

## **Subspecies-specific sequence detection for differentiation of *Mycobacterium abscessus* complex**

Alina Minias <sup>1\*</sup>, Lidia Żukowska <sup>2</sup>, Jakub Lach <sup>3</sup>, Tomasz Jagielski <sup>4</sup>, Dominik Strapagiel <sup>3</sup>, Su-Young Kim <sup>5</sup>, Won-Jung Koh <sup>#5</sup>, Heather Adam <sup>6</sup>, Ruth Bittner <sup>6</sup>, Sara Truden <sup>7</sup>, Manca Žolnir-Dovč <sup>7</sup>, Jarosław Dziadek <sup>1</sup>

<sup>1</sup> Institute of Medical Biology, Polish Academy of Sciences, Lodz, Poland

<sup>2</sup> BioMedChem Doctoral School of the University of Lodz and the Institutes of the Polish Academy of Sciences in Lodz, Lodz, Poland

<sup>3</sup> Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Lodz, Lodz, Poland

<sup>4</sup> Department of Medical Microbiology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Warsaw, Poland

<sup>5</sup> Division of Pulmonary and Critical Care Medicine, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

<sup>6</sup> Diagnostic Services, Shared Health, Manitoba, Canada

<sup>7</sup> National Reference Laboratory for Mycobacteria, University Clinic of Respiratory and Allergic Diseases, Golnik, Slovenia

<sup>#</sup> deceased

\*Corresponding author: Alina Minias; ul. Lodowa 106, 93-232 Lodz, Poland; tel: +48 42 272 36 06, fax: +48 42 27 23 630, email: [aminias@cbm.pan.pl](mailto:aminias@cbm.pan.pl)

