Day3 session 1.3- Mycobacterium tuberculosis NGS made easy: data analysis step-by-step

We would like you to think for a couple of minutes about what you have heard. Also, we would like to assess the quality of our teaching in order to improve the quality of this workshop.

Day 3 - Webinar "The concept of clustering & Genetic Distance thresholds"

Please briefly describe one situation where you could put the concept of clustering in practice and name one critical aspect of applying clustering to infer transmission. Thank you!

Name	Briefly describe one situation where you could put the concept of clustering in practice	Name one critical aspect of applying clustering to infer transmission	Indicate here concepts that you did not fully understood
Eva Sodja	Tuberculosis outbreak in nursing home	- Sampling (bias), imported cases - De novo vs transmitted resistance - resolution of clustering (genotyping)	

		approach	
Halima Said	1.To understand transmission dynamic 2. Outbreak investigation 3. Control and evaluate performance of a program 4. Strain diversity	1.Sampling 2.Time 3.Location (proximity)	
Justice Tresor NGOM NGOM	To understand the Nosocomial outbreak, we can use the concept of clustering		
Ilinca Memelis	In a healthcare setting (hospital or nursing home).	Defining the clusters and sampling them. The area/location. To understand the transmission.	
Farzana Ismail	Suspected outbreak of MDR TB in a school (or any other setting, e.g. old age home etc.)	Sampling, time, geographical area	
Justice Ohene	Tuberculosis	Applying	

Amofa	outbreak in rural and hotspot areas	resolution of clustering and genotyping in resolution. Engaging in De Novo Assembly processes	
Jagadeeshwari Uppada	Outbreak of TB in a particular city (TB hospitals) by comparing the samples from various health centers	It will help in tracking the transmission and also helps in analyzing the transmission of drug resistant strains	
Shamsuddeen Yusuf Ma’aruf	Suspected outbreak of tuberculosis in a remote area Suspected outbreak of tuberculosis in primary schools	Independent evolvement of mutation across phylogenetics	
Jeanne Crescence	One practical application of clustering would be to identify genetically distinct subpopulations of TB within a species based on their genetic	Geographical area, sampling method	-What clustering methods are commonly used to study the distribution and patterns of genetic markers? - How are genetic distance thresholds defined and used in

	marker profiles.		analyses? - What are the main factors taken into account when determining relevant genetic distance thresholds for comparative studies?
Kevim Bordignon Guterres	Zoonotic tuberculosis	Cases of tuberculosis in a Zoo	If genomic results differ from phenotypes? For example, if sample 1 contains a non-synonymous mutation that confers resistance to drug X, and this sample responds in vitro. Could this be related to population heterogeneity? Or microevolutionary?
Esto Bahizire	Back home we have high frequency of TB/MDR TB in artisanal miners	People in these areas are moving around then one might not be able to catch where transmission is	

		coming from. The same patients are going back in their family and communities where they could have got TB or bring it there from mining zones	
Juliana Maira Watanabe Pinhata	To investigate if TB cases in a prison are related	Sampling bias, case importation from another region	
Nazmi Mehmeti	In the context of public health, clustering analysis can be crucial in identifying and controlling the spread of infectious diseases such as tuberculosis. **Example Situation:** **Monitoring**	One critical aspect of applying clustering to infer transmission is selecting the appropriate SNP threshold. This threshold defines the maximum genetic distance between	**Continuum in SNP Distances:** The concept of how SNP distances can form a continuum in certain settings, making it challenging to distinguish between recent and older transmissions, is somewhat unclear. How exactly does this affect clustering decisions?

	Tuberculosis Transmission **Scenario:** Imagine a city health department is trying to understand the recent spread of tuberculosis (TB) to implement effective control measures. By performing clustering analysis on TB cases, they can determine which cases are linked by recent transmission and identify potential outbreak sources. **Steps:** 1. **Sample Collection and Sequencing:** Collect samples from TB patients and perform whole genome	bacterial strains considered to be linked by recent transmission. Choosing the right threshold is crucial because: 1. **Accuracy of Transmission Links:** A too-low threshold might miss connections, underestimating the outbreak's extent, while a too-high threshold can group unrelated cases, leading to false positives. 2. **Context Dependence:** Optimal thresholds	**Within-Host Diversity Impact:** The extent to which within-host diversity can skew clustering results, particularly how different genotypes within the same patient can impact the inference of transmission pathways, needs further clarification. How frequently does this issue arise, and what are its typical consequences?

	sequencing to obtain the bacterial DNA sequences. 2. **SNIP Analysis:** Calculate the single nucleotide polymorphism (SNP) distances between the genomes of the TB strains from different patients. 3. **Threshold Setting:** Define a SNP threshold to determine recent transmission links (e.g., cases with less than 10 SNP differences	vary based on the pathogen, mutation rate, and epidemiological context, requiring careful calibration to specific situations. For example, in tuberculosis, a commonly used SNP threshold might be 10 SNPs, but this may need adjustment based on the setting and strain characteristics.	

