université PARIS-SACLAY

ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Title : Comparative genomics and phylogenomics of *Mycobacterium tuberculosis* with a special emphasis on *Mycobacterium africanum*

Keywords : *Mycobacterium tuberculosis* complex, L5, L6, *Mycobacterium africanum*, Evolutionary Genomics, Phylogeography, Nigeria

Abstract : *Mycobacterium africanum* (MAF) as a member of MTBC causes a substantial amount of human tuberculosis in West Africa, but little is known on their genetic diversity. There is low recovery rate of MAF resulting from difficulties in growing and isolation in routine practice due to metabolic specificities and deficiencies. Contrarily to other lineages such as L2 and L4, there are still scarce whole-genome-sequence (WGS) data on L5-L6, the two different lineages within *Mycobacterium africanum* in public genome repositories. For this reason, we highlight the need for sequencing additional MTBC genomes from Nigeria. After acceptance of our study by a local ethical committee, we extracted the DNA of 325 MTBC clinical isolates from patient suffering from tuberculosis in Northern Nigeria. The goal was to screen L5 and L6 from this region to study (1) the northern Nigeria MTBC diversity, (2) L5 taxonomy and phylogenetics, (3) L5 and L6 geographical distribution and potential links with human genetics, and (4) the presence in Nigeria of the L4.6.2.2 sublineage Cameroon. We provide an in-depth analysis of the phenotypic drug-resistance, screened L5 L6 by MIRU-VNTR locus 24 analysis, and using line-probe assay as a confirmation of strain identity. We obtained WGS of 33 isolates from Nigeria belonging to L4, L5, and L6. Only 6(out of 325) equivalent to 1.85% represented L5 or L6 in the study area. More than 75% were MDR isolates based on Phenotypic drug-susceptibility testing results. Bioinformatical analysis using *TB profiler*, *TB annotator*, and *Genotube* pipelines of the new sequences Read Archives (SRAs) to previously published ones from available Bioprojects run in Nigeria was performed. We also performed RAxML phylogenetic reconstructions of larger samples that include public NCBI SRAs from some neighboring countries (Cameroon, Ghana). To confront phylogenetic reconstruction to metadata, we used the new *TB-Annotator* pipeline. We propose a new L5 taxonomical scheme with more than 12 sublineages among which 6 are new. The scheme allow us to classify 97-98% of isolates as compared to 82-83% formerly used. Our results also suggest an alarming rate of drug resistance of the L4.6.2.2 lineage and a high diversity of L5. There is a prevalence rate of 70-80% Rifampin resistance and 41-45.5% Isoniazid resistance. There is no link between drug resistance, geographical location, and clinical characteristics of patients. One L2/Beijing isolate (pre-XDR) from one Nigerian state was found in Yobe. L4.6.2.2 was found to be RIF and INH resistant in Nigeria. L6 is weakly represented and *Mycobacterium africanum* lineages are less associated with multiresistance. L5 genomes in Northern Nigeria belong to new clades. We also describe the L4.6.2.2 lineage in Nigeria, Cameroon, and Ghana. We provide computations on the divergence time of L4.6.2 and suggest a new hypothesis on its origin .