AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

MYCOBACTERIOLOGY AND AEROBIC ACTINOMYCETES

Identification of *Mycobacterium abscessus* Subspecies by MALDI-TOF Mass Spectrometry and Machine Learning

David Rodríguez-Temporal,[a,b] Laura Herrera,[c] Fernando Alcaide,[d,e] Diego Domingo,[f] Genevieve Héry-Arnaud,[g,h] Jakko van Ingen,[i] An Van den Bossche,[j] André Ingebretsen,[k] Clémence Beauruelle,[g,h] Eva Terschlüsen,[i] Samira Boarbi,[j] Neus Vila,[d] Manuel J. Arroyo,[l] Gema Méndez,[l] Patricia Muñoz,[a,b,m,n] Luis Mancera,[l] María Jesús Ruiz-Serrano,[a] Belén Rodríguez-Sánchez[a,b]

[a]Clinical Microbiology and Infectious Diseases Department, Hospital General Universitario Gregorio Marañón, Madrid, Spain
[b]Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[c]Servicio de Bacteriología, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Spain
[d]Servei de Microbiologia, Hospital Universitari de Bellvitge-IDIBELL, Hospitalet de Llobregat, Spain
[e]Departament de Patologia i Terapèutica Experimental, Universitat de Barcelona, Hospitalet de Llobregat, Spain
[f]Servicio de Microbiología, Hospital Universitario La Princesa, Madrid, Spain
[g]Unit of Bacteriology, Brest University Hospital, Brest, France
[h]Brest University, INSERM, UMR 1078, GGB, Microbiota Axis, Brest, France
[i]Department of Medical Microbiology, Radboud University Medical Centre, Nijmegen, The Netherlands
[j]Division of Human Bacterial Diseases, Sciensano, Brussels, Belgium
[k]Oslo University Hospital, Oslo, Norway
[l]Clover Bioanalytical Software, Granada, Spain
[m]CIBER de Enfermedades Respiratorias (CIBERES CB06/06/0058), Madrid, Spain
[n]Medicine Department, Faculty of Medicine, Universidad Complutense de Madrid, Madrid, Spain

ABSTRACT *Mycobacterium abscessus* is one of the most common and pathogenic nontuberculous mycobacteria (NTM) isolated in clinical laboratories. It consists of three subspecies: *M. abscessus* subsp. *abscessus*, *M. abscessus* subsp. *bolletii*, and *M. abscessus* subsp. *massiliense*. Due to their different antibiotic susceptibility pattern, a rapid and accurate identification method is necessary for their differentiation. Although matrix-assisted laser desorption/ionization–time of flight mass spectrometry (MALDI-TOF MS) has proven useful for NTM identification, the differentiation of *M. abscessus* subspecies is challenging. In this study, a collection of 325 clinical isolates of *M. abscessus* was used for MALDI-TOF MS analysis and for the development of machine learning predictive models based on MALDI-TOF MS protein spectra. Overall, using a random forest model with several confidence criteria (samples by triplicate and similarity values >60%), a total of 96.5% of isolates were correctly identified at the subspecies level. Moreover, an improved model with Spanish isolates was able to identify 88.9% of strains collected in other countries. In addition, differences in culture media, colony morphology, and geographic origin of the strains were evaluated, showing that the latter had an impact on the protein spectra. Finally, after studying all protein peaks previously reported for this species, two novel peaks with potential for subspecies differentiation were found. Therefore, machine learning methodology has proven to be a promising approach for rapid and accurate identification of subspecies of *M. abscessus* using MALDI-TOF MS.

KEYWORDS *Mycobacterium abscessus*, subspecies differentiation, MALDI-TOF, mass spectrometry, machine learning, *Mycobacterium bolletii*, *Mycobacterium massiliense*

Nontuberculous mycobacteria (NTM) are a group of mycobacteria present in the environment that, in some cases, can cause different types of infections in humans, such as pulmonary infections, skin and soft tissue infections, and disseminated infections (1). *Mycobacterium abscessus* is one of the most common and

Editor Melissa B. Miller, The University of North Carolina at Chapel Hill School of Medicine

Address correspondence to David Rodríguez-Temporal, david.rodriguez@iisgm.com.

The authors declare no conflict of interest.

Received 28 July 2022
Returned for modification 6 September 2022
Accepted 14 December 2022

pathogenic NTM species isolated in clinical laboratories, causing respiratory infections, particularly in patients with cystic fibrosis (2). *M. abscessus* contains three subspecies: *M. abscessus* subsp. *abscessus*, *M. abscessus* subsp. *massiliense*, and *M. abscessus* subsp. *bolletii* (3). Here, they will be referred to as *M. abscessus*, *M. massiliense*, and *M. bolletii*, respectively.

The three subspecies show different susceptibility to clarithromycin, a decisive antibiotic for the treatment of these infections. Thus, *M. bolletii* and most strains of *M. abscessus* show resistance to clarithromycin, whereas *M. massiliense* is susceptible (4), although some exceptions can occur on rare occasions (5). The different antibiotic susceptibility pattern, in addition to the recommendation of the American Thoracic Society and Infectious Diseases Society of America (ATS/IDSA) to identify NTMs at the species level, makes it necessary to implement novel approaches for rapid and accurate discrimination of these three subspecies (6).