	are considered linked). 4. **Cluster Formation:** Group the cases into clusters based on SNP distances, identifying which patients are part of the same transmission chain. 5. **Interpretation and Action:** Use the clusters to pinpoint areas with ongoing transmission and target public health interventions, such as contact tracing, vaccination campaigns,		

	or heightened surveillance in those areas. **Outcome:** This approach helps public health officials identify transmission hotspots, understand the dynamics of the outbreak, and deploy resources more effectively to curb the spread of TB.		
Saidou MAHMOUDOU	Concept of clustering in the TB transmission and none transmission		
Michellin Baje	Outbreak of TB in prison Analyze clustering in provinces and islands like the Philippines	Appropriate sampling, and geographic location Consider taking medical history of participants	

Nneka Onyejepu	In a high MDR-TB State like Lagos Nigeria to understand transmission dynamics of primary MDR-TB in the community by sampling a subset of possible TB patients	Appraising the right SNP' s threshhold since some studies have shown a continuum in transmission threshold (above the commonly used SNP threshold 5,10 and 12SNPs)	
Caitlin Williams	Clustering can be used to understand transmission patterns in a community especially when it is combined with epidemiological data.	Adjusting SNP thresholds to fit the community being evaluated rather than sticking with 15 or 10 as your threshold.	I understood the content.
Kunle-Ope Chioma	When there an outbreak of TB disease, clustering can be applied to determine the chain of transmission	Sampling and Geographical Regions	I did not understand pitfalls of SNP Threshold, I will need more explanation
Hassan Ramadan Rajab	MTB outbreak near the border of South Sudan and Kenya	Population sample business may affect Mtb outbreak and transmission.	N/A

		Deciding the correct SNPs thresholds to confer clustering of individuals in a sample population	
Sofia Moraes	Clustering can be used to investigate the transmission patterns during a TB outbreak in a specific location such as a country /city/clinical setting/etc	Choosing the appropriate SNP threshold, which will define which strains are considered to be linked by transmission	
Dejenie Shiferaw			
Vanessa Falvo	For identify is a TB outbreak is due to a imported, reactivate o new mutation	SNP threshold definition Clustering is very much influenced by sampling. Lower sampling proportions and shorter sampling timeframes lead to lower clustering rates that shoud not be confounded with lack of transmission.	

Michael Reigh Guevarra	Identifying transmission dynamics between PLHIV with MTB coinfection receiving ART medicine in a tertiary hospital	Setting the statistically acceptable SNP threshold, thorough evaluation of medical history, doing background check do determine probability of transmission (geographical location of patients, finding common grounds)	The lesson was delivered properly. It is easy to understand.
Makonk Najah	Clustering can be used to group cases into strains based on differences in the number of SNPs	It can be applied to contact tracing.	
Alice Bontempi Bispo	Clustering can be used to understand the transmission profile in a public hospital for example, allowing us to get the data from the patients and	Even though we can assemble patients in one cluster because of the strain of phenotype, it is very important to access the genotype and	I understood the lesson, thank you.

	locating the main pathways of transmission of that strain.	understand what type of mutation occurred so we can better understand if we are dealing with, in the case of DR-TB, de novo evolution or transmitted DR, if the patients should or not be in the same cluster. What also has to be in our minds is the possibility of a latent tuberculosis patient having the same resistance but from a totally different mutation.	
Richard Ojedele	From my understanding, clustering is used to understand transmission patterns in a community especially when it is combined with epidemiological data.		

Nick Foster	Determining transmission patterns in isolated communities	Access of the genotype is important to having accurate and reproducible analyses	n/a
SOUAD ELMI	**Cluster analysis is a widely-used technique in data science and statistics, which aims to group similar objects within a dataset. By identifying these relationships, researchers and analysts can gain important insights into the underlying structure of the data, enabling better decision-making and more accurate predictions.**	1. The diameter of each cluster. 2. The sum of its branch lengths. 3. Chains of pairwise distances.	
Siti Ramli	DR TB outbreak in old folks home	1-De novo resistant vs transmitted resistance 2. resolution of clustering approach 3-SNP threshold determination	
Daniel	-TB in school	-SNP	-Within host diversity

Mekonnen	-TB in prison TB in -congregate setting	threshold/Training tool -sampling issue	-the culture isolates Vs sputum isolates diversity -The term recent interms of time -Transmission chain VS clustering
Buhari Yusuf	TB outbreak investigation, particularly involving isolates with varying genetic characteristics.	Genetic distance threshold	
Taime Sylvester	Investigation of possible transmission with a rise in MDR cases in an area	-Sampling bias -Ensuring accurate differentiation between recent transmission and reactivation of latent infections- based on the SNP threshold (10 SNPS).	Continuous transmission SNP threshold.
Bernice Twenewaa Sekyere	Identifying the outbreak of cholera in a remote area	Appropriate sampling in the geographic location	

