Improved subspecies identification in clinical Mycobacterium abscessus complex isolates using whole genome sequencing

Dachuan Lin
Shenzhen University
Ruifeng Sun
PolyU
Yaoju Tan
State key laboraory of Respiratory disease
Jing Wang
Yangzhou University
Xinchun Chen (chenxinchun@szu.edu.cn)
Yun Li
GuangzhouChest Hospital

Research article

Keywords: Mycobacterium abscessus, taxogenomics, whole genome sequence

Posted Date: March 26th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-19328/v1

Abstract

Background

Mycobacterium abscessus complex, which is frequently reported to cause a variety of skin and soft tissues diseases in humans, is composed of three subspecies, namely, M. abscessus subsp. abscessus, M. abscessus subsp. massiliense and M. abscessus subsp. bolletii. Currently, the differentiation of these three subspecies in clinical isolates still largely depends on single gene identification methods, including the genes hsp65 and 16s, with limited accuracy.

Results

This study confirmed the limitations of the single gene-based method of subspecies identification. We performed a comprehensive analysis of MABC genomes in the NCBI database and tried to build an accurate and user-friendly identification method. Here, we describe an improved assay for Mycobacterium abscessus complex rapid identification using WGS data, based on the identities of rpoB, erm(41) and rpls. Comprehensive analysis has been performed to compare our software results with the traditional method. The results showed that the method built in this study could achieve 100% identification of subspecies for the Mycobacterium abscessus complex in a public genome database (893 genomes from the NCBI database and 6 clinical isolates from this study). This software can be easily integrated into a routine workflow to quickly and precisely provide subspecies-level identification and discrimination of different MABC subspecies in clinical isolates by WGS.

Conclusions

This assay will facilitate accurate molecular identification of species from the MABC complex in a variety of clinical specimens and diagnostic contexts.

Background

Mycobacterium abscessus complex (MABC), which causes a range of diseases from skin infections to pulmonary infections [1, 2], is a rapidly growing mycobacterium that has become an emerging pathogen. MABC is notorious not only because it can accelerate inflammatory damage, leading to increased morbidity and mortality, but also because it causes cystic fibrosis (CF), which has become the most lethal and frequent infection. The notorious MABC complex causes diseases that are called nightmares in the clinical field, not only because M. abscessus is the second most common nontuberculous mycobacterial species associated with lung disease but also because of the intrinsic and acquired resistance of Mycobacterium abscessus to commonly used antibiotics, which limits the chemotherapeutic options for infections caused by these mycobacteria. Therefore, the infections caused by MABC, especially multidrug resistance strains, are very difficult to treat [3–7] and sometimes impossible to treat, even in developed countries [8, 9].

Despite their high genome similarity, members of the MABC usually have distinct phenotypes in culture, specifically antibiotic resistance patterns, particularly regarding the critical first line antibiotic treatment clarithromycin[10, 11], and cause differing treatment outcomes for patients infected with M. abscessus subsp. abscessus versus M. abscessus subsp. massiliense[12]. Different treatment requirements and outcomes are thought to vary among different subspecies[13]; thus, it is clinically significant to differentiate MABC. Therefore, the main purpose of this study was to construct an accurate and user-friendly method for MABC subspecies identification.

The MABC represent a diverse and clinically important family of bacteria. Although controversy still exists about the taxonomy and nomenclature of M. abscessus subspecies, most researchers believe that MABC comprises three subspecies: M. abscessus subsp. abscessus, M. abscessus subsp. massiliense and M. abscessus subsp. bolletii[14, 15]. With technical improvement, debates (reclassification, elevated to species level) regarding the classification of Mycobacterium abscessus have existed since its discovery in 1953[16], and even during the past 30 years, the MABC taxonomic classification has undergone a series of changes from single species to different species, then to three subspecies[17]. The debate still continues today[18], so one of the purposes of this study was to further clarify the debate by undertaking genome comparison.

Prior work demonstrated that the commonly used 16S rRNA was not sufficient to identify MABC[19], and only using rpoB could lead to misidentification for Mycobacterium[23]. Thus, in this study we evaluated the current taxonomic methods based on 16S rRNA, rpoB, erm(41) and erm(42) in order to find a good MABC taxonomic classification method by implementing the combination of these marker genes. Furthermore, it is critical to assess species boundaries among different species or subspecies for taxonomy, and thus a special cut-off value is needed for different species, even for the traditional identification genes. Therefore, this study comprehensively compared the genomes in databases and selected the cut-off for MABC taxonomic classification. Additionally, to our knowledge, no specific genes have been identified to discriminate M. abscessus subsp. bolletii (the traditional rpoB gene is not good enough for subspecies identification). WGS-based protocols have so far not been developed for the subspecies level, similar to MABC identification. Therefore, one of the purposes of this study was to identify specific genes for M. abscessus subsp. bolletii identification by implementing the combination of several genes and to build a powerful and reliable taxonomical tool for MABC based on minimum and maximum identity values.