Currently, *M. abscessus* subspecies can only be identified by molecular methods, such as a commercial kit based on PCR-reverse hybridization (7) or by multiple gene sequencing (*hsp65*, *rpoB*, *erm*(*41*), etc.) (8, 9). On the other hand, the use of matrix-assisted laser desorption/ionization–time of flight mass spectrometry (MALDI-TOF MS) allows the reliable identification of most NTMs and has become the main identification method in several clinical laboratories (10, 11). However, differentiation of closely related species (like *M. abscessus* subspecies) remains a challenge. Although some studies have attempted subspecies identification by protein peak analysis (12–17), there is no consensus on the best strategy to follow. In last years, new approaches for data analysis from MALDI-TOF mass spectra have been applied, such as machine learning methods, which have the potential to get additional information than simple species identification (18).

The aim of this study was to evaluate MALDI-TOF MS and machine learning algorithms for the differentiation of *M. abscessus* subspecies. This study represents the first proof of concept for the identification of these species by applying MALDI-TOF MS and machine learning.

MATERIALS AND METHODS

Mycobacterial isolates. A total of 325 clinical isolates of *M. abscessus* obtained from 232 different patients were included in this study. They encompassed 157 *M. abscessus*, 116 *M. massiliense*, and 52 *M. bolletii* isolates. The isolates were obtained from Hospital General Universitario Gregorio Marañón (HGM; Madrid, Spain), Hospital Universitario La Princesa (HLP; Madrid, Spain), Instituto de Salud Carlos III-Centro Nacional de Microbiología (ISCIII; Majadahonda, Spain), Hospital Universitari de Bellvitge (HUB; Hospitalet de Llobregat, Spain), Oslo University Hospital (Oslo, Norway), Brest University Hospital (Brest, France), Radboud University Medical Centre (Nijmegen, The Netherlands), and Sciensano (Brussels, Belgium). All isolates are described in Table S1 in the supplemental material.

Bacterial cultures and protein extraction procedure. All isolates were previously identified by PCR-reverse hybridization (GenoType NTM-DR; Hain Lifescience, Nehren, Germany) and/or whole-genome sequencing by Kmer finder (19). All HGM, HLP, and ISCIII isolates were cultured from frozen stocks on 7H11 agar plates until growth was observed. Among HUB isolates, 38 were cultured on 7H11 agar plates and 48 on Löwenstein-Jensen (bioMérieux; Marcy l'Etoile, France) media. In all cases, the isolates were incubated at 37°C until growth was observed (4 to 7 days). The protein extraction procedure for MALDI-TOF MS analysis was performed as previously described (11). First, a 1 μL loopful of biomass was suspended in 300 μL of high-performance liquid chromatography quality water and then heat inactivated in a dry bath at 95°C for 30 min. After this, 900 μL of ethanol was added, the tubes were centrifuged at 13,000 rpm for 2 min, and the supernatant was discarded. After the supernatant was centrifuged and discarded again, the pellet was dried at room temperature. Then, 0.5-mm silica/zirconia beads were added together with 10 μL of acetonitrile. The tubes were vortexed briefly and sonicated for 15 min. After sonication, 10 μL of formic acid was added, and the tubes were vortexed for 10 s and centrifuged at 13,000 rpm for 2 min. One microliter of the supernatant was deposited onto the MALDI target plate (Bruker Daltonics; Bremen, Germany) in triplicates, allowed to dry, and covered with 1 μL of α-cyano-4-hydroxycinnamic acid.

Spectra acquisition by MALDI-TOF MS and data processing. Acquisition of protein spectra was performed using the MBT Smart MALDI Biotyper (Bruker Daltonics) in the range of 2,000 to 20,000 Da. All spots were read three times, resulting in nine protein spectra per isolate. The spectra were exported and processed with Clover MS Data Analysis software (Clover Biosoft, Granada, Spain). The processing pipeline consisted of (i) smoothing by Savitzky-Golay filter (window length = 11; polynomial order = 3); (ii) baseline subtraction by Top-Hat filter (factor = 0.02); (iii) alignment of spectra with 2 Da of constant tolerance and 300 ppm of linear mass tolerance; and (iv) normalization by total ion current.