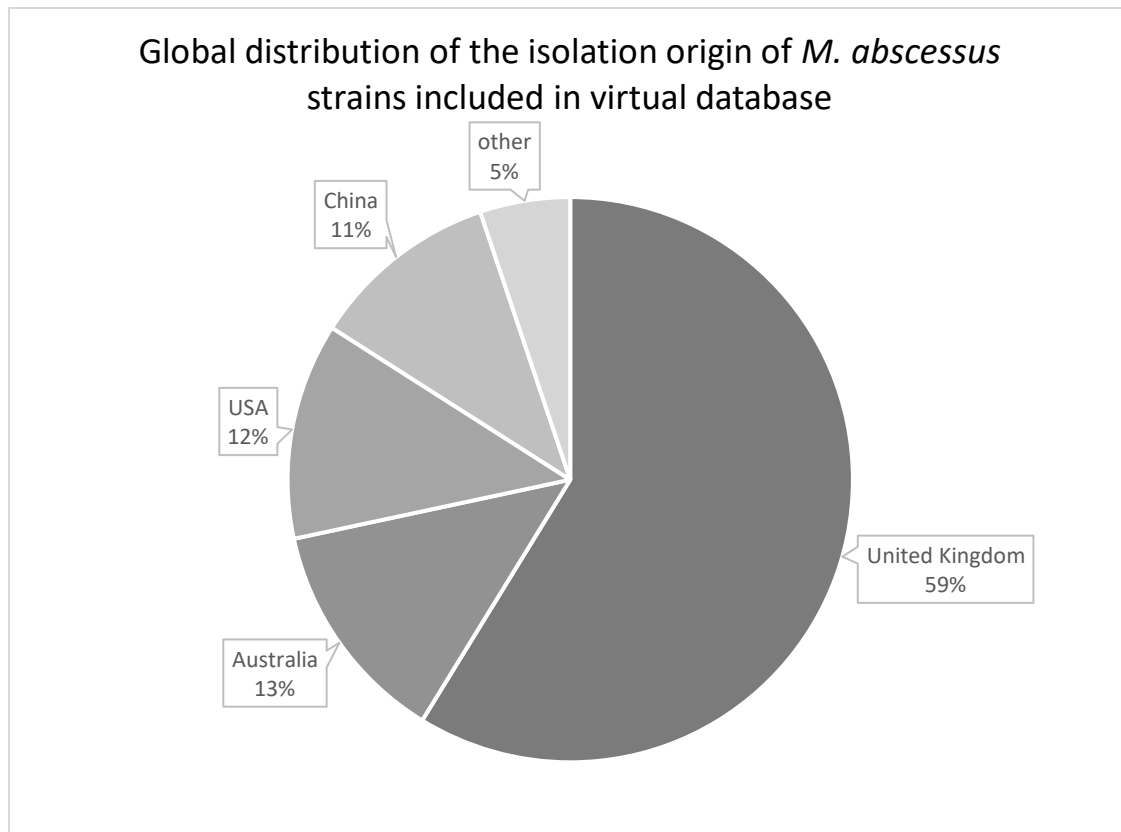


Fig. S1. The geographic origin of the *M. abscessus* strains included in virtual database. The majority of strains included in the database were isolated in United Kingdom. Four continents- Europe, Australia, North America and Asia were represented in the database by at least ten percent of the strains each.

