



**HAL**  
open science

## Contribution of machine learning for subspecies identification from *Mycobacterium abscessus* with MALDI-TOF MS in solid and liquid media

Alexandre Godmer, Lise Bigey, Quentin Gai Gianetto, Gautier Pierrat, Noshine Mohammad, Faiza Mougari, Renaud Piarroux, Nicolas Veziris, Alexandra Aubry

### ► To cite this version:

Alexandre Godmer, Lise Bigey, Quentin Gai Gianetto, Gautier Pierrat, Noshine Mohammad, et al.. Contribution of machine learning for subspecies identification from *Mycobacterium abscessus* with MALDI-TOF MS in solid and liquid media. *Microbial Biotechnology*, 2024, 17 (9), pp.e14545. 10.1111/1751-7915.14545 . pasteur-04854502

**HAL Id: pasteur-04854502**

**<https://pasteur.hal.science/pasteur-04854502v1>**

Submitted on 23 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.



L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

## RESEARCH ARTICLE

# Contribution of machine learning for subspecies identification from *Mycobacterium abscessus* with MALDI-TOF MS in solid and liquid media

Alexandre Godmer<sup>1,2</sup>  | Lise Bigey<sup>1,3</sup> | Quentin Gai-Gianetto<sup>4,5</sup> |  
Gautier Pierrat<sup>2</sup> | Noshine Mohammad<sup>6</sup>  | Faiza Mougari<sup>7</sup> | Renaud Piarroux<sup>6</sup> |  
Nicolas Veziris<sup>1,2,8</sup> | Alexandra Aubry<sup>1,8</sup>

<sup>1</sup>U1135, Centre d'Immunologie et des Maladies Infectieuses (Cimi-Paris), Sorbonne Université, Paris, France

<sup>2</sup>AP-HP, Sorbonne Université (Assistance Publique Hôpitaux de Paris), Département de Bactériologie, Groupe Hospitalier Universitaire, Sorbonne Université, Hôpital, Paris, France

<sup>3</sup>DER (Département d'Enseignement et de Recherche) de Biologie, ENS Paris-Saclay, Université Paris-Saclay, Gif-sur-Yvette, France

<sup>4</sup>Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics HUB, Paris, France

<sup>5</sup>Institut Pasteur, Université Paris Cité, Proteomics Platform, Mass Spectrometry for Biology Unit, UAR CNRS 2024, Paris, France

<sup>6</sup>Inserm, Institut Pierre-Louis d'Epidémiologie et de Santé Publique, IPLESP, AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de Parasitologie-Mycologie, Sorbonne Université, Paris, France

<sup>7</sup>Service de Mycobactériologie spécialisée et de référence, Centre National de Référence des Mycobactéries (Laboratoire associé), APHP GHU Nord, Université Paris Cité, INSERM IAME UMR, Paris, France

<sup>8</sup>AP-HP, Sorbonne Université (Assistance Publique Hôpitaux de Paris), Centre National de Référence des Mycobactéries et de la Résistance des Mycobactéries aux Antituberculeux, Paris, France

## Correspondence

Alexandra Aubry, U1135, Centre d'Immunologie et des Maladies Infectieuses (Cimi-Paris), Sorbonne Université, Paris, France.

Email: [alexandra.aubry@sorbonne-universite.fr](mailto:alexandra.aubry@sorbonne-universite.fr)

## Funding information

Annual grant from National Public Health Agency (Santé Publique France)

## Abstract

*Mycobacterium abscessus* (MABS) displays differential subspecies susceptibility to macrolides. Thus, identifying MABS's subspecies (*M. abscessus*, *M. bolletii* and *M. massiliense*) is a clinical necessity for guiding treatment decisions. We aimed to assess the potential of Machine Learning (ML)-based classifiers coupled to Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) MS to identify MABS subspecies. Two spectral databases were created by using 40 confirmed MABS strains. Spectra were obtained by using MALDI-TOF MS from strains cultivated on solid (Columbia Blood Agar, CBA) or liquid (MGIT®) media for 1 to 13 days. Each database was divided into a dataset for ML-based pipeline development and a dataset to assess the performance. An in-house programme was developed to identify discriminant peaks specific to each subspecies. The peak-based approach successfully distinguished *M. massiliense* from the other subspecies for strains grown on CBA. The ML approach achieved 100% accuracy for subspecies identification on CBA, falling to 77.5% on MGIT®. This study validates the usefulness of ML, in particular the Random Forest algorithm, to discriminate MABS subspecies by MALDI-TOF MS. However, identification in MGIT®, a medium largely used in mycobacteriology laboratories, is not yet reliable and should be a development priority.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Microbial Biotechnology* published by John Wiley & Sons Ltd.

## INTRODUCTION

Non-tuberculous mycobacteria (NTM) are opportunistic human pathogens associated with a variety of infections (Henkle & Winthrop, 2015). Such NTMs frequently include strains of *Mycobacterium abscessus* (MABS) that are primarily responsible for skin and lung infections (Der Werf et al., 2014; Forbes et al., 2018; McGrath et al., 2010; Mougari et al., 2016) and exhibit multiresistance, which makes them particularly challenging to treat. Antibiotic treatment with macrolides such as clarithromycin (CLA) or azithromycin (AZI) is the cornerstone of MABS infection therapy (Forbes et al., 2018; Nessar et al., 2012). However, MABS comprises three subspecies: *M. abscessus* subsp. *abscessus* (*M. abscessus*), *M. abscessus* subsp. *massiliense* (*M. massiliense*) and *M. abscessus* subsp. *bolletii* (*M. bolletii*), each exhibiting distinct CLA/AZI susceptibility profiles (Bastian et al., 2011; Brown-Elliott et al., 2015; Jeon et al., 2009; Koh et al., 2011, 2017; Nash et al., 2009; Richard et al., 2020; Yoshida et al., 2015). Clinical strains of *M. bolletii*, as well as approximately 70%–80% of *M. abscessus* strains, possess an inducible erythromycin ribosome methyltransferase encoded by the *erm*(41) gene, conferring inducible macrolide resistance (Bastian et al., 2011; Brown-Elliott et al., 2015; Koh et al., 2017). In contrast, most strains of *M. massiliense* are naturally susceptible to CLA due to non-functional *erm*(41) gene (Bastian et al., 2011; Brown-Elliott et al., 2015). These genetic differences lead to differences in the response to treatment (Harada et al., 2012; Huang et al., 2010).

In clinical laboratories, liquid culture methods are widely employed for promoting mycobacterial growth and expediting microbiological diagnosis (Pfyffer et al., 1997). Nowadays, the identification of MABS subspecies is performed by molecular methods, such as PCR-reverse hybridization using commercial kits or multiple gene sequencing. Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) is a powerful tool for the identification of bacteria (Faron et al., 2015; Hou et al., 2019), but the identification of mycobacteria, particularly MABS, has certain limitations and commercial systems such as MALDI Biotyper® (MBT) (Bruker®) and VITEK® MS (bioMérieux) are unable to distinguish the different MABS subspecies (Brown-Elliott et al., 2019; Fangous et al., 2014; Panagea et al., 2015; Suzuki et al., 2015; Teng et al., 2013; Tseng et al., 2013; Wassilew et al., 2017). The identification of mycobacteria using MALDI-TOF MS is challenging because of the distinctive composition of the mycobacterial wall, which necessitates a complex mechanical and chemical extraction protocol (Alcaide et al., 2018; Marrakchi et al., 2014). Also, liquid media such as MGIT® (Mycobacteria Growth Indicator

Tube, BD, Sparks, MD, USA) have the disadvantage of achieving lower rates of correct identification by MALDI-TOF MS (Balázová et al., 2014; Buchan et al., 2014; Kehrmann, Schoerding, et al., 2016; Lotz et al., 2010). Efforts to improve the identification of MABS subspecies have mainly focused on the search for distinguishing peaks in MALDI-TOF MS spectra (Fangous et al., 2014; Panagea et al., 2015; Suzuki et al., 2015; Teng et al., 2013; Tseng et al., 2013). However, this presents reproducibility challenges, as the presence of these distinguishing peaks varies depending on the growth medium used for mycobacteria (Popović et al., 2021) and the technique fails to reliably differentiate *M. bolletii* from *M. abscessus* (Kehrmann, Wessel, et al., 2016; Suzuki et al., 2015).

Artificial intelligence (AI) approaches, however, have the potential to enhance bacterial identification through MALDI-TOF MS. Of the AI approaches, Machine Learning (ML) focuses on developing algorithms that learn from datasets to enhance performance in resolving specific problems based on the processed data. The utilization of ML-based algorithms, known as classifiers, has improved the performance of MALDI-TOF MS for bacterial identification (Godmer et al., 2021; Rodríguez-Temporal et al., 2023). An alternative approach of ML methodology would be employing a metaclassifier to generate a final prediction by aggregating outputs from several classifiers. With regard to the identification of MABS subspecies, Rodríguez-Temporal et al. evaluated several ML-based algorithms to improve the performance of subspecies identification. They showed that the Random Forest (RF) algorithm, which is based on decision trees, achieved a high level of accuracy (95.9%) in the identification of MABS subspecies using MALDI-TOF MS from strains grown on different solid media (Rodríguez-Temporal et al., 2023). Furthermore, studies are needed to confirm the effectiveness of these ML techniques, not only for strains grown on solid media but also for strains grown on liquid media, which are widely used in laboratories.