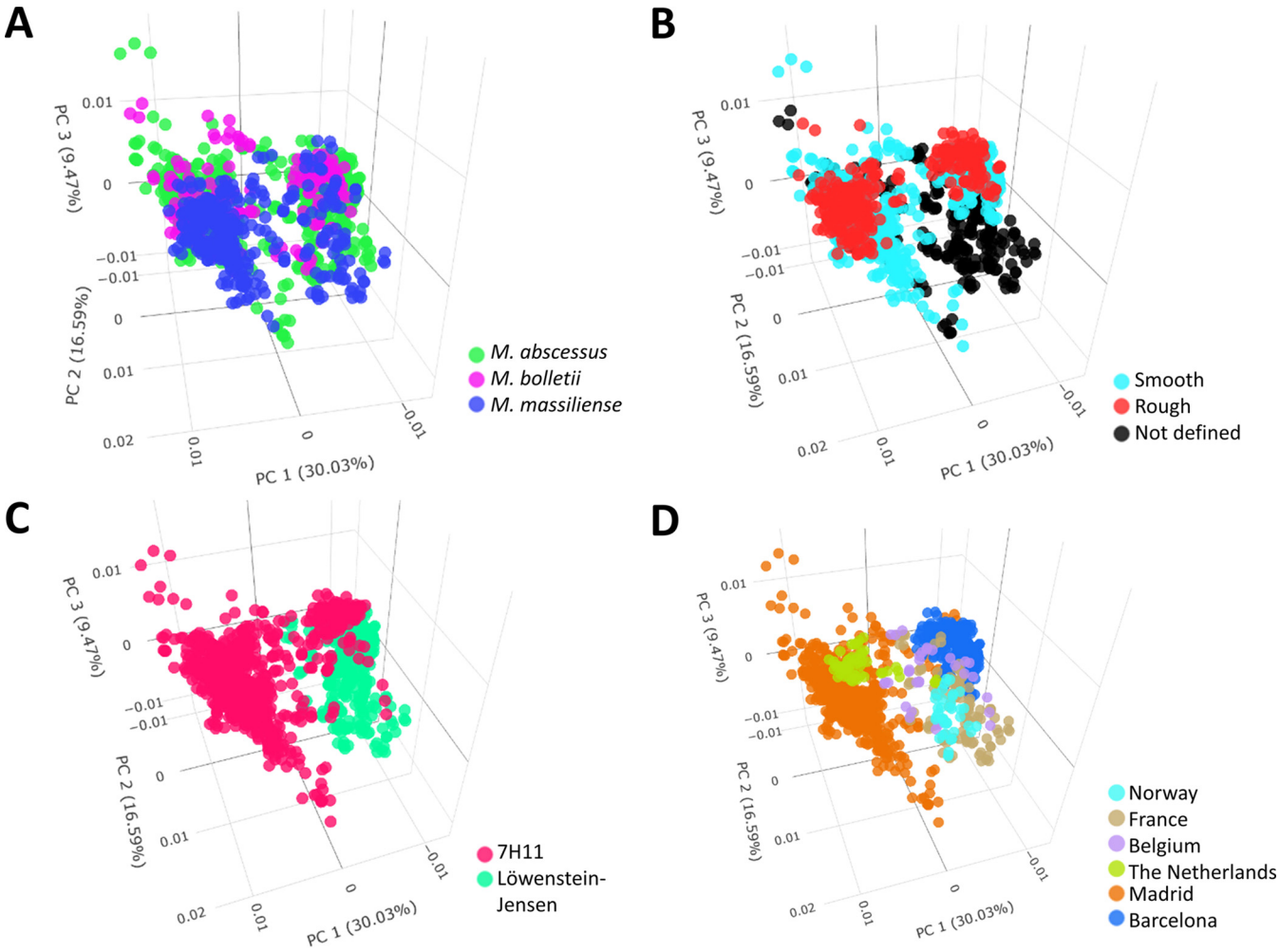

FIG 1 Principal component (PC) analysis of all isolates included in the study, colored according to different characteristics. (A) Comparison of *M. abscessus* subspecies. (B) comparison of colony morphology. (C) Comparison of culture media. (D) Comparison of geographical zone of origin.

Predictive models and external validation. Once the spectra were processed, unsupervised principal component analysis (PCA) and hierarchical cluster analysis and supervised partial least squares discriminant analysis (PLS-DA), support vector machine (SVM), random forest (RF), and K-nearest neighbors (KNN) algorithms were applied for the creation of predictive models using Clover MS Data Analysis software. A total of 43 isolates (20 *M. abscessus*, 15 *M. massiliense*, and 8 *M. bolletii*) collected in HGM, ISCIII, and HUB were included in the test set for the creation of the predictive models; they represented a total of 539 mass spectra. These isolates were randomly selected to represent all the variability observed previously (subspecies, morphology, culture media, and geographical origin). *M. massiliense* and *M. bolletii* spectra were balanced by oversampling to obtain the same number of spectra for each category. Internal validation was performed by 10-fold cross-validation. For external validation, 282 isolates collected in all centers were used (137 *M. abscessus*, 101 *M. massiliense*, and 44 *M. bolletii*), and the identification obtained in each of the three spots used was considered.

Ethics statement. The Ethics Committee of the Gregorio Marañón Hospital (CEIm) evaluated this project and considered that all the conditions for waiving informed consent were met, since the study was conducted with microbiological samples and not with human products.

RESULTS

Analysis of isolates by unsupervised algorithms. Through the analysis of all isolates by PCA (Fig. 1), different variables included in the study were examined: the *M. abscessus* subspecies, the morphology of the colonies, the type of culture media, and the geographical origin of the isolates. As can be observed in Fig. 1, the geographical origin of the isolates formed clusters, especially from strains collected in Madrid hospitals (HGM, HLP, and ISCII), Barcelona (HUB), Norway, and The Netherlands. On the other hand, isolates from different subspecies and different morphology overlapped in both

TABLE 1 Accuracy results for internal 10-fold cross validation and external validation using Spanish isolates

Algorithm	*M. abscessus* (%)	*M. bolletii* (%)	*M. massiliense* (%)	Total (%)
10-Fold cross validation				
PLS-DA	99.4	100	100	99.8
SVM	99.4	100	100	99.8
RF	100	100	99.4	99.8
KNN	79.9	97.2	92.2	89.8
External validation				
PLS-DA	58.0	84.4	73.9	67.7
SVM	89.0	81.3	90.8	88.1
RF	93.0	78.1	91.8	90.1
KNN	44.0	25.0	61.8	47.8

clusters, as well as isolates from HUB, which were analyzed in two different culture media.

Analysis of isolates by supervised algorithms. (i) Internal validation of predictive models. The results for each algorithm after applying a 10-fold cross-validation are shown in Table 1. The algorithms PLS-DA (Fig. S1), SVM, and RF showed the same accuracy (99.8%), with only 1 spectrum of the 539 misclassified as other category (Table S2), while KNN was the algorithm with lower accuracy (Fig. S1).

