, Outbreak investigations

Registration ID

OHS26-70

Professional Status of the Speaker

Graduate Student

Junior Scientist Status

Yes, I am a Junior Scientist.

Autor: ROSCHEK, Caroline (Genome Competence Centre (MF1), Robert Koch Institute)

Co-Autoren: Frau DASSIO, Clara (Genome Competence Centre (MF1), Robert Koch Institute); Herr DRILLER POVILL, Maximilian (Genome Competence Centre (MF1), Robert Koch Institute); Dr. HOUWAART, Torsten (Genome Competence Centre (MF1), Robert Koch Institute); Dr. WOLF, Silver (Genome Competence Centre (MF1), Robert Koch Institute)

Vortragende(r): ROSCHEK, Caroline (Genome Competence Centre (MF1), Robert Koch Institute)

Track Klassifizierung: Genomics & Proteomics

Typ des Beitrags: Poster presentation

Eingereicht von **ROSCHEK, Caroline** <roschekc@rki.de> am **Freitag, 19. Juni 2026**