Methods

Bacterial isolates A total of 6 MABC isolates were obtained from different patients from the same hospital during the same time period (2009). Of them, 2 strains were isolated from sputum samples, and 4 were obtained from bronchoalveolar lavage fluid (BALF). All of the isolates had been classified as M. abscessus based on the results of biochemistry following Leao's method[20].

DNA Extraction Bacterial DNA extraction was performed as described previously[21]. In brief, the bacterium was harvested after growing in Middlebrook 7H9 liquid medium for 5 days, and then the samples were crushed with zirconia beads (1 mm in diameter) in a tissue disintegrator instrument. Total genomic DNA was extracted from the crushed suspension using a commercial ethanol precipitation kit according to the manufacturer's instructions and stored at -20 °C. Samples were sequenced on the Illumina HiSeq X10 sequencer, using 150 base-pair paired-end reads.

Genome assembly After checking the length and quality of the reads with FastQC (Version 0.11.8)[22], the reads were de novo assembled with SPAdes (v3.13.0), using the '-careful' setting and k-mers 21, 33, 55, 77, and 99[23, 24], and the assembled genome was manually trimmed for short contigs (length less than 3,000 bp) and very low coverage contigs (less than 10). The results of the assembly were evaluated using QUAST (Version v.5.0.2, http://quast.bioinf.spbau.ru/;)[25].

In vitro drug susceptibility testing The microdilution method was used following the recommendations of the Clinical and Laboratory Standards Institute (CLSI) for rapidly growing mycobacteria[26]. The susceptibility results to ciprofloxacin, moxifloxacin and clarithromycin were judged by the established breakpoints from the CLSI document (M24-A2-2011).

To evaluate the traditional identification taxonomy method, 16S rRNA (1,468 bp), rpoB (409 bp), erm(41) (GenBank accession number: CU458896.1), and erm(42) (GenBank accession number: FJ358487.1) were selected for comparison[27, 28]. Nucleotide sequences were extracted from the reference sequences of M. abscessus subsp. abscessus ATCC 19977 and Mycobacterium_abscessus_subsp_bolletii CIP 108541 using traditional primers (the primer used in this study is listed in supplemental table 1). For comparison, the erm(41) fragment was selected with the same beginning and end as the erm(42) fragment. The identities among the fragments were obtained by BLASTN searches with 1E-5 as a cut-off value [29]. Additionally, the fragments from the genomes were obtained by TBtools[30].

To test the method used in this study, all of the genomes (complete, chromosome, scaffold) were downloaded from NCBI genome (https://www.ncbi.nlm.nih.gov/genome/genomes/1360?). To evaluate the subspecies of MABC, average nucleotide identity (ANI), average amino acid identity (AAI) and genome to genome distance (GGDC) were calculated among the typical strains using the Kostas Lab two-way average AAI calculator (http://enve- omics.ce.gatech.edu/aai/)[31], ANI (http://enve-omics.ce.gatech.edu/)[32], OrthoANI calculator[33], and Genome-to-Genome Distance Calculator (http://ggdc.dsmz.de/ggdc.php)[34], and fastANI[40] was also used to estimate the ANI among the genome datasets constructed in this study. To obtain the most effective gene to distinguish M. bolletii from other MABC, the Up to date Bacterial Core Gene (UBCG) tool was chosen to select genes from among the 92 bacterial core genes[35]. The phylogenetic tree was visualized using iTOL[36].

Results

Patients were primarily male (4/6 patients) and greater than 65 years of age (4/6 patients). Antimicrobial resistance patterns of M. abscessus subsp. abscessus and M. abscessus subsp. massiliense isolated

from respiratory specimens are shown in supplemental table 2. The results confirmed the innate resistance of MABC to several drug classes, such as cefoxitin. Fortunately, those isolates were still sensitive to clarithromycin within 3 days. Isolate 1 was the only one with observed resistance to clarithromycin at 7 days. The genome analysis results also showed that no known resistance-conferring mutation in the 23S rRNA gene was observed in those isolates.

Then, the study first verified whether MABC belonged to the same species or different species. The AAI, ANI, and GGDC results of the typical isolates of MABC are shown in table 1. Among the representative genome datasets as previously reported (unpublished result), using the full length ropB obtained from M. abscessus subsp. abscessus ATCC 19977 as the reference, the similarities among MABC complex representative isolates ranged from 97.78%-100% (table 1 and Fig. 1). All results showed that the MABC complex belonged to the same species.

