

Description of a new ancestral sublineage of the *Mycobacterium tuberculosis* complex Lineage 2 in Japan

Christophe Guyeux¹, Gaetan Senelle¹, Guislaine Refrégier^{2,3} Emmanuelle Cambau^{4,5}, Christophe Sola^{2,4*}

¹FEMTO-ST Institute, UMR 6174 CNRS, DISC Computer Science Department, Univ. Bourgogne Franche-Comté (UBFC), 16 Route de Gray, 25000 Besançon, ²Université Paris-Saclay, Saint-Aubin, France, ³Université Paris-Saclay, CNRS, AgroParisTech, UMR ESE, 91405, Orsay, France, ⁴Université de Paris, IAME, UMR1137, INSERM, Paris, France, ⁵AP-HP, Laboratoire associé au Centre National de Référence Tuberculose, Paris, France.

The “Beijing” lineage 2 remains an important threat of tuberculosis worldwide. By gathering a growing collection of 135 public SRAs from the *Mycobacterium tuberculosis* complex (MTC) belonging to the L2 lineage, and using a new bioinformatic pipeline, “TB-Annotator”, that analyzes more than 50,000 characters, we describe in central Japan a new L2 sublineage. These isolates harbor a number of specific criteria including an adaptative mutation in *rpoC*, a phylogenetic SNP in Rv3401, various spoligotype and MIRU-VNTR profiles, and share *IS6110* copies. They are distant by pairwise genome comparison between 200 to 400 SNPs, and do not fulfill, either the “Modern Beijing” lineage criteria (L2.2.1.2.2) or the other currently known “Ancestral Beijing lineages” described so far in the literature. We named this new clone as “Asia Ancestral 5”. Using this study, we confirm the specific intermediate position of the “Asia Ancestral 4” recently described. The Asia Ancestral 4 possesses intermediate “modern Beijing” characteristics such as the classical *mutT2* SNP at codon 58 but not the *ogt* SNP at codon 12, and as such could be better renamed as *intermediate*. Asia Ancestral 5 could be older than Asia Ancestral 3 and 4. This study shows that the history of the tuberculosis L2 pandemia in Asia is likely to reveal complex local historical patterns since its emergence, and suggests that the global epidemic history of L2 remains to be deciphered in more details.

*christophe.sola@inserm.fr