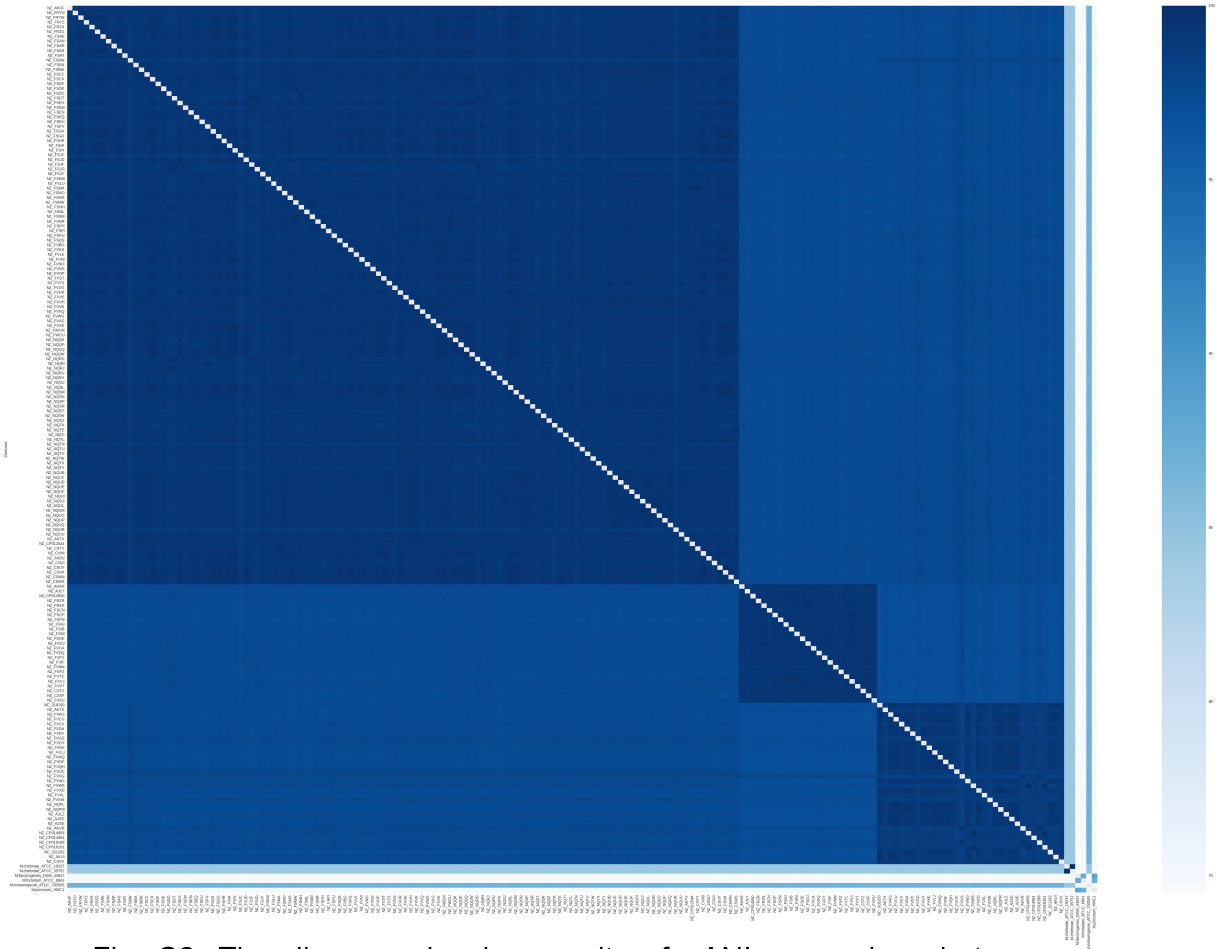


Fig. S2. The diagram showing results of gANI comparison between individual strains of *M. abscessus* complex strains for which the results of differentiation by different methods were inconsistent. The dataset included reference strains for each subspecies- NC\_018150 representing *M. abscessus* subsp. *massiliense*, NZ\_CP014950 representing *M. abscessus* subsp. *bolletii* and NC\_010397 representing *M. abscessus* subsp. *abscessus*. The blue band on the right of the chart shows the level of similarity between the strains, where dark color represents high similarity of the genomes and light color represents less similar genome sequences.

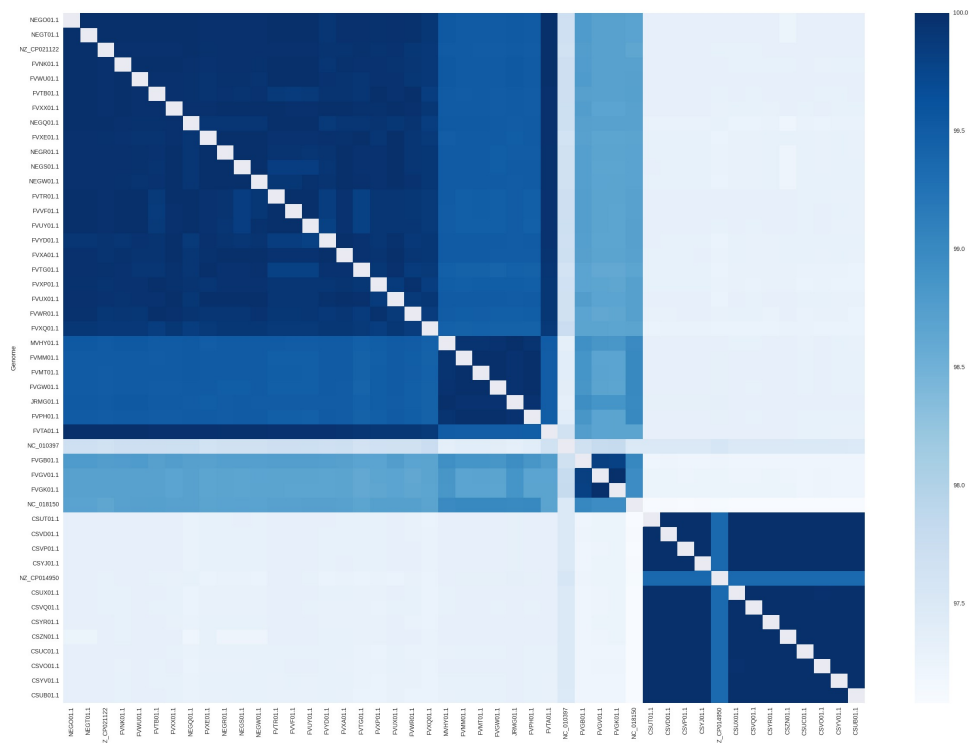


Fig. S3. The diagram showing results of gANI comparison between individual strains of *M. abscessus* complex strains for which the results of differentiation by different methods were inconsistent. The dataset included reference strains for each subspecies- NC\_018150 representing *M. abscessus* subsp. *massiliense*, NZ\_CP014950 representing *M. abscessus* subsp. *bolletii* and NC\_010397 representing *M. abscessus* subsp. *abscessus*. The blue band on the right of the chart shows the level of similarity between the strains, where dark color represents high similarity of the genomes and light color represents less similar genome sequences.