Then, this study tested whether WGS comparison could further distinguish subspecies among the MABC complex. The ANI values for Mycobacterium abscessus subsp. massiliense ranged from 96.72%-100% (Supplemental table 6), and the ANI for Mycobacterium abscessus subsp. bolletii ranged from 96.1263%-100% (Supplemental table 7). Because of the overlapping value of ANI, the ANI value could not be used to distinguish subspecies for the MABC complex. Supplemental Fig. 1 also shows that although the ANI value could distinguish the species level, it could not distinguish the sub species that are within the same species. Then, another WGS-based method, UBCG, was used for subspecies verification. In accordance with our findings, the phylogenomic analyses of the UBCG (92 genes) of isolates from the public genome database showed the clear distance separating the three subspecies of the M. abscessus complex (Fig. 2A). The genome with the subspecies names labeled is shown in Fig. 2. From Fig. 2A, we can see that although many genomes were not labeled with subspecies in the NCBI dataset, all genomes with clear subspecies could be separated from other subspecies by different branches. Therefore, the UBCG is a good standard for subspecies verification. However, using the UBCG method for subspecies classification takes a long time and possesses ambiguous boundaries (there is no clear cut-off value for the UBCG method). Thus, this study further tested the representative genes for species identification.

First, in order to obtain the most accurate gene for identifying the MABC complex from other Mycobacterium spp, this study first evaluated the currently used gene fragment (Supplemental table 1). When we tested the traditional marker for MABC identification, the hsp65 and 16S–23S ITS primers could not be found in the genomes for the representative strains (Supplemental table 1). Therefore, this study only compared the 16S and rpoB genes. Then, this study found that when using the fragment obtained by the primers rpoB-mycoF and rpoB-mycoR, the identity between the fragments from M. abscessus subsp. abscessus ATCC 19977 and M. abscessus subsp. bolletii CIP 108541 was 96.81. As the identity was below 97%, which was settled by the traditional cut-off value, this study then used the full-length rpoB genes for comparison. The similarities of the 16S and rpoB genes are shown in Fig. 1. As shown in Fig. 1, compared with 16S identity, rpoB gene identity could better distinguish the MABC from other Mycobacterium spp; 16S identity could not distinguish the MABC complex from similar species such as Mycobacterium_salmoniphilum, Mycobacterium_immunogenum, Mycobacterium_stephanolepidis,

Mycobacterium_franklinii, Mycobacterium_chelonae, and Mycobacterium_saopaulense. To verify whether the cut-off value could suit all the MABC complex isolates, the genome dataset downloaded from NCBI was used for testing. As shown in Supplemental table 4, the rpoB identities ranged from 97.697%- 100% (893 isolates from 910 total isolates; 7 of the genomes in the public databases were labeled as contaminated and were excluded from the dataset. Another 10 isolates in the genome database are shown in Supplemental table 4, and all genome-based identification methods showed that they were very far from the MABC complex). Therefore, the full-length rpoB gene and the cut-off value of 97% could be used for MABC complex identification.

Then, this study tested whether rpoB could distinguish subspecies. This study first utilized the dataset of M. abscessus subsp. bolletii from NCBI to test whether the full-length rpoB gene could be used as a subspecies marker. However, when using the rpoB gene from M. abscessus subsp. abscessus ATCC 19977, the identities ranged from 97.697–97.896%, and using the rpoB gene from M. abscessus subsp. massiliense, the identities ranged from 98.052–99.742%. The M. abscessus subsp. bolletii CIP 108541 identities ranged from 97.811 to 100%. As the ranges overlapped with each other, using different rpoB genes could not distinguish other subspecies.

After using rpoB for MABC species identification, this study then verified the current method for subspecies identification. This study first selected the erythromycin ribosomal methylase gene, erm(41), to further distinguish the subspecies. The identities of erm(41) and erm(42) are shown in supplemental table 8. When using the erm(41) fragment from Mycobacterium abscessus subsp. abscessus ATCC 19977 as a reference, the identities ranged from 97.893%-100% (586/893), and the range for the erm(42) fragment from Mycobacterium abscessus subsp. massiliense (GenBank accession number: FJ358487.1) was 99.58%-100% (307/893). Because there were no genomes that shared identities with both erm(41) and erm(42), this study used the identity of 99.58% for erm(42) as the cut-off value for selecting Mycobacterium abscessus subsp. massiliense from the MABC complex. Then, already labeled Mycobacterium abscessus subsp. were tested for classification, and the results are shown in Fig. 2B. From Fig. 2B, we can see there were no crosses between erm(41) and erm(42) (the genomes that contained both erm(41) and erm(42) are located in different branches). The figure shows that the erm(42) genes exactly matched Mycobacterium abscessus subsp massiliense. This is also the case for erm(41) with Mycobacterium abscessus subsp bolletii and the isolates without special labels. Therefore, the erm(42) fragment with the cut-off value of 99.58% was used for the identification of Mycobacterium abscessus subsp. massiliense.