In this study, we aimed to evaluate the ability of several ML-based approaches associated with MALDI-TOF MS to improve the identification of MABS subspecies from solid and liquid media.

## EXPERIMENTAL PROCEDURES

### Bacterial isolates and cultures

Thirty-seven clinical isolates from the French National Reference Center of Mycobacteria collection belonging to all MABS subspecies were used: 5 *M. bolletii*, 11 *M. massiliense* and 21 *M. abscessus* (Table S1). Three references strains (*M. bolletii* CIP 108541, *M. abscessus* T28 CIP 104536 and *M. abscessus* C28 CR 5701) were also included in our study (Table S1). Identification was

performed by using the GenoType NTM-DR (Hain Lifescience, Nehren, Germany) line probe assay.

The classification of mycobacteria has undergone significant changes in recent years, following an important study in 2018 (Gupta et al., 2018). This study proposed that mycobacteria should be divided into five groups: *Mycobacterium*, including *tuberculosis* complex and other species such as those in the *M. avium* complex; *Mycobacteroides*, focusing on subspecies of *M. abscessus*; *Mycolicibacterium*, including mycobacteria in the *M. fortuitum* and *M. smegmatis* complexes; *Mycolicibacter*, highlighting *M. terrae*; and *Mycolicibacillus*, including *M. trivalis*. Although initially accepted by some scientific journals and databases, this new classification was controversial (Meehan et al., 2021; Oren & Garrity, 2020; Tortoli et al., 2019). Eventually, the International Journal of Systematic Evolutionary Microbiology reverted to *Mycobacterium* as the main genus. In this context, we decided to follow traditional nomenclature, keeping *Mycobacterium* as the main genus of our study (Meehan et al., 2021). Regarding the subspecies of *M. abscessus* complex, we propose to name *Mycobacterium abscessus* subsp. *abscessus* as *Mycobacterium abscessus*, *Mycobacterium abscessus* subsp. *bolletii* as *Mycobacterium bolletii* and *Mycobacterium abscessus* subsp. *massiliense* as *Mycobacterium massiliense* in order to simplify the reading of the manuscript.

Strains were subcultured from frozen stocks in aerobic atmosphere at 37°C on solid media (Columbia blood agar +5% horse blood [CBA, bioMerieux®, France]) and further plated on different media to prepare the MALDI-TOF samples. First, the strains from the CBA medium were collected at one time point after 4 to 7 days of incubation. Second, from the CBA subculture, an initial inoculum of 0.5 McFarland was made and 1 mL was inoculated in MGIT® (Mycobacteria Growth to Indicator Tube) vials (BD, Sparks, MD, USA) supplemented with 5% OADC (BD, Sparks, MD, USA) and antimicrobials (polymyxin B, amphotericin B, nalidixic acid, trimethoprim and azlocillin). The strains from the MGIT medium were collected at one time point after 1 to 2 days of incubation.

## MALDI-TOF MS sample preparation

One mL and one loopful of a calibrated microbial loop of 1 µL were collected for each strain from both the MGIT® and CBA media, respectively. The protein extraction was performed using the MycoEx® protocol (Bruker® Daltonics, Bremen, Germany) (Pastrone et al., 2023). Rough strains had an additional step of 2 minutes with ultrasonic Sonifier (Vibra-Cell™, VC 505, Newtown, USA) to dissociate the aggregates. For each extraction, a total of 8 deposits were performed.

Dried spots were overlaid with 1 µL of MALDI matrix (10 mg/mL of α-cyano-4-hydroxy-cinnamic acid [α-HCCA] in 50% acetonitrile-2.5% trifluoroacetic acid; Bruker® Daltonics, Bremen, Germany).

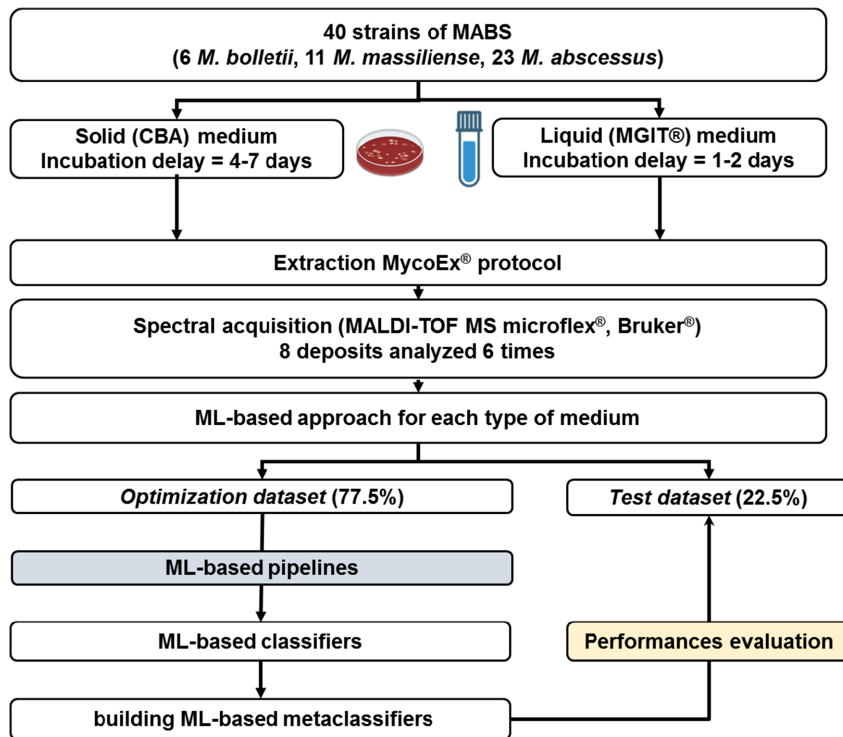
## Mass spectrum acquisition and analysis

Mass spectra were acquired with a Microflex® LT instrument (Bruker® Daltonics, Bremen, Germany) using the default parameters of the standardized CE-IVD (Conformité Européenne – *In Vitro* Diagnostic) method recommended by Bruker®. Each deposit was analysed 6 times, corresponding to 48 spectra per strain. An external calibration standard (Bacterial Test Standard; Bruker® Daltonics, Bremen, Germany) was used for mass calibration.

According to the Bruker® recommendations for building a database, spectra from the same strain were visually analysed to discard those of poor quality (Fergusson et al., 2020; Kostrzewa & Maier, 2017). If less than 5 correct spectra were obtained from the 48 spectra produced, protein extraction was repeated. The spectra were then imported into R with the MALDIForeign package (Gibb & Franceschi, 2022). The post-acquisition signal processing steps were carried out in the R environment using the MSclassifR and MALDIquant packages (Gibb & Strimmer, 2012; Godmer et al., 2022). The following pipeline was used: (i) square root intensity transformation, (ii) spectrum smoothing (Undecimated Wavelet Transform [UDWT] algorithm), (iii) baseline processing (SNIP for statistics-sensitive non-linear iterative peak-clipping algorithm), (iv) intensity calibration (TIC for Total Ion Current algorithm) and (v) spectrum alignment (500 ppm) using a standard spectrum obtained from 15 deposits of the reference strain (named *M. abscessus* T28 CIP 104536) of MABS and peak selection with signal-to-noise (S/N) greater than 3. All the generated spectra were examined and analysed using PCA for the culture media and subspecies.

## Predictive models and validation by using machine learning

After post-acquisition signal processing, the spectra from strains were separated according to the medium (Figure 1). Next, spectra from each medium were randomly divided into two datasets for each database: the “*optimization dataset*” corresponding to the spectra from 77.5% of the strains ( $n=31$ ) was used to train the ML-based classifiers and to evaluate the performances of different ML-based pipelines (see below) and the “*test dataset*” corresponding to the spectra from 22.5% of the strains ( $n=9$ ) was used to estimate the performances of the ML-based classifiers.



**FIGURE 1** Workflow of the Machine Learning-based approach used in this study. CBA, Columbia Blood Agar; d, day; MABS, *Mycobacterium abscessus*; MALDI-TOF MS, Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry; MGIT®, Mycobacteria Growth Indicator Tube; ML, Machine Learning.