(ii) External validation of predictive models using Spanish isolates. Blind analysis of the 201 Spanish isolates used for external validation showed that PLS-DA and KNN produced low identification rates, while the RF algorithm yielded 90.1% correct classification (Table 1). With this algorithm, *M. bolletii* obtained the lower identification rate, with 18 spectra misclassified (Table 2). Since RF yielded the highest identification rate (Fig. 2A), results obtained with this algorithm were further analyzed. Among the three identifications obtained in each spot for each isolate, the subspecies obtained in at least two spots were considered the final identification. A total of 184 (91.5%) of the isolates obtained uniform identification results (Table 3), and only 1 isolate obtained different subspecies identification for each spot. Among the isolates with identical identification, the accuracy rate of

TABLE 2 Accuracy of external validation for all algorithms tested over 201 isolates, representing 603 mass spectra

Actual/predicted	*M. abscessus*	*M. bolletii*	*M. massiliense*	Percent correct
PLS-DA				
M. abscessus	174	122	4	58.0%
M. bolletii	3	78	15	81.2%
M. massiliense	0	54	153	73.9%
Total PLS-DA				67.2%
SVM				
M. abscessus	264	7	29	88.0%
M. bolletii	15	77	4	80.2%
M. massiliense	13	6	188	90.8%
Total SVM				87.7%
RF				
M. abscessus	277	5	18	92.3%
M. bolletii	15	78	3	81.2%
M. massiliense	9	8	190	91.8%
Total RF				90.4%
KNN				
M. abscessus	134	58	108	44.7%
M. bolletii	29	23	44	24.0%
M. massiliense	47	32	128	61.8%
Total KNN				47.3%

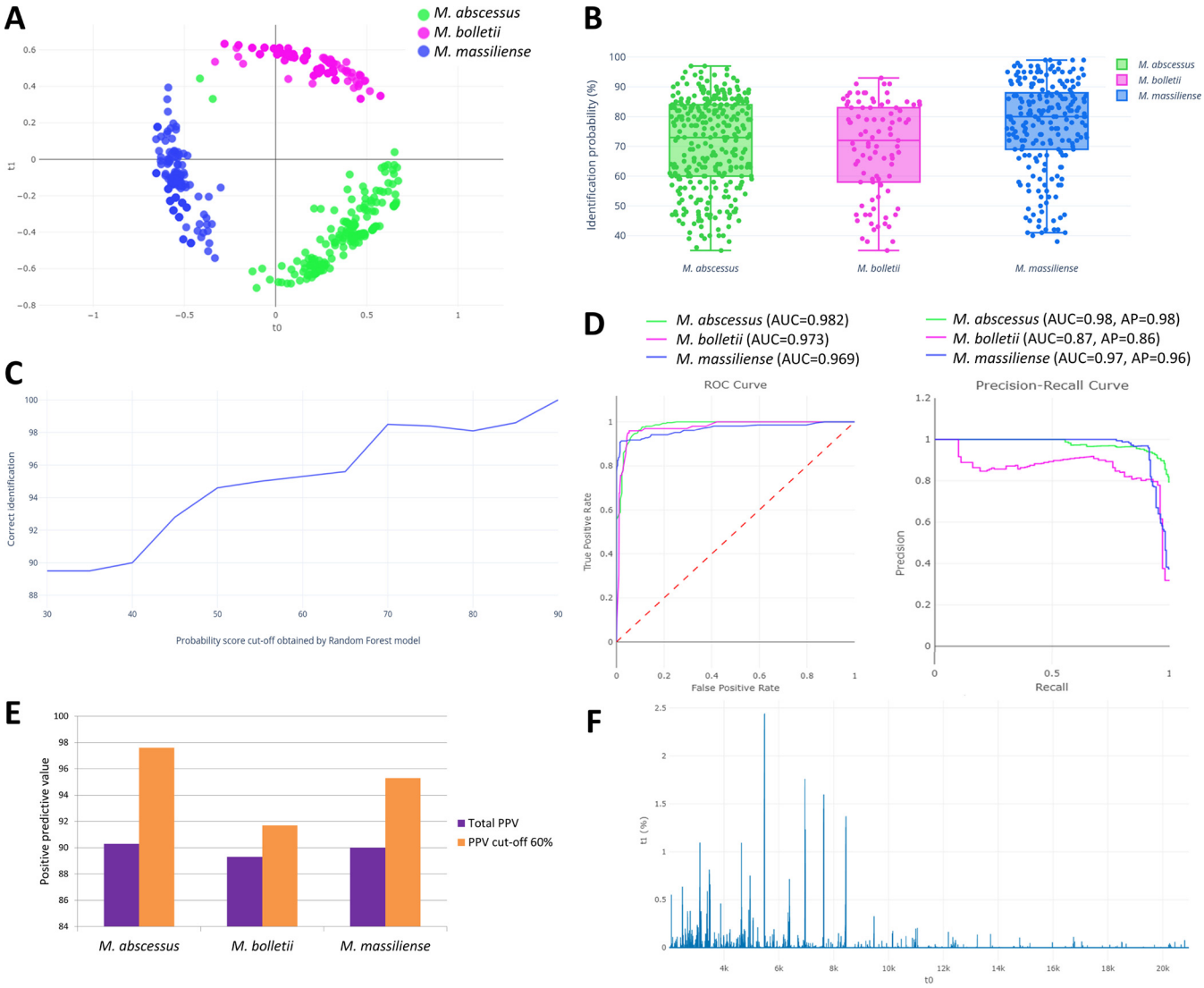

FIG 2 Analysis of mass spectra by random forest (RF) algorithm. (A) RF plot of the model. (B) Percentages of identification probably obtained by RF on validation isolates. (C) Number of correctly identified isolates according to probability cutoff obtained by RF. (D) Receiver operating characteristics (ROC) and precision recall curves for validation isolates by RF. AP, average precision. Precision-recall curve shows the trade-off between the precision (low false-positive rate) and recall (low false-negative rate). (E) Total positive predictive value (PPV) for RF results and PPV using a 60% probability cutoff. (F) Feature importances of mass peaks for RF model.