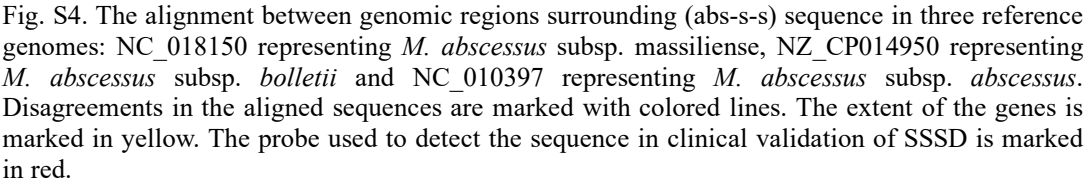


Fig. S4. The alignment between genomic regions surrounding (abs-s-s) sequence in three reference genomes: NC\_018150 representing *M. abscessus* subsp. massiliense, NZ\_CP014950 representing *M. abscessus* subsp. bolletii and NC\_010397 representing *M. abscessus* subsp. abscessus. Disagreements in the aligned sequences are marked with colored lines. The extent of the genes is marked in yellow. The probe used to detect the sequence in clinical validation of SSSD is marked in red.



Fig. S5. The alignment between genomic regions surrounding (*bol-s-s*) sequence in three reference genomes: NC\_018150 representing *M. abscessus* subsp. *massiliense*, NZ\_CP014950 representing *M. abscessus* subsp. *bolletii* and NC\_010397 representing *M. abscessus* subsp. *abscessus*. Disagreements in the aligned sequences are marked with colored lines. The extent of the genes is marked in yellow. The probe used to detect the sequence in clinical validation of SSSD is marked in red.



Fig. S6. The alignment between genomic regions surrounding (*mas1-s-s*) sequence in three reference genomes: NC\_018150 representing *M. abscessus* subsp. *massiliense*, NZ\_CP014950 representing *M. abscessus* subsp. *bolletii* and NC\_010397 representing *M. abscessus* subsp. *abscessus*. Disagreements in the aligned sequences are marked with colored lines. The extent of the genes is marked in yellow. The probe used to detect the sequence in clinical validation of SSSD is marked in red.

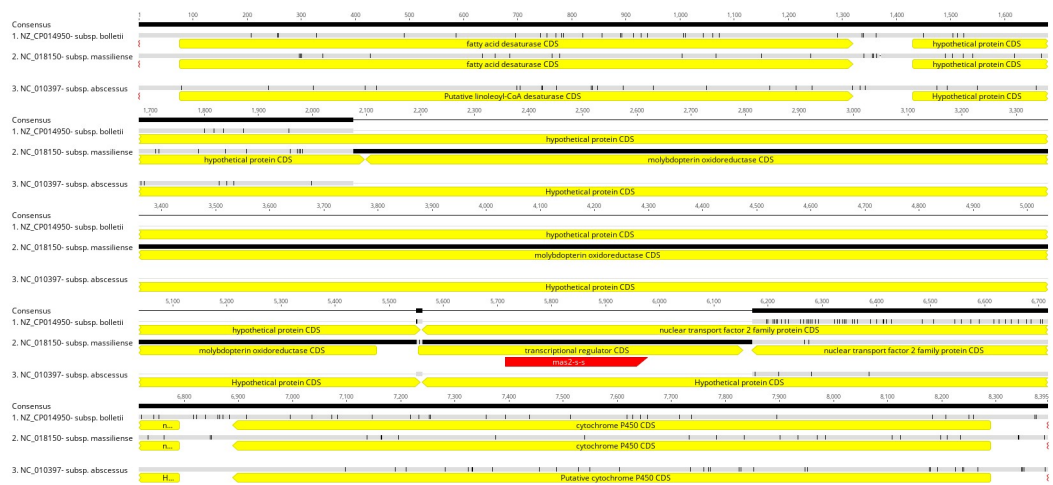


Fig. S7. The alignment between genomic regions surrounding (*mas2-s-s*) sequence in three reference genomes: NC\_018150 representing *M. abscessus* subsp. *massiliense*, NZ\_CP014950 representing *M. abscessus* subsp. *bolletii* and NC\_010397 representing *M. abscessus* subsp. *abscessus*. Disagreements in the aligned sequences are marked with colored lines. The extent of the genes is marked in yellow. The probe used to detect the sequence in clinical validation of SSSD is marked in red.