The spectra from the “*optimization dataset*” were divided into two subdatasets: the “*train dataset*” (70% of the spectra or 995 spectra) was used to train the ML-based algorithms, and the “*validation dataset*” (30% of the spectra or 427 spectra) was used to measure the performance of the generated ML-based classifiers from the different ML-based pipelines. A total of 12 ML-based pipelines was used to create various ML-based classifiers using the MSclassifR R package (Godmer et al., 2022). The ML-based pipelines were trained according to the culture medium. The first step aimed to eliminate non-informative peaks (corresponding mainly to background noise) using three ML algorithm methods: (i) Recursive Feature Elimination coupled to Random Forest (RFE-RF) to iteratively eliminate non-discriminant peaks using a Random Forest (RF) algorithm; (ii) selection of discriminant peaks statistically (SelectionVarStat) using an analysis of variance test to determine whether the intensities of peaks were significantly different between groups; (iii) sparse partial least squares discriminant analysis (sPLS-DA) with  $k$ -folds cross-validation ( $k=5$ ). In this process, in a first iteration, the first fold ( $k=1$ ) is used as a test for the algorithm, while the others ( $k=4$ ) are used for training; this process is then repeated until every fold has been used as a test set. The second step to evaluate the performance of the ML-based pipelines consisted of injecting the previously selected informative peaks into ML algorithms trained with  $k$ -folds cross-validation ( $k=5$ ). Four ML algorithms were implemented in this study: linear regression (multinom), single-hidden-layer neural network (nnet), RF (Random Forest) and SVM (linear Support Vector Machine). According to

the predictions of each ML algorithm from the different pipelines, the performance (Cohen's Kappa coefficient, called Kappa coefficient in this study) was estimated for each “*validation dataset*” for each ML pipeline. Each ML-based pipeline was run 10 times with randomly searched hyperparameters (which determined the behaviour of the model). Due to imbalanced data, three resampling methods were used during the training process: the down-sampling method, which randomly removes spectra from majority classes; the up-sampling method, which randomly replicates spectra from minority classes; and the Synthetic Minority Sampling TEchnique (SMOTE), which uses an ML algorithm ( $K$ -nearest neighbours) to generate new spectra for the minority classes. The combination of the ML-based methods was labelled as “Selection variables Method-ML-algorithm”.

We developed 2 meta-classifiers (called Metaclassifier-ML-CBA or Metaclassifier-ML-MGIT, according to the medium) that combine predictions using averages to produce a final prediction using a ‘soft voting’ approach. The performances of these meta-classifiers were evaluated on the “*test dataset*”. In constructing these meta-classifiers, the following approach was applied according to each type of medium: selection of the best-performing ML-based algorithm identified by statistically significant and mostly higher Kappa coefficients compared to other pipelines. If none of the ML-based algorithms performed better than the others, a combination of ML-based classifiers from the ML-based pipeline whose classifier had obtained the best observed performance was carried out. These meta-classifiers were

called Metaclassifier-ML-CBA or Metaclassifier-ML-MGIT, according to the medium.

## Evaluation of the performance of the ML-based pipelines and statistical analysis

To evaluate comprehensively the ML-based classifiers within the ML-based pipelines, several metrics were used to estimate performances on the “*test dataset*”: accuracy (proportion of correctly classified spectra out of the total), sensitivity (corresponding to the proportion of well-identified spectra in this study per subspecies) and the reliability of identification (represents the percentage of certainty of correct identification and corresponding to the positive predictive value).

Metrics specific to the ML field were also used to account for unbalanced data: the Kappa coefficient (measuring inter-rater reliability and ranging from  $-1$  [complete disagreement] to  $1$  [complete agreement]) and F1-score (i.e. the harmonic mean of positive predictive value [referred to as precision in the ML field] and recall [referred to as sensitivity in the ML field]). The performances of the different methods (measured by the Kappa coefficient, which assesses the reliability of identification between each ML-based classifiers and the molecular reference method) were compared by considering ML-based pipelines, selection variables methods and ML-based algorithms using the one-tailed Wilcoxon signed ranks test with the Benjamini–Hochberg correction.

## Specific peak analysis for differentiation of subspecies

Specific peaks with a frequency  $\geq 95\%$  in at least one subspecies and  $\leq 5\%$  in another subspecies were identified using an in-house programme in the environment R. To visualize the peaks, an average spectrum was generated by using the R plotly package (Sievert 2020), which includes aligned masses (tolerance = 150 ppm) and the average intensities for each species accessible online at <https://agodmer.github.io/MABSc/>. The peaks were considered identical with a maximum deviation of 5 Da.

## RESULTS

### MABS isolates identification by the Bruker® MALDI Biotyper® (Table S2)

Ninety-seven per cent (i.e. 1777/1826 spectra) of the spectra identified with the MBT 3.1 and Mycobacteria Library version 5.0, regardless of culture medium, had

a score  $\geq 1.8$ , indicating a high probability of MABS in accordance with the Bruker® manufacturer's recommendations for identification of mycobacteria at the complex level. Regardless of the associated score (minimum: 1.46; maximum: 2.36), all spectra tested resulted in identification as MABS by the MBT, whereas only a few strains were correctly identified at the species level when the top match gave a subspecies identification. For the strains grown on CBA, the software MBT has rendered an identification as a subspecies for 59% (577/977) of the spectra obtained, and the rates of correct identification for *M. bolletii*, *M. massiliense* and *M. abscessus* were 5.9% (7/119), 67.9% (95/140) and 81.1% (258/318), respectively. For strains grown on MGIT®, the software MBT has rendered an identification as a subspecies for 90.6% (769/849) of the spectra, and the rates of correct identifications for *M. bolletii*, *M. massiliense* and *M. abscessus* were 9.2% (10/109), 35.5% (76/214) and 93.9% (419/446), respectively.

### Specific peak analysis for subspecies discrimination of MABS

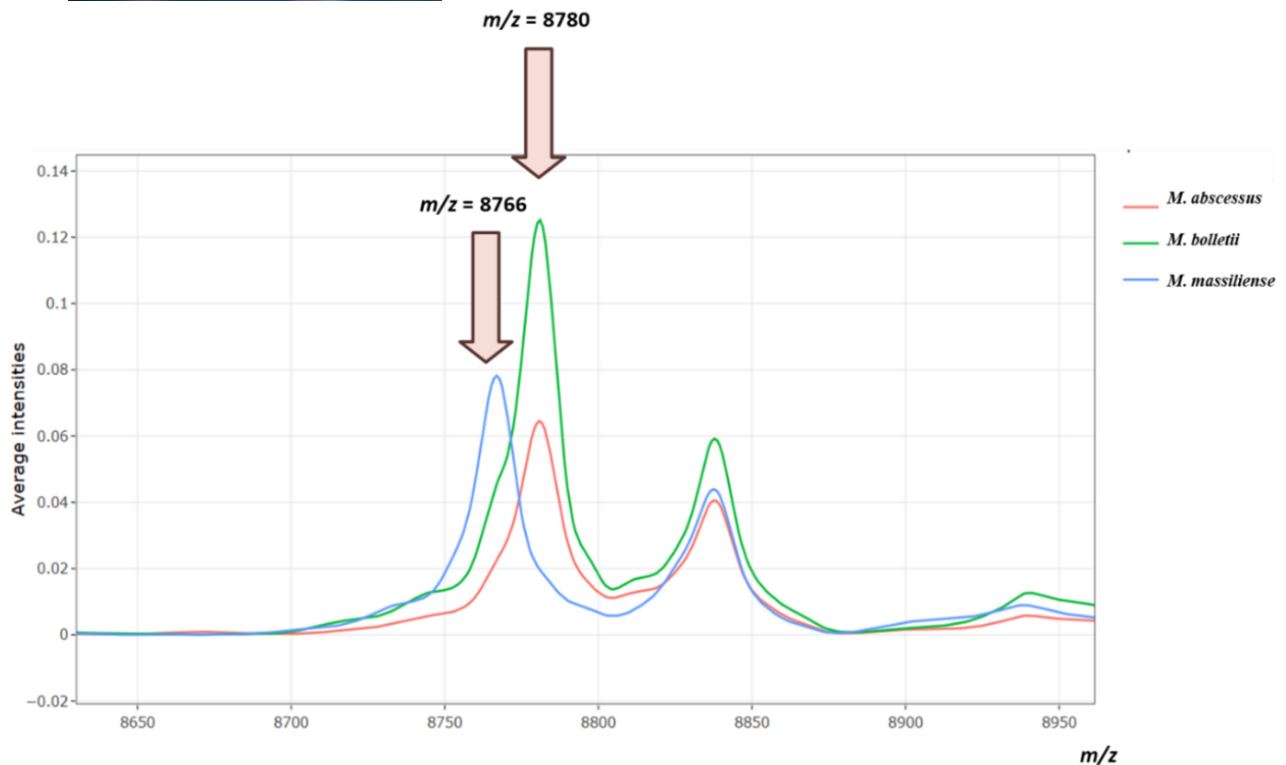
All protein peaks reported in previous studies were searched for among the analysed isolates. No unique subspecies specific peaks were found from spectra obtained from strains grown on MGIT®, whereas two discriminant peaks were found from spectra obtained from strains grown on CBA. The two peaks ( $m/z = 8766$  and  $8780$ ) were useful to discriminate *M. massiliense* from the other subspecies, but no discriminant peak was found to distinguish between *M. bolletii* and *M. abscessus* (Figure 2). In particular, the peaks  $m/z = 2081$ ,  $3378$  and  $7637$  considered to be discriminating in previous studies were found to have comparable frequencies in the three subspecies in our study (Table 1).

### Analysis of isolates by unsupervised algorithms

Through the analysis of all isolates by principal component analysis (PCA), different variables were analysed: the MABS subspecies, the type of culture media. Types of culture media formed clusters (Figure 3A), whereas the strains belonging to different MABS subspecies overlapped in both clusters (Figure 3B).