subspecies-level identification was higher than in those with only two matching identifications. Moreover, the probability of correct identification provided by RF was evaluated to establish a confidence cutoff. For all subspecies, 172 (85.6%) isolates obtained a probability higher than 60% (Fig. 2B), so this cutoff was proposed for a confident result. Considering the categorical result, 89.6% of isolates were correctly identified at the subspecies level, while establishing the confidence cutoff at 60% of probability, the accuracy rate increased to 95.9% (Fig. 2C). Moreover, when both parameters were considered (same identification in 3 spots and confidence higher than 60%), out of the 170 isolates that met these criteria, 164 (96.5%) were correctly identified (Table 3). Then, a total of 6 isolates were misidentified: 2 *M. abscessus* identified as *M. massiliense*, 2 *M. massiliense* as *M. abscessus*, and 2 *M. massiliense* identified as *M. bolletii*. The three subspecies performed similarly, with a similar area under the curve (AUC) between them (Fig. 2D). Finally, positive predictive values (PPV) were evaluated for each subspecies. Considering all identification results, the PPV obtained was 90.3% for *M. abscessus*, 89.3% for *M. bolletii*, and 90.0% for *M. massiliense*. When we considered only those isolates with a probability result higher than 60%, the PPV increased to 97.6%, 91.7%, and 95.3% for each subspecies, respectively (Fig. 2E).

TABLE 3 Random forest accuracy according to the identification obtained in each spot[a]

RF identification	No. isolates (%)	No. correct (%)
Same ID in 3 spots	184 (91.5)	174 (94.6)
Same ID in 2 spots	16 (8.0)	7 (43.8)
Three different ID	1 (0.5)	0 (0)
Isolates with confidence >60%	172 (85.6)	165 (95.9)
Same ID (3 spots) and confidence >60%	170 (84.6)	164 (96.5)

[a]RF, random forest; ID, indentification.

(iii) External validation with isolates from other countries. To evaluate the usefulness of the model for its application in other laboratories, a larger classification model was developed using the 120 isolates (40 of each subspecies) sourced from Spanish hospitals and it was applied to the 81 strains obtained from 4 European countries. This new RF model increased the correct subspecies-level identification rate to 88.9%. Using the RF algorithm, the previous model obtained 69.1% of correct classification at the subspecies level (Table S3).

Specific peak analysis for subspecies discrimination. All protein peaks reported in previous studies were searched among the analyzed isolates (Table 4). No unique subspecies specific peaks were found, although some of them were present in most strains of certain subspecies. Thus, almost all *M. abscessus* isolates showed peaks at 2,081, 3,378, and 7,637 *m/z*; *M. bolletii* showed at 2,081, 3,123, 3,463, and 7,637 *m/z*; and *M. massiliense* showed at 3,378, 4,385, and 6,711 *m/z*. In addition, two novel potential peaks were found in this study: 2,673 *m/z* (Fig. S3), which was present in 88.9% of *M. abscessus* isolates, 17.1% of *M. bolletii*, and 7.3% of *M. massiliense*; and 6,960 *m/z* (Fig. S3), which was present in 90.5% of *M. abscessus* isolates, 9.8% of *M. bolletii*, and 26.0% of *M. massiliense* (Table 4).

DISCUSSION

Differentiation of *M. abscessus* subspecies by MALDI-TOF MS has been attempted in previous studies using conventional peak analysis (12–17). However, many variables can

TABLE 4 Presence of all protein peaks reported in previous studies and in the present one among isolates of each subspecies[a]

	M. abscessus		*M. bolletii*		*M. massiliense*	
Peaks	Present study (%)	Previous studies (%)	Present study (%)	Previous studies (%)	Present study (%)	Previous studies (%)
Previously reported peaks (*m/z*)						
2,081	96.0	31.7–96.5	100	100	53.1	0
3,108	10.3	0	17.1	0	79.2	100
3,123	84.1	100	95.1	100	9.4	0
3,354	4.0	1.4	12.2	NA	3.1	100
3,378	99.2	100	36.6	0	97.9	100
3,463	23.8	0	90.2	66.7	31.2	0
4,385	31.7	0–1.4	29.3	0	89.6	89.5–100
4,391	53.2	98.6–100	41.5	100	9.4	0–5.2
6,711	32.5	0	73.2	NA	90.6	100
7,637	92.6	93.2–100	90.2	100	62.5	0
7,667	11.9	3.4	2.4	0	41.7	88.1–100
8,508	7.1	4.9	14.6	NA	9.4	84.6
8,768	2.4	0–0.7	7.3	0	70.8	38.4–100
8,782	72.2	89.2–100	61.0	100	6.2	0–2.8
9,475	78.6	17–100	73.2	100	16.7	0–9.5
Novel potential peaks (*m/z*)						
2,673	88.9	NA	17.1	NA	7.3	NA
6,960	90.5	NA	9.8	NA	26.0	NA

[a]NA, not analyzed.

hinder this objective, such as the culture media used, the morphology of the colonies, or the geographic origin of the strains (15, 20). Moreover, due to the large number of protein peaks that are usually found in mass spectra, accurate identification based on only a few peaks may not be entirely reliable. Therefore, it is necessary to apply novel strategies capable of analyzing a large amount of data, such as machine learning methodologies (18, 21).

In the present study, we applied machine learning using both unsupervised and supervised algorithms. The first approach by unsupervised methodology (PCA) did not provide subspecies differentiation (Fig. 1). In the case of morphology, no important spectral differences between smooth and rough variants were observed. The main difference of these two variants is the expression of glycopeptidolipids on the surface (22, 23), and due to MALDI-TOF MS analyzing mainly ribosomal proteins, these differences were not detected. On the other hand, because this is a proof of concept and identification of mycobacteria from liquid media could be more complex (24), only solid culture media were evaluated, and differences between 7H11 and Löwenstein-Jensen were not observed. Interestingly, several clusters obtained by PCA corresponded to the geographic origin of the isolates. All strains obtained from the three Madrid hospitals (HGM, HLP, and ISCIII) were grouped together, while those from Barcelona (HUB) were separated. All isolates from Madrid were analyzed in the same hospital (HGM) and Barcelona isolates in HUB by the same operator, so differences in experience and preparation of the protein extracts were discarded. Thus, although the MALDI-TOF MS model in both centers was the same (MBT Smart Biotyper) and the acquisition of spectra was performed with the same technical parameters, some differences were still present in the mass spectra of these isolates, as observed in Fig. 1D. Moreover, strains collected in Norway and The Netherlands clustered together. Therefore, these results may suggest that there could be differences in the mass spectra of *M. abscessus* from different origins, suggestive of ecotypes, and highlight the importance of including strains from diverse origins in this type of study.

The application of supervised machine learning algorithms was targeted to the differentiation of the three subspecies. Among the four algorithms tested, the lower results were obtained by PLS-DA (Fig. S2A), SVM (Fig. S2B), and KNN (Fig. S2C), while Random Forest was able to identify a greater number of isolates at the subspecies level (Table 1). As recommended by other studies, the identification of NTM by MALDI-TOF MS should be performed in two or three replicates (11), so for more accurate identifications, we used three spots for each isolate. When the identification of the three spots was considered, the accuracy was higher in those cases where the same subspecies was obtained in all spots (Table 3). The categorical result of RF is accompanied by a probability result, so we aimed to establish a confidence cutoff to reach higher accuracy of identification. Without applying probability cutoff, RF correctly identified 90.1% of the isolates. Most of the isolates (172; 85.6%) obtained probability results above 60% (Fig. 2B), so when the cutoff was established at 60%, the accuracy rate increased to 95.9% (Fig. 2C). Moreover, by applying this cutoff, the PPV for all subspecies increased to higher than 90% (Fig. 2E).

The current model was evaluated on strains from other European countries to evaluate its application in other laboratories. Due to the identification rate being lower than that obtained with Spanish isolates, the largest model with more isolates increased this rate to 88.9% of correct identification. This fact shows that the development of bigger models can increase their accuracy and it could be easily exported to more laboratories in the future.

To analyze the protein peaks found in this study, all previously reported peaks were searched and compared with our isolates. In most cases, the detection rate of the peaks was similar to previous reports (Table 4) and, in addition, the most important peaks were found in the range of 2,000 to 10,000 Da (Fig. 2F). However, remarkable discrepancies in a few cases were found. Some peaks were found with a lower presence than reported previously: this is the case of the peak at 4,391 *m/z* (Fig. S4A) in *M. abscessus* and *M. bolletii*, with only half of our isolates presenting it; the peak around 8,782 *m/z* (Fig. S4B) that was present in 61% of *M. bolletii* isolates in comparison with 100% reported by Suzuki et al. (13) and Kehrmann et al. (16); and the peak around

7,667 *m/z* (Fig. S4C) in *M. massiliense* that was found in 47.1% of our isolates. Strikingly, peaks at 3,354 and 8,508 *m/z* were found only in a few *M. massiliense* isolates, while they were previously reported in most isolates of this subspecies (14, 15). On the other hand, peaks that were reported as absent in some subspecies, were found in some of our isolates. That was the case of 3,108 (Fig. S4D) and 4,385 *m/z* (Fig. S4A) in *M. abscessus* and *M. bolletii*; 3,123 *m/z* (Fig. S4D) in *M. massiliense*; 3,378 *m/z* in *M. bolletii* (Fig. S3A); 3,463 *m/z* (Fig. S4E) in *M. abscessus* and *M. massiliense*; and 6,711 *m/z* (Fig. S4F) in *M. abscessus*. The greater differences were in peaks at 2,081 (Fig. S3B) and 7,637 *m/z*, which have never been reported in *M. massiliense* (13, 17), and we found them in more than 50% of *M. massiliense* isolates. All these differences could have been influenced by two factors. First, it is important to include a high number of strains, representing the three subspecies to confirm that the peaks found are specific to them. The second factor is the geographic origin of the isolates. There have been reported differences in peak patterns according to the origin of the strains (15), so multicentric studies are needed to search common peaks worldwide and create accurate identification algorithms. On the other hand, two novel potential peaks have been found: 2,673 (Fig. S3C) and 6,960 *m/z* (Fig. S3D), both of them present in most *M. abscessus* isolates and in the low number of isolates from the other subspecies.

Due to the variability in the detection of peaks observed previously, the present study showed that the application of novel methodologies for data analysis, such as machine learning, could be an innovative way to improve the accuracy of MALDI-TOF MS in the identification of *M. abscessus* subspecies. Recently, other novel strategies have been evaluated for the same purpose. Jia Khor et al. (25) used the MALDI Biotyper Sirius system (Bruker Daltonics) for the detection of subspecies-specific lipids and were able to differentiate a few *M. abscessus* isolates. On the other hand, Bajaj et al. (26) evaluated for the first time the liquid chromatography-mass spectrometry for the identification of *M. abscessus* subspecies. However, these novel methods need to be validated with larger collections of clinical isolates to confirm their utility in a microbiology laboratory setting.