### Identification of MABS isolates using ML-based pipelines through the “test dataset”

A total of 1826 spectra were generated and analysed from the 40 strains included in the study, representing 977 and 849 spectra from strains grown on CBA and



**FIGURE 2** Visualization of two discriminant peaks able to distinguish *Mycobacterium massiliense* from the two other subspecies from *Mycobacterium abscessus* on solid medium (CBA for Columbia Blood Agar). This visualization tool is available at <https://agodmer.github.io/MABSc/>

MGIT®, respectively. The performances (Kappa coefficient and accuracy) were higher for the ML-based classifiers trained with spectra obtained from strains grown on CBA compared to MGIT® ( $p < 2.10^{-145}$ ). This confirms a significant difference between the spectra depending on the culture medium, as seen previously in the PCA (Figure 3).

For spectra obtained from strains grown on CBA and across all ML-based pipelines tested, an average of 89% of spectra were accurately identified, consistent with the molecular reference method (corresponding to a mean  $\pm$  standard deviation accuracy of  $0.89 \pm 0.01$ ). In addition, the mean  $\pm$  standard deviation Kappa coefficient deviation was at  $0.85 \pm 0.15$ . The RF algorithm obtained significantly better performances than the other algorithms ( $p < 10^{-26}$ ) (Table S3). More specifically, the ML-based pipelines trained with RF algorithms exhibited mean F1-scores and recall (sensitivity) exceeding 0.89 irrespective of the subspecies (Table S4). These performances translated to a mean correct identification rate of at least 98% for *M. abscessus* and *M. massiliense* and between 90 and 95% for *M. bolletii* (Table S4). Given these results, the ML-based classifiers trained with RF were used to develop a metaclassifier.

For spectra obtained from strains grown on MGIT®, ML-based pipelines trained with RF algorithms exhibited lower mean F1-score, sensitivity and specificity

than those obtained for spectra obtained for strains grown on CBA (Table S4). The global performances translated to a correct identification of up to 80%, 67% and 78% for *M. abscessus*, *M. massiliense* and *M. bolletii*, respectively. The differences between ML-based algorithms (excluding method for selecting non-informative peaks) were not significantly different ( $p > 0.2$ ) (Table S3). However, unlike the spectra obtained from CBA medium, whose best-performing ML-based classifier was RF, the one designed with the (sparse Partial Least Squares-linear Support Vector Machine) pipeline obtained the highest performance with Kappa coefficient and accuracy of 0.66 and 0.81, respectively. Given these results, the ML-based classifiers trained sPLSDA-SVM pipeline were used to develop a metaclassifier.

### Identification of MABS isolates using metaclassifiers through the “test dataset”

For the spectra obtained from strains grown on CBA, predictions from the ML-based classifiers performed with the RF algorithm resulted in the Metaclassifier-ML-CBA (Table 2). For the spectra obtained from strains grown on MGIT®, predictions from the ML-based pipeline (sPLSDA-SVM classifiers) were used to develop the Metaclassifier-ML-MGIT (Table 2). Overall,

TABLE 1 Discriminant peaks and their associated frequencies in MABS, in the literature and in our study (in bold).

Media	Peak position (m/z)	Frequency depending of the MABS species (%)			MALDI-TOF system	Geographical area	Reference
		<i>M. abscessus</i>	<i>M. massiliense</i>	<i>M. boletii</i>			
7H11	[2081.00–2084.63]	31.7	7.7	NI	VITEK MS	USA	Panagea et al. (2015)
COS		97.8	0	100	MALDI Biotyper	France	Fangous et al. (2014)
LJ		+ <sup>a</sup>	–	NI	MALDI Biotyper	Taiwan	Tseng et al. (2013)
7H11 and LJ		96	100	100	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
CBA		<b>7.6</b>	<b>4.4</b>	<b>20</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
MGIT		<b>87.8</b>	<b>88.4</b>	<b>82.5</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
7H11 and LJ	[2673–2676.66]	88.9	7.3	17.1	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
CBA		<b>22.9</b>	<b>36.9</b>	<b>16.4</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
MGIT		<b>1.7</b>	<b>0.44</b>	<b>0</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
LJ	[2764.02–2766.57]	–	+ <sup>a</sup>	NI	MALDI Biotyper	Taiwan	Tseng et al. (2013)
CBA		<b>1.1</b>	<b>0.4</b>	<b>0.7</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
MGIT		<b>8.9</b>	<b>9</b>	<b>7</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
LJ	[3107.97–3108.00]	0	96	0	MALDI Biotyper	France	Fangous et al. (2014)
7H11 and LJ		10.3	79.2	17.1	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
CBA		<b>0.82</b>	<b>0.4</b>	<b>0.7</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
MGIT		<b>0.3</b>	<b>0</b>	<b>0</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
COS	3123	100	4	100	MALDI Biotyper	France	Fangous et al. (2014)
7H11 and LJ		84.1	9.4	95.1	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
CBA		NA**	NA**	NA**	MALDI Biotyper	France	<b>Our study</b>
MGIT		NA**	NA**	NA**	MALDI Biotyper	France	<b>Our study</b>
LJ	[3354.40–3355.04]	1.4	100	NI	VITEK MS	China	Luo et al. (2016)
7H11 and LJ		4	3.1	12	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
CBA		<b>8.9</b>	<b>88.1</b>	<b>70</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
MGIT		<b>14</b>	<b>63.2</b>	<b>56.6</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>
LJ	[3374.393378.00]	100	96	4.8	MALDI Biotyper	France	Fangous et al. (2014)
7H11 and LJ		99.2	99.9	36.6	MALDI Biotyper	France	Rodriguez-Temporal et al. (2023)
CBA		<b>42.4</b>	<b>23.8</b>	<b>31.4</b>	<b>MALDI Biotyper</b>	<b>France</b>	<b>Our study</b>

(Continues)



TABLE 1 (Continued)

Media	Peak position (m/z)	Frequency depending of the MABS species (%)			MALDI-TOF system	Geographical area	Reference
		<i>M. abscessus</i>	<i>M. massiliense</i>	<i>M. boletii</i>			
MGIT		29.2	14.3	34.1	MALDI Biotyper	France	Our study
LJ	[3463.00–3463.49]	2.2	0	80.9	MALDI Biotyper	France	Fangous et al. (2014)
7H11 and LJ		23.8	31.2	90.2	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
COS		5.6	8.3	6.4	MALDI Biotyper	France	Our study
MGIT		1.9	37.7	3.9	MALDI Biotyper	France	Our study
7H11	[4383.5–4386.24]	0	100	0	MALDI Biotyper	Japan	Suzuki et al. (2015)
LJ		1.4	97.2	NI	VITEK MS	China	Luo et al. (2016)
7H11		–	<sup>a</sup>	NI	MALDI Biotyper	Taiwan	Teng et al. (2013)
CBA		0.3	25.8	0.7	MALDI Biotyper	France	Our study
7H11 and LJ		31.7	89.6	29.3	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
MGIT		0	89.5	0	VITEK MS	Germany	Kehrmann, Wessel, et al. (2016)
MGIT		NA	NA	NA	MALDI Biotyper	France	Our study
7H11	[4390–4391.24]	100	0	100	MALDI Biotyper	Japan	Suzuki et al. (2015)
LJ		98.6	2.8	NA	VITEK MS	China	Luo et al. (2016)
7H11 and LJ		53.2	9.4	41.5	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
CBA		84.4	0.4	99 0.2	MALDI Biotyper	France	Our study
MGIT		100	5.3	100	VITEK MS	Germany	Kehrmann, Wessel, et al. (2016)
MGIT		99.2	100	100	MALDI Biotyper	France	Our study
LJ		0	100	NI	VITEK MS	China	Luo et al. (2016)
7H11 and LJ	[6711.10–6712.47]	32.5	90.6	73.2	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
CBA		9.5	99.6	84.3	MALDI Biotyper	France	Our study
MGIT		32.7	85.7	86.8	MALDI Biotyper	France	Our study
7H11 and LJ	[6960–6961.7]	90.5	26	9.8	MALDI Biotyper	Europe	Rodriguez-Temporal et al. (2023)
CBA		82.2	23.4	16.4	MALDI Biotyper	France	Our study
MGIT		99.8	56.1	51.9	MALDI Biotyper	France	Our study
7H11	[7636.43–7639.70]	93.2	0	100	MALDI Biotyper	Japan	Suzuki et al. (2015)

TABLE 1 (Continued)

Media	Peak position (m/z)	Frequency depending of the MABS species (%)			MALDI-TOF system	Geographical area	Reference
		<i>M. abscessus</i>	<i>M. massiliense</i>	<i>M. boletii</i>			
7H11		+ <sup>a</sup>	–	NI	Taiwan	Teng et al. (2013)	
7H11 and LJ		92.6	62.5	90.2	Europe	Rodriguez-Temporal et al. (2023)	
CBA		<b>38.9</b>	<b>0.8</b>	<b>42.8</b>	France	<b>Our study</b>	
MGIT		<b>85.3</b>	<b>31.4</b>	<b>72.1</b>	France	<b>Our study</b>	
7H11	[7665.69–7669.20]	3.4	88.1	0	Japan	Suzuki et al. (2015)	
7H11		–	+ <sup>a</sup>	NI	Taiwan	Teng et al. (2013)	
7H11 and LJ		11.9	41.7	2.4	Europe	Rodriguez-Temporal et al. (2023)	
CBA		<b>NA</b>	<b>NA</b>	<b>NA</b>	France	<b>Our study</b>	
MGIT		<b>67.9</b>	<b>73.5</b>	<b>19.4</b>	France	<b>Our study</b>	
7H11	[8508.66–8509.09]	4.9	84.6	NI	USA	Panagea et al. (2015)	
CBA		91.8	97.2	85.7	France	<b>Our study</b>	
7H11 and LJ		7.1	9.4	14.6	Europe	Rodriguez-Temporal et al. (2023)	
MGIT		38.5	75.8	89.9	France	<b>Our study</b>	
7H11	[8766.90–8771.13]	0	38.4	NI	USA	Panagea et al. (2015)	
7H11		0	100	0	Japan	Suzuki et al. (2015)	
7H11		–	+ <sup>a</sup>	NI	Taiwan	Teng et al. (2013)	
LJ		–	+ <sup>a</sup>	NI	Taiwan	Tseng et al. (2013)	
LJ		0.7	97.2	NI	China	Luo et al. (2016)	
7H11 and LJ		2.4	70.8	7.3	Europe	Rodriguez-Temporal et al. (2023)	
CBA		<b>4.1</b>	<b>99</b>	<b>0</b>	France	<b>Our study</b>	
MGIT		0	57.9	0	Germany	Kehrmann, Wessel, et al. (2016)	
MGIT		<b>15.7</b>	<b>72.2</b>	<b>9.3</b>	France	<b>Our study</b>	
7H11	[8780.67–8783.84]	100	0	100	Japan	Suzuki et al. (2015)	
7H11		+ <sup>a</sup>	–	NI	Taiwan	Teng et al. (2013)	
LJ		+ <sup>a</sup>	–	NI	Taiwan	Tseng et al. (2013)	
MGIT		89.3	0	100	Germany	Kehrmann, Wessel, et al. (2016)	

(Continues)

TABLE 1 (Continued)

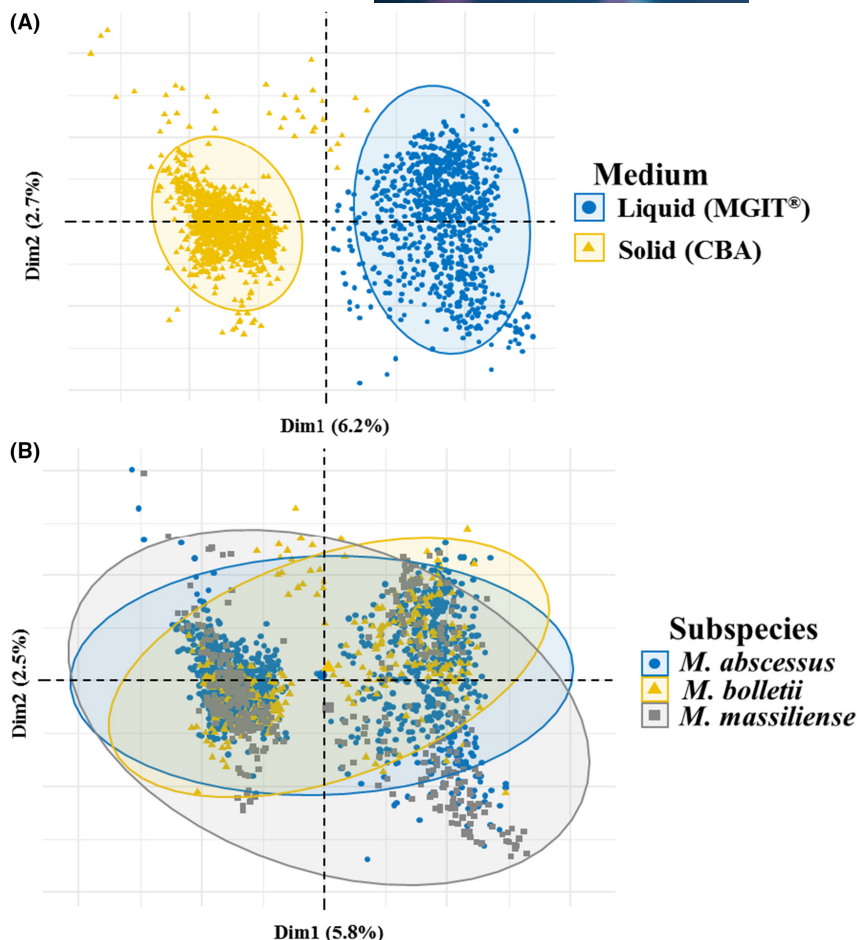
Media	Peak position (m/z)	Frequency depending of the MABS species (%)			MALDI-TOF system	Geographical area	Reference
		<i>M. abscessus</i>	<i>M. massiliense</i>	<i>M. boletii</i>			
7H11 and LJ		72.2	6.2	61	MALDI Biotyper	Europe	Rodríguez-Temporal et al. (2023)
LJ		99.3	2.8	NI	VITEK MS	China	Luo et al. (2016)
CBA		99.1	0	96	MALDI Biotyper	France	Our study
MGIT		87.6	32.7	100	MALDI Biotyper	France	Our study
7H11	[9472.85–9475]	17	0	NI	VITEK MS	USA	Panagea et al. (2015)
7H11		64.4	9.5	100	MALDI Biotyper	Japan	Suzuki et al. (2015)
7H11 and LJ		78.6	16.7	73.2	MALDI Biotyper	Europe	Rodríguez-Temporal et al. (2023)
CBA		96.1	9.5	100	MALDI Biotyper	France	Our study
MGIT		99.8	46.1	100	MALDI Biotyper	France	Our study
7H11	[9477–9477.48]	+ <sup>a</sup>	–	NI	MALDI Biotyper	Taiwan	Teng et al. (2013)
CBA		NA	NA	NA	MALDI Biotyper	France	Our study
MGIT		NA	NA	NA	MALDI Biotyper	France	Our study
LJ	[9500.20–9502.93]	–	+ <sup>a</sup>	NA	MALDI Biotyper	Taiwan	Tseng et al. (2013)
CBA		5.6	82.5	0.7	MALDI Biotyper	France	Our study
MGIT		21.9	69	14	MALDI Biotyper	France	Our study

Note: + Indicates presence and – indicates absence of the peaks.

Abbreviations: CBA, Columbia agar +5% horse blood; COS, Columbia agar with 5% sheep blood; LJ, Löwenstein–Jensen; M7H11, Middlebrook 7H11 agar; NA, Peak was not found in the study; NI, Subspecies was not included in the study.

<sup>a</sup>Frequency was not specified in the study.

**FIGURE 3** Principal component analysis (PCA) of all isolates included in the study coloured according to: (A) the culture media (solid medium, CBA for Columbia Blood Agar in yellow triangles) and liquid medium (MGIT® in blue circles); (B) the subspecies with spectra (*M. abscessus* in blue circles, *M. bolletii* in yellow triangles and *M. massiliense* in grey squares).



**TABLE 2** Performance of identification of generated spectra from MABS isolates using meta-classifiers through the “test dataset” versus the molecular reference method.

Medium	MABS subspecies identification according to the meta-classifiers	MABS subspecies identification according to the reference method			Total	Reliability <sup>b</sup>
		<i>M. abscessus</i>	<i>M. bolletii</i>	<i>M. massiliense</i>		
CBA	<i>M. abscessus</i>	119			119	100% <sup>b</sup>
	<i>M. bolletii</i>		24		24	100% <sup>b</sup>
	<i>M. massiliense</i>			72	72	100% <sup>b</sup>
	Total	119	24	72	215	100% <sup>b</sup>
	<b>Rate of correct identification<sup>a,c</sup></b>	<b>100%<sup>a</sup></b>	<b>100%<sup>a</sup></b>	<b>100%<sup>a</sup></b>	<b>100%<sup>c</sup></b>	
MGIT®	<i>M. abscessus</i>	99	16	24	139	71.2% <sup>b</sup>
	<i>M. bolletii</i>		4		4	100% <sup>b</sup>
	<i>M. massiliense</i>	4		42	46	91.3% <sup>b</sup>
	Total	101	20	66	187	
	<b>Rate of correct identification<sup>a,c</sup></b>	<b>98%<sup>a</sup></b>	<b>20%<sup>a</sup></b>	<b>63.6%<sup>a</sup></b>	<b>77.5%<sup>c</sup></b>	

Abbreviations: CBA, Columbia Blood Agar; MABS, *Mycobacterium abscessus*; MGIT®, Mycobacteria Growth Indicator Tube.

<sup>a</sup>Rate of correct identification corresponds to sensitivity (proportion of correct identification for a subspecies/total of identification for a subspecies).

<sup>b</sup>Reliability corresponds to positive predictive value (proportion of correct identifications among positive predictions for a subspecies).

<sup>c</sup>Global rate of correct identification (number of corrected spectra identified/total spectra).

values in bold: rate of correct identification (by subspecies or overall) and reliability

the rates of correct identification were 100% (215/215) and 77.5% (145/187) using the Metaclassifier-ML-CBA and Metaclassifier-ML-MGIT, respectively. Using the Metaclassifier-ML-CBA, all the strains were correctly identified whatever the subspecies, whereas using the Metaclassifier-ML-MGIT, the correct identification rates were lower: 98% (99/101), 20% (4/20) and 63.6% (42/66) for *M. abscessus*, *M. bolletii* and *M. massiliense*, respectively.