In conclusion, the high correct identification rate of *M. abscessus* subspecies obtained in this study, states the utility of machine learning strategy for identification purposes. The validation of the developed method could allow in the near future the automatic discrimination of *M. abscessus* subspecies in the routine of the clinical microbiology laboratory by the addition of a greater number and diversity of isolates.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 1.5 MB.

ACKNOWLEDGMENTS

D.R.-T.: conceptualization, experimentation, formal analysis, data collection, validation, visualization, original draft preparation, and review/editing. L.H., F.A., D.D., G.H.-A., J.v.I., A.V.d.B., A.I., C.B., E.T., S.B., N.V., P.M., and M.J.R.-S.: submission of isolates, writing, and review/editing. M.J.A., G.M., and L.M.: data analysis, validation, writing, and review/editing. B.R.S.: conceptualization, project administration, formal analysis, supervision, validation, visualization, original draft preparation, and review/editing.

This work was supported by projects PI15/01073, PI18/00997, and PI18/01068 from the Health Research Fund (Instituto de Salud Carlos III, Plan Nacional de I+D+I 2013–2016) of the Carlos III Health Institute (ISCIII, Madrid, Spain) partially financed by the European Regional Development Fund (FEDER) "A way of making Europe." This work was partially funded by a grant from the Spanish Society of Clinical Microbiology and Infectious Diseases (SEIMC). B.R.-S. is the recipient of a Miguel Servet contract (CPII19/00002) supported by the Health Research Foundation. D.R.-T. was funded by the Intramural Program of the Gregorio Marañón Health Research Institute.

The authors declare no conflict of interest. M.J.A., G.M., and L.M. are employees of Clover Bioanalytical Software, Ltd.