## DISCUSSION

MABS's subspecies identification is crucial for timely medical treatment, since the different subspecies display different susceptibility profiles to macrolides that are the cornerstone of treatment. Since the available tools to identify NTM are mainly based on molecular methods that are cumbersome and require trained staff, a simple and rapid method for the identification of MABS subspecies such as MALDI-TOF MS would be of great medical interest. A previous study had the same objective as ours, that is to develop the discriminating potential of MALDI-TOF combined with ML techniques for MABS subspecies (Rodríguez-Temporal et al., 2023).

As previously reported, the initial observation revealed that the commonly utilized commercial MBT Bruker® system in microbiology laboratories lacks the capability to routinely differentiate the different subspecies of MABS (Brown-Elliott et al., 2019). First, the initial match reported a subspecies name only for three-quarters of cases (73.7%, 1346/1826). Second, when an identification was proposed, it was correct in only two-third of cases (64.3%, 865/1346). Thus, identification was correct in only half of cases (47.4%, 865/1826) as compared with molecular techniques.

The search for discriminating peaks, utilized to enhance MALDI-TOF MS performance in identifying MABS subspecies, was initially chosen for simplicity purpose. However, spectral variations based on the medium and instrument employed may impact the presence or absence of distinctive peaks (Table 1) (Popović et al., 2021). Our assessment evaluated the identification of all three MABS subspecies, unlike some studies that have focused on only two subspecies (Panagea et al., 2015; Teng et al., 2013). We confirmed that the discriminant peaks at  $m/z = 8766$  and  $8780$  enable the differentiation of *M. massiliense* from *M. abscessus* and *M. bolletii* (Table 1). However, these peaks were discriminatory only for strains grown on CBA and did not show the same discriminatory power for strains grown on MGIT®. Previous research from Japan, Taiwan and China has already identified these peaks as discriminatory markers (Suzuki et al., 2015; Teng et al., 2013; Tseng et al., 2013). Also, a study conducted in Germany

showed a significant association between one of these two peaks ( $m/z = 8780$ ) and a frequency exceeding 89% in *M. abscessus* and *M. bolletii*, which was absent in *M. massiliense* strains grown in MGIT® medium (Table 1) (Kehrmann, Schoerding, et al., 2016). However, our findings differ from those of a US study and a multicentre European study, which reported lower frequencies and less discrimination associated with these peaks (Panagea et al., 2015; Rodríguez-Temporal et al., 2023) (Table 1). These differences in discriminating peaks may be explained by the presence of hybrid proteomes due to horizontal gene transfer (Fangous et al., 2014; Jeon et al., 2014; Luo et al., 2016; Sapriel et al., 2016), but also by the different epidemiology of circulating strains in different countries, which probably do not have the same proteome (Ruis et al., 2021).

Another important result of our study is the higher rate of correct identification for strains grown on CBA medium compared to strains grown on MGIT® (Figure 3). This difference has been observed previously for mycobacteria (Balážová et al., 2014; Lotz et al., 2010) and may be due to the fact that liquid medium contains supplements that might alter spectra. For example, the MGIT® media usually contains supplemental components, such as OADC (oleic acid–albumin–dextrose–catalase) that are difficult to rinse out, and can alter the spectra (Anderson et al., 2011; Garrigos et al., 2021; Quinlan et al., 2015). In addition, the rinsing stages during the preparation of the strains for MALDI-TOF MS are responsible for a loss of biomass, which might have an impact in MGIT® as compared to solid media (Anderson et al., 2011; Garrigos et al., 2021; Quinlan et al., 2015). This major limit for mycobacterial diagnosis already observed in old studies is not solved by ML. However, in this study, we evaluated three different ML-based algorithms to eliminate non-informative peaks. This approach enabled us to determine that different ML method should be used depending on the medium used. Indeed, the RFE-RF technique has a higher efficiency than the other methods for analysing spectra derived from solid media ( $p < 10^{-9}$ , Table S3), whereas the “SelectionVarStat” method provided the best performance for analysing spectra from liquid media, ( $p < 10^{-17}$ , Table S3). Interestingly, we did not observe differences in PCA according to morphology of the strains (smooth or rough) contrarily to Rodríguez-Temporal et al. (data not shown), possibly because we add an additional step in the extraction process (i.e. the use of ultrasonic Sonifier), which participated to obtain more homogeneous extractions (Rodríguez-Temporal et al., 2023).

The main goal of our study was to evaluate the benefit of the ML-based approach for MABS subspecies identification. In contrast to the approach based on the search for discriminating peaks, the ML-based approach seems reliable for distinguishing the three

MABS subspecies grown on solid medium. In this regard, the Metaclassifier-ML-CBA gave encouraging results in our study (accuracy at 1 for spectra from the “test dataset”). These results are in line with a recent study that has shown a rate of correct identification as high as 88.6% for MABS subspecies isolated from four European countries (Rodríguez-Temporal et al., 2023). In this European study, using two types of solid media (Löwenstein–Jensen and Middlebrook 7H11 agar), no spectral differences were observed based on these solid media types, and the authors demonstrated a high accuracy rate of 96.5% for 164/170 isolates with identical identification on three spots on strains grown on solid media. The current study confirms the superiority of solid over liquid media, even when using a non-mycobacteria dedicated medium such as CBA, we obtained better performances than with the mycobacteria specific liquid medium MGIT®. These results suggest solid media to be the media of choice for the identification of MABS subspecies by MALDI-TOF MS. A combination of factors may account for these performances, including the presence of the discriminatory peaks and the improved spectral quality of solid media.

Another point of interest is that the highest performance in our study was obtained with the RF algorithm, which is in line with the findings of Rodrigues et al. (Rodríguez-Temporal et al., 2023). Our results confirm that the use of this algorithm is particularly effective for the discrimination of MABS subspecies. The RF algorithm also showed good performance for the discrimination of closely related species from the *Enterobacter cloacae* complex using MALDI-TOF MS (Candela et al., 2023). Given the absence of guidelines for applying ML to MALDI-TOF data analysis in microbiology (Greener et al., 2022), it may be advisable to use this algorithm consistently in such analyses.

Although our study brings new data to the field, there are still important limits to the routine use of MALDI-TOF MS for mycobacteria identification in clinical laboratories. First, despite the ML-based approach, the identification of mycobacteria grown in MGIT® media remains a significant challenge. This is evidenced by the fact that none of the ML-based classifiers trained on MGIT® spectra produced results considered suitable for routine laboratory use. Several approaches that could be considered to improve the performance of the ML-based classifiers, including the use of a higher volume of medium to increase the amount of extracted biomass and other ML methods, which might allow intricate patterns in complex datasets to be found (Candela et al., 2022). Second, horizontal gene transfer influences the evolution of MABS subspecies (Macheras et al., 2014), leading to certain strains acquiring a hybrid proteome with biomarkers from both *M. massiliense* and *M. abscessus*, potentially impacting the accuracy of subspecies differentiation (Fangous et al., 2014; Macheras et al., 2014). Notably, the reference method

in this study is approximately 92% reliable in identifying MABS's subspecies, necessitating confirmation results through multiple targets in molecular techniques (Huh et al., 2019; Kehrmann, Kurt, et al., 2016). Finally, the Metaclassifier-ML-CBA was performed on CBA medium because it is commonly used in bacteriology and recommended by Bruker® for the construction of the spectral database (Egli et al., 2015; Kehrmann, Kurt, et al., 2016; Huh et al., 2019). Although ML-based algorithms show promise for improving MALDI-TOF performance, they are still research tools. Therefore, they need to be evaluated on large and various data sets as well as having user-friendly interface to be integrated in routine use (Abdrabou et al., 2023; Candela et al., 2023; Rodríguez-Temporal et al., 2023).

## AUTHOR CONTRIBUTIONS

**Alexandre Godmer:** Conceptualization; investigation; validation; methodology; software; formal analysis; data curation; writing – review and editing; writing – original draft. **Lise Bigey:** Methodology; writing – review and editing; formal analysis; software; data curation. **Quentin Gaii-Gianetto:** Software; writing – review and editing; visualization. **Gautier Pierrat:** Methodology; visualization; formal analysis. **Noshine Mohammad:** Methodology; visualization. **Faiza Mougari:** Resources; visualization. **Renaud Piarroux:** Methodology; writing – review and editing; visualization; formal analysis. **Nicolas Veziris:** Conceptualization; methodology; resources; formal analysis; writing – review and editing; visualization. **Alexandra Aubry:** Conceptualization; investigation; methodology; validation; formal analysis; supervision; resources; writing – review and editing; visualization.

## ACKNOWLEDGEMENTS

Direct financial support was provided by an annual grant from Santé Publique France. The funder had no role in the study design, data collection and interpretation, or the decision to submit the work for publication. The authors would like to thank Dr Andrew Lane (Lane Medical Writing) for editorial support for English language and grammar.

## CONFLICT OF INTEREST STATEMENT

All authors report no potential conflicts of interest with respect to the research, authorship and publication of this article.

## DATA AVAILABILITY STATEMENT


The data that support the findings of this study are available from the corresponding author upon reasonable request.


## ETHICS STATEMENT

All specimens were processed and anonymized in accordance with ethical and legal standards, and patients

were not physically involved in this study. Informed consent was not needed for this study.

## ORCID

Alexandre Godmer  <https://orcid.org/0000-0002-5211-5796>

Noshine Mohammad  <https://orcid.org/0000-0002-3891-0216>

## REFERENCES

- Abdrabou, A.M.M., Sy, I., Bischoff, M., Arroyo, M.J., Becker, S.L., Mellmann, A. et al. (2023) Discrimination between hypervirulent and non-hypervirulent ribotypes of *Clostridioides difficile* by MALDI-TOF mass spectrometry and machine learning. *European Journal of Clinical Microbiology & Infectious Diseases*, 42, 1373–1381.
- Alcaide, F., Amlerová, J., Bou, G., Ceysens, P.J., Coll, P., Corcoran, D. et al. (2018) How to: identify non-tuberculous mycobacterium species using MALDI-TOF mass spectrometry. *Clinical Microbiology and Infection*, 24, 599–603.
- Anderson, N.W., Buchan, B.W., Riebe, K.M., Parsons, L.N., Gnacinski, S. & Ledebor, N.A. (2011) Effects of solid-medium type on routine identification of bacterial isolates by use of matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Journal of Clinical Microbiology*, 50, 1008–1013.
- Balážová, T., Makovcová, J., Šedo, O., Slaný, M., Faldyna, M. & Zdráhal, Z. (2014) The influence of culture conditions on the identification of mycobacterium species by MALDI-TOF MS profiling. *FEMS Microbiology Letters*, 353, 77–84.
- Bastian, S., Veziris, N., Roux, A.L., Brossier, F., Gaillard, J.L., Jarlier, V. et al. (2011) Assessment of clarithromycin susceptibility in strains belonging to the *Mycobacterium abscessus* group by erm(41) and rrl sequencing. *Antimicrobial Agents and Chemotherapy*, 55, 775–781.
- Brown-Elliott, B.A., Fritsche, T.R., Olson, B.J., Vasireddy, S., Vasireddy, R., Iakhiaeva, E. et al. (2019) Comparison of two commercial matrix-assisted laser desorption/ionization–time of flight mass spectrometry (MALDI-TOF MS) systems for identification of nontuberculous mycobacteria. *American Journal of Clinical Pathology*, 152, 527–536.
- Brown-Elliott, B.A., Vasireddy, S., Vasireddy, R., Iakhiaeva, E., Howard, S.T., Nash, K. et al. (2015) Utility of sequencing the erm(41) gene in isolates of *Mycobacterium abscessus* subsp. *abscessus* with low and intermediate clarithromycin MICs. *Journal of Clinical Microbiology*, 53, 1211–1215.
- Buchan, B.W., Riebe, K.M., Timke, M., Kostrzewa, M. & Ledebor, N.A. (2014) Comparison of MALDI-TOF MS with HPLC and nucleic acid sequencing for the identification of Mycobacterium species in cultures using solid medium and broth. *American Journal of Clinical Pathology*, 141, 25–34.
- Candela, A., Arroyo, M.J., Sánchez-Molleda, Á., Méndez, G., Quiroga, L., Ruiz, A. et al. (2022) Rapid and reproducible MALDI-TOF-based method for the detection of vancomycin-resistant *Enterococcus faecium* using classifying algorithms. *Diagnostics*, 12, 328.
- Candela, A., Guerrero-López, A., Mateos, M., Gómez-Asenjo, A., Arroyo, M.J., Hernandez-García, M. et al. (2023) Automatic discrimination of species within the *Enterobacter cloacae* complex using matrix-assisted laser desorption ionization–time of flight mass spectrometry and supervised algorithms. *Journal of Clinical Microbiology*, 61, e01049-22.
- Der Werf, M.V., Kodmon, C., Katalinic-Jankovic, V., Kummik, T., Soini, H., Richter, E. et al. (2014) Inventory study of non-tuberculous mycobacteria in the European Union. *BMC Infectious Diseases*, 14, 62.
- Egli, A., Tschudin-Sutter, S., Oberle, M., Goldenberger, D., Frei, R. & Widmer, A.F. (2015) Matrix-assisted laser desorption/ionization time of flight mass-spectrometry (MALDI-TOF MS) based typing of extended-spectrum  $\beta$ -lactamase producing *E. coli* – a novel tool for real-time outbreak investigation. *PLoS One*, 10, e0120624.
- Fangous, M.-S., Mougari, F., Gouriou, S., Calvez, E., Raskine, L., Cambau, E. et al. (2014) Classification algorithm for subspecies identification within the *Mycobacterium abscessus* species, based on matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Journal of Clinical Microbiology*, 52, 3362–3369.
- Faron, M.L., Buchan, B.W., Hyke, J., Madisen, N., Lillie, J.L., Granato, P.A. et al. (2015) Multicenter evaluation of the Bruker MALDI Biotyper CA system for the identification of clinical aerobic gram-negative bacterial isolates. *PLoS One*, 10, e0141350.
- Fergusson, C.H., Coloma, J.M.F., Valentine, M.C., Haackl, F.P.J. & Linnington, R.G. (2020) Custom matrix-assisted laser desorption ionization–time of flight mass spectrometric database for identification of environmental isolates of the genus Burkholderia and related genera. *Applied and Environmental Microbiology*, 86, e00354-20.
- Forbes, B.A., Hall, G.S., Miller, M.B., Novak, S.M., Rowlinson, M.C., Salfinger, M. et al. (2018) Practice guidelines for clinical microbiology laboratories: Mycobacteria. *Clinical Microbiology Reviews*, 31, e00038-17.
- Garrigos, T., Neuwirth, C., Chapuis, A., Bador, J., Amoureux, L., Andre, E. et al. (2021) Development of a database for the rapid and accurate routine identification of Achromobacter species by matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS). *Clinical Microbiology and Infection*, 27, 126.e1–126.e5.
- Gibb, S. & Franceschi, P. (2022) MALDIquantForeign: import/export routines for “MALDIquant”.
- Gibb, S. & Strimmer, K. (2012) Maldiquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics*, 28, 2270–2271.
- Godmer, A., Benzerara, Y., Normand, A.C., Veziris, N., Gallah, S., Eckert, C. et al. (2021) Revisiting species identification within the *Enterobacter cloacae* complex by matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Microbiology Spectrum*, 9, e0066121.
- Godmer, A., Benzerara, Y., Veziris, N., Matondo, M., Aubry, A. & Gianetto, Q.G. (2022) MSclassifR: an R package for supervised classification of mass spectra with machine learning methods. *bioRxiv* 2022.03.14.484252.
- Greener, J.G., Kandathil, S.M., Moffat, L. & Jones, D.T. (2022) A guide to machine learning for biologists. *Nature Reviews. Molecular Cell Biology*, 23, 40–55.
- Gupta, R.S., Lo, B. & Son, J. (2018) Phylogenomics and comparative genomic studies robustly support division of the genus mycobacterium into an emended genus mycobacterium and four novel genera. *Frontiers in Microbiology*, 9, 67.
- Harada, T., Akiyama, Y., Kurashima, A., Nagai, H., Tsuyuguchi, K., Fujii, T. et al. (2012) Clinical and microbiological differences between *Mycobacterium abscessus* and *Mycobacterium massiliense* lung diseases. *Journal of Clinical Microbiology*, 50, 3556–3561.
- Henkle, E. & Winthrop, K.L. (2015) Nontuberculous mycobacteria infections in immunosuppressed hosts. *Clinics in Chest Medicine*, 36, 91–99.
- Hou, T., Chuan, C. & Teng, S. (2019) Current status of MALDI-TOF mass spectrometry in clinical microbiology. *Journal of Food and Drug Analysis*, 27, 404–414.
- Huang, Y.-C., Liu, M.-F., Shen, G.-H., Lin, C.-F., Kao, C.-C., Liu, P.-Y. et al. (2010) Clinical outcome of *Mycobacterium abscessus* infection and antimicrobial susceptibility testing. *Journal of Microbiology, Immunology, and Infection*, 43, 401–406.