REFERENCES

1. Falkinham JO, III. 2016. Current epidemiologic trends of the nontuberculous mycobacteria (NTM). Curr Environ Health Rep 3:161–167. https://doi.org/10.1007/s40572-016-0086-z.
2. Johansen MD, Herrmann JL, Kremer L. 2020. Non-tuberculous mycobacteria and the rise of Mycobacterium abscessus. Nat Rev Microbiol 18:392–407. https://doi.org/10.1038/s41579-020-0331-1.
3. Tortoli E, Kohl TA, Brown-Elliott BA, Trovato A, Leao SC, Garcia MJ, Vasireddy S, Turenne CY, Griffith DE, Philley JV, Baldan R, Campana S, Cariani L, Colombo C, Taccetti G, Teri A, Niemann S, Wallace RJ, Jr, Cirillo DM. 2016. Emended description of Mycobacterium abscessus, Mycobacterium abscessus subsp. abscessus and Mycobacteriumabscessus subsp. bolletii and designation of Mycobacteriumabscessus subsp. massiliense comb. nov. Int J Syst Evol Microbiol 66:4471–4479. https://doi.org/10.1099/ijsem.0.001376.
4. Lopeman RC, Harrison J, Desai M, Cox JAG. 2019. Mycobacterium abscessus: environmental bacterium turned clinical nightmare. Microorganisms 7:90. https://doi.org/10.3390/microorganisms7030090.
5. Shallom SJ, Gardina PJ, Myers TG, Sebastian Y, Conville P, Calhoun LB, Tettelin H, Olivier KN, Uzel G, Sampaio EP, Holland SM, Zelazny AM. 2013. New rapid scheme for distinguishing the subspecies of the *Mycobacterium abscessus* group and identifying *Mycobacterium massiliense* isolates with inducible clarithromycin resistance. J Clin Microbiol 51:2943–2949. https://doi.org/10.1128/JCM.01132-13.
6. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, Holland SM, Horsburgh R, Huitt G, Iademarco MF, Iseman M, Olivier K, Ruoss S, von Reyn CF, Wallace RJ, Jr, Winthrop K, Infectious Disease Society of America. 2007. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. Am J Respir Crit Care Med 175:367–416. https://doi.org/10.1164/rccm.200604-571ST.
7. Kehrmann J, Kurt N, Rueger K, Bange FC, Buer J. 2016. GenoType NTM-DR for identifying *Mycobacterium abscessus* subspecies and determining molecular resistance. J Clin Microbiol 54:1653–1655. https://doi.org/10.1128/JCM.00147-16.
8. Griffith DE, Brown-Elliott BA, Benwill JL, Wallace RJ, Jr. 2015. Mycobacterium abscessus. "Pleased to meet you, hope you guess my name". Ann Am Thorac Soc 12:436–439. https://doi.org/10.1513/AnnalsATS.201501-015OI.
9. Macheras E, Roux AL, Ripoll F, Sivadon-Tardy V, Gutierrez C, Gaillard JL, Heym B. 2009. Inaccuracy of single-target sequencing for discriminating species of the *Mycobacterium abscessus* group. J Clin Microbiol 47:2596–2600. https://doi.org/10.1128/JCM.00037-09.
10. Rodriguez-Temporal D, Alcaide F, Marekovic I, O'Connor JA, Gorton R, van Ingen J, Van den Bossche A, Hery-Arnaud G, Beauruelle C, Orth-Holler D, Palacios-Gutierrez JJ, Tudo G, Bou G, Ceyssens PJ, Garrigo M, Gonzalez-Martin J, Greub G, Hrabak J, Ingebretsen A, Mediavilla-Gradolph MC, Oviano M, Palop B, Pranada AB, Quiroga L, Ruiz-Serrano MJ, Rodriguez-Sanchez B. 2022. Multicentre study on the reproducibility of MALDI-TOF MS for nontuberculous mycobacteria identification. Sci Rep 12:1237. https://doi.org/10.1038/s41598-022-05315-7.
11. Alcaide F, Amlerová J, Bou G, Ceyssens PJ, Coll P, Corcoran D, Fangous M-S, González-Álvarez I, Gorton R, Greub G, Hery-Arnaud G, Hrábak J, Ingebretsen A, Lucey B, Marekovi· I, Mediavilla-Gradolph C, Monté MR, O'Connor J, O'Mahony J, Opota O, O'Reilly B, Orth-Höller D, Oviaño M, Palacios JJ, Palop B, Pranada AB, Quiroga L, Rodríguez-Temporal D, Ruiz-Serrano MJ, Tudó G, Van den Bossche A, van Ingen J, Rodriguez-Sanchez B, European Study Group on Genomics and Molecular Diagnosis (ESGMD). 2018. How to: identify non-tuberculous Mycobacterium species using MALDI-TOF mass spectrometry. Clin Microbiol Infect 24:599–603. https://doi.org/10.1016/j.cmi.2017.11.012.
12. Teng SH, Chen CM, Lee MR, Lee TF, Chien KY, Teng LJ, Hsueh PR. 2013. Matrix-assisted laser desorption ionization–time of flight mass spectrometry can accurately differentiate between *Mycobacterium masilliense* (*M. abscessus* subspecies *bolletti*) and *M. abscessus* (*Sensu Stricto*). J Clin Microbiol 51:3113–3116. https://doi.org/10.1128/JCM.01239-13.
13. 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). Diagn Microbiol Infect Dis 83:365–370. https://doi.org/10.1016/j.diagmicrobio.2015.08.011.
14. Panagea T, Pincus DH, Grogono D, Jones M, Bryant J, Parkhill J, Floto RA, Gilligan P. 2015. Mycobacterium abscessus complex identification with matrix-assisted laser desorption ionization–time of flight mass spectrometry. J Clin Microbiol 53:2355–2358. https://doi.org/10.1128/JCM.00494-15.
15. Luo L, Liu W, Li B, Li M, Huang D, Jing L, Chen H, Yang J, Yue J, Wang F, Chu H, Zhang Z. 2016. Evaluation of matrix-assisted laser desorption ionization–time of flight mass spectrometry for identification of *Mycobacterium abscessus* subspecies according to whole-genome sequencing. J Clin Microbiol 54:2982–2989. https://doi.org/10.1128/JCM.01151-16.
16. Kehrmann J, Wessel S, Murali R, Hampel A, Bange FC, Buer J, Mosel F. 2016. Principal component analysis of MALDI TOF MS mass spectra separates M. abscessus (sensu stricto) from M. massiliense isolates. BMC Microbiol 16:24. https://doi.org/10.1186/s12866-016-0636-4.
17. Fangous MS, Mougari F, Gouriou S, Calvez E, Raskine L, Cambau E, Payan C, Hery-Arnaud G. 2014. Classification algorithm for subspecies identification within the *Mycobacterium abscessus* species, based on matrix-assisted laser desorption ionization–time of flight mass spectrometry. J Clin Microbiol 52:3362–3369. https://doi.org/10.1128/JCM.00788-14.
18. Weis C, Cuenod A, Rieck B, Dubuis O, Graf S, Lang C, Oberle M, Brackmann M, Sogaard KK, Osthoff M, Borgwardt K, Egli A. 2022. Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. Nat Med 28:164–174. https://doi.org/10.1038/s41591-021-01619-9.
19. Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Moller N, Aarestrup FM. 2014. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. J Clin Microbiol 52:139–146. https://doi.org/10.1128/JCM.02452-13.
20. Dhieb C, Normand AC, Al-Yasiri M, Chaker E, El Euch D, Vranckx K, Hendrickx M, Sadfi N, Piarroux R, Ranque S. 2015. MALDI-TOF typing highlights geographical and fluconazole resistance clusters in Candida glabrata. Med Mycol 53:462–469. https://doi.org/10.1093/mmy/myv013.
21. Weis CV, Jutzeler CR, Borgwardt K. 2020. Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review. Clin Microbiol Infect 26:1310–1317. https://doi.org/10.1016/j.cmi.2020.03.014.
22. Gutierrez AV, Viljoen A, Ghigo E, Herrmann JL, Kremer L. 2018. Glycopeptidolipids, a double-edged sword of the Mycobacterium abscessus complex. Front Microbiol 9:1145. https://doi.org/10.3389/fmicb.2018.01145.
23. Howard ST, Rhoades E, Recht J, Pang X, Alsup A, Kolter R, Lyons CR, Byrd TF. 2006. Spontaneous reversion of Mycobacterium abscessus from a smooth to a rough morphotype is associated with reduced expression of glycopeptidolipid and reacquisition of an invasive phenotype. Microbiology (Reading) 152:1581–1590. https://doi.org/10.1099/mic.0.28625-0.
24. Rodriguez-Temporal D, Rodriguez-Sanchez B, Alcaide F. 2020. Evaluation of MALDI biotyper interpretation criteria for accurate identification of nontuberculous mycobacteria. J Clin Microbiol 58:e01103-20. https://doi.org/10.1128/JCM.01103-20.
25. Jia Khor M, Broda A, Kostrzewa M, Drobniewski F, Larrouy-Maumus G. 2021. An improved method for rapid detection of mycobacterium abscessus complex based on species-specific lipid fingerprint by routine MALDI-TOF. Front Chem 9:715890. https://doi.org/10.3389/fchem.2021.715890.
26. Bajaj AO, Slechta ES, Barker AP. 2022. Rapid and accurate differentiation of Mycobacteroides abscessus complex species by liquid chromatography-ultra-high-resolution orbitrap mass spectrometry. Front Cell Infect Microbiol 12:809348. https://doi.org/10.3389/fcimb.2022.809348.