- Huh, H.J., Kim, S.-Y., Shim, H.J., Kim, D.H., Yoo, I.Y., Kang, O.-K. et al. (2019) GenoType NTM-DR performance evaluation for identification of *Mycobacterium avium* complex and *Mycobacterium abscessus* and determination of clarithromycin and amikacin resistance. *Journal of Clinical Microbiology*, 57, e00516–e00519.
- Jeon, K., Kwon, O.J., Nam, Y.L., Kim, B.J., Kook, Y.H., Lee, S.H. et al. (2009) Antibiotic treatment of *Mycobacterium abscessus* lung disease: a retrospective analysis of 65 patients. *American Journal of Respiratory and Critical Care Medicine*, 180, 896–902.
- Jeon, S.M., Lim, N.R., Kwon, S.J., Shim, T.S., Park, M.S., Kim, B.J. et al. (2014) Analysis of species and intra-species associations between the *Mycobacterium abscessus* complex strains using pulsed-field gel electrophoresis (PFGE) and multi-locus sequence typing (MLST). *Journal of Microbiological Methods*, 104, 19–25.
- Kehrmann, J., Kurt, N., Rueger, K., Bange, F.C. & Buer, J. (2016) GenoType NTM-DR for identifying *Mycobacterium abscessus* subspecies and determining molecular resistance. *Journal of Clinical Microbiology*, 54, 1653–1655.
- Kehrmann, J., Schoerding, A.-K., Murali, R., Wessel, S., Koehling, H.L., Mosel, F. et al. (2016) Performance of Vitek MS in identifying nontuberculous mycobacteria from MGIT liquid medium and Lowenstein-Jensen solid medium. *Diagnostic Microbiology and Infectious Disease*, 84, 43–47.
- Kehrmann, J., Wessel, S., Murali, R., Hampel, A., Bange, F.-C., Buer, J. et al. (2016) Principal component analysis of MALDI TOF MS mass spectra separates *M. abscessus* (sensu stricto) from *M. massiliense* isolates. *BMC Microbiology*, 16, 24.
- Koh, W.J., Jeon, K., Lee, N.Y., Kim, B.J., Kook, Y.H., Lee, S.H. et al. (2011) Clinical significance of differentiation of *Mycobacterium massiliense* from *Mycobacterium abscessus*. *American Journal of Respiratory and Critical Care Medicine*, 183, 405–410.
- Koh, W.-J., Jeong, B.-H., Kim, S.-Y., Jeon, K., Park, K.U., Jhun, B.W. et al. (2017) Mycobacterial characteristics and treatment outcomes in *Mycobacterium abscessus* lung disease. *Clinical Infectious Diseases*, 64, 309–316.
- Kostrzewa, M. & Maier, T. (2017) Criteria for development of MALDI-TOF mass spectral database. In: H. N. Shah & S. E. Gharbia (Eds.), *MALDI-TOF and tandem MS for clinical microbiology*. Chichester: John Wiley & Sons, Ltd, pp. 39–54.
- Lotz, A., Ferroni, A., Beretti, J.-L., Dauphin, B., Carbonnelle, E., Guet-Revillet, H. et al. (2010) Rapid identification of mycobacterial whole cells in solid and liquid culture media by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology*, 48, 4481–4486.
- Luo, L., Liu, W., Li, B., Li, M., Huang, D., Jing, L. et al. (2016) Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of *Mycobacterium abscessus* subspecies according to whole-genome sequencing. *Journal of Clinical Microbiology*, 54, 2982–2989.
- Macheras, E., Konjek, J., Roux, A.-L., Thiberge, J.-M., Bastian, S., Leão, S.C. et al. (2014) Multilocus sequence typing scheme for the *Mycobacterium abscessus* complex. *Research in Microbiology*, 165, 82–90.
- Marrakchi, H., Lanéelle, M.-A. & Daffé, M. (2014) Mycolic acids: structures, biosynthesis, and beyond. *Chemistry & Biology*, 21, 67–85.
- McGrath, E.E., Blades, Z., McCabe, J., Jarry, H. & Anderson, P.B. (2010) Nontuberculous mycobacteria and the lung: from suspicion to treatment. *Lung*, 188, 269–282.
- Meehan, C.J., Barco, R.A., Loh, Y.-H.E., Cogneau, S. & Rigouts, L. (2021) Reconstituting the genus mycobacterium. *International Journal of Systematic and Evolutionary Microbiology*, 71, 004922.
- Mougari, F., Guglielmetti, L., Raskine, L., Sermet-Gaudelus, I., Veziris, N. & Cambau, E. (2016) Infections caused by *Mycobacterium abscessus*: epidemiology, diagnostic tools and treatment. *Expert Review of Anti-Infective Therapy*, 14, 1139–1154.
- Nash, K.A., Brown-Elliott, A.B. & Wallace, R.J. (2009) A novel gene, erm(41), confers inducible macrolide resistance to clinical isolates of *Mycobacterium abscessus* but is absent from *Mycobacterium chelonae*. *Antimicrobial Agents and Chemotherapy*, 53, 1367–1376.
- Nessar, R., Cambau, E., Reyrat, J.M., Murray, A. & Gicquel, B. (2012) *Mycobacterium abscessus*: a new antibiotic nightmare. *Journal of Antimicrobial Chemotherapy*, 67, 810–818.
- Oren, A. & Garrity, G. (2020) List of new names and new combinations previously effectively, but not validly, published. *International Journal of Systematic and Evolutionary Microbiology*, 70, 4043–4049.
- Panagea, T., Pincus, D.H., Grogono, D., Jones, M., Bryant, J., Parkhill, J. et al. (2015) *Mycobacterium abscessus* complex identification with matrix-assisted laser desorption ionization – time of flight mass spectrometry. *Journal of Clinical Microbiology*, 53, 2355–2358.
- Pastrone, L., Curtoni, A., Criscione, G., Scaiola, F., Bottino, P., Guarrasi, L. et al. (2023) Evaluation of two different preparation protocols for MALDI-TOF MS nontuberculous mycobacteria identification from liquid and solid media. *Microorganisms*, 11, 120.
- Pfyffer, G.E., Welscher, H.M., Kissling, P., Cieslak, C., Casal, M.J., Gutierrez, J. et al. (1997) Comparison of the mycobacteria growth indicator tube (MGIT) with radiometric and solid culture for recovery of acid-fast bacilli. *Journal of Clinical Microbiology*, 35, 364–368.
- Popović, N.T., Kazazić, S.P., Bojanić, K., Strunjak-Perović, I. & Čož-Rakovac, R. (2021) Sample preparation and culture condition effects on MALDI-TOF MS identification of bacteria: a review. *Mass Spectrometry Reviews*, 42, 117–134.
- Quinlan, P., Phelan, E. & Doyle, M. (2015) Matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry (MS) for the identification of mycobacteria from MBBacT ALERT 3D liquid cultures and Lowenstein-Jensen (LJ) solid cultures. *Journal of Clinical Pathology*, 68, 229–235.
- Richard, M., Gutiérrez, A.V. & Kremer, L. (2020) Dissecting erm(41)-mediated macrolide-inducible resistance in *Mycobacterium abscessus*. *Antimicrobial Agents and Chemotherapy*, 64(3), e01930-19.
- Rodríguez-Temporal, D., Herrera, L., Alcaide, F., Domingo, D., Héry-Arnaud, G., van Ingen, J. et al. (2023) Identification of *Mycobacterium abscessus* subspecies by MALDI-TOF mass spectrometry and machine learning. *Journal of Clinical Microbiology*, 61, e0111022.
- Ruis, C., Bryant, J.M., Bell, S.C., Thomson, R., Davidson, R.M., Hasan, N.A. et al. (2021) Dissemination of *Mycobacterium abscessus* via global transmission networks. *Nature Microbiology*, 6, 1279–1288.
- Sapriel, G., Konjek, J., Orgeur, M., Bouri, L., Frézal, L., Roux, A.-L. et al. (2016) Genome-wide mosaicism within *Mycobacterium abscessus*: evolutionary and epidemiological implications. *BMC Genomics*, 17, 118.
- Suzuki, H., Yoshida, S., Yoshida, A., Okuzumi, K., Fukusima, A. & Hishinuma, A. (2015) A novel cluster of *Mycobacterium abscessus* complex revealed by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF MS). *Diagnostic Microbiology and Infectious Disease*, 83, 365–370.
- Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC. <https://plotly-r.com>
- Teng, S.H., Chen, C.M., Lee, M.R., Lee, T.F., Chien, K.Y., Teng, L.J. et al. (2013) Matrix-assisted laser desorption ionization-time of flight mass spectrometry can accurately differentiate between *Mycobacterium massiliense* (*M. abscessus* subspecies



- bolletti) and *M. abscessus* (Sensu Stricto). *Journal of Clinical Microbiology*, 51, 3113–3116.
- Tortoli, E., Brown-Elliott, B.A., Chalmers, J.D., Cirillo, D.M., Daley, C.L., Emler, S. et al. (2019) Same meat, different gravy: ignore the new names of mycobacteria. *The European Respiratory Journal*, 54, 1900795.
- Tseng, S.-P., Teng, S.-H., Lee, P.-S., Wang, C.-F., Yu, J.-S. & Lu, P.-L. (2013) Rapid identification of *M. abscessus* and *M. massiliense* by MALDI-TOF mass spectrometry with a comparison to sequencing methods and antimicrobial susceptibility patterns. *Future Microbiology*, 8, 1381–1389.
- Wassilew, N., Hillemann, D., Maurer, F.P., Kohl, T.A., Merker, M., Brinkmann, F. et al. (2017) Evaluation of the GenoType® NTM DR for subspecies identification and determination of drug resistance in clinical *M. abscessus* isolates. *Clinical Microbiology: Open Access*, 6, Article 1000286.
- Yoshida, S., Arikawa, K., Tsuyuguchi, K., Kurashima, A., Harada, T., Nagai, H. et al. (2015) Investigation of the population structure of *Mycobacterium abscessus* complex strains using 17-locus variable number tandem repeat typing and the further distinction of *Mycobacterium massiliense* hsp65 genotypes. *Journal of Medical Microbiology*, 64, 254–261.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Godmer, A., Bigey, L., Giai-Gianetto, Q., Pierrat, G., Mohammad, N., Mougari, F. et al. (2024) Contribution of machine learning for subspecies identification from *Mycobacterium abscessus* with MALDI-TOF MS in solid and liquid media. *Microbial Biotechnology*, 17, e14545. Available from: <https://doi.org/10.1111/1751-7915.14545>