To further distinguish Mycobacterium abscessus subsp. abscessus and M. abscessus subsp. bolletii, this study first tested a traditional marker: the rpoB fragment. According to the traditional method, M. abscessus subsp. bolletii genomes should be grouped together according to the identity of the rpoB fragment. This study tested the identities of the rpoB fragment from M. abscessus subsp. bolletii with the rpoB fragment from all three species of the MABC complex. The rpoB identities among the M. abscessus subsp. bolletii were from 95.88–99.73% when using the fragment from Mycobacterium abscessus subsp. massiliense GO 06. The identities were from 94.16–100% when using the fragment from Mycobacterium

abscessus subsp. abscessus ATCC 19977, and the identities were from 95.88–100% when using the fragment from M. abscessus subsp. bolletii CIP 108541 (Supplemental table 9). As the ropB gene could not distinguish M. abscessus subsp. bolletii, this study tried to find a marker gene for M. abscessus subsp. bolletii from the UBCG-selected genes. All 92 genes tested in UBCG were evaluated in this study. From these individual gene trees, when using the rpls gene to draw the phylogenetic tree, the fewest genomes had been labeled with Mycobacterium abscessus subsp. bolletii (Fig. 3A). Then, in order to obtain the cut-off value for the rpls gene for Mycobacterium abscessus subsp. bolletii identification, this study first used full-length rpls gene identities (Supplemental table 10 A). Because the identities of two genomes were a little far from the other genomes, this study selected the partial common rpls gene fragment shared by Mycobacterium abscessus subsp. bolletii for a reference (Supplemental file 11). Although Mycobacteroides abscessus subsp. bolletii 50594 was labeled as Mycobacteroides abscessus subsp. bolletii, it carried erm(42) instead of erm(41). Furthermore, this strain was also labeled as Mycobacterium massiliense 50594 (heterotypic synonym). Therefore, this study excluded this genome from the Mycobacteroides abscessus subsp. bolletii dataset, and the cut-off value for rpls fragment identity was 99.53% (Supplemental table 10B). Then, the study used the fragment and the cut-off value to test all MABC complex genomes from NCBI. The genomes with rpls fragment identity > 99.53% and erm(42) fragment identity < 99.58% were exactly the branch of Mycobacteroides abscessus subsp. bolletii (Fig. 2C, A and Supplemental table 10 C). Therefore, the rpls gene fragment and the cut-off value of 99.53% were selected for Mycobacterium abscessus subsp. bolletii identification.

The clinical isolates involved in this study were submitted to NCBI with the accession number PRJNA594106. Finally, this study summarized the data analyzing (Fig. 4) and contructed the software NucleotideQuery, which is accurate and user-friendly (no need for installation with the ability to detect genomes directly), for MABC complex classification (Attachment file 12).

Discussion

All three M. abscessus subspecies had been reported to cause serious infections such as pulmonary infections [37], and controversies about the taxonomic status of MABC have persisted. The proposed and generally accepted species boundaries for AAI, ANI and dDDH values are 95%, 95 ~ 96% and 70%, respectively [34, 38, 39]. As shown in Table 1, the AAI values among the MABC typical strains were 97.61%-97.95% and were all larger than the cut-off value 95%. The ANI values among the MABC typical strains were 96.94–97.39% (the results obtained by different software programs were similar as shown in Table 1), and they were all larger than the cut-off value of 96%. The GGDC values among the typical MABC strains were 73.4–77.2%. The above results of this study supported the current classification of M. abscessus subsp. abscessus, M. abscessus subsp. massiliense and M. abscessus subsp. bolletii as belonging to the same species.

The most common method to identify the MABC complex isolates is based on rpoB, but this method has been questioned[40]. Additionally, the ranges for the cut-off value of rpoB cross over each other, so this study believed rpoB fragments could not be used for M. abscessus subsp. bolletii identification. A

previous study also confirmed that a single gene was not sufficient for subspecies taxonomy[51]. Other methods such as pulsed-field gel electrophoresis are complex and costly and are not suitable for the clinical field[52]. To identify the most effective gene for distinguishing M. abscessus. subsp. bolletii, all 92 genes selected by the UBCG were evaluated by the number of the smallest group including M. abscessus. subsp. bolletii. Then, this study obtained the cut-off value for M. abscessus. subsp. bolletii. After verifying the selected gene rpls and the cut-off value using the MABC complex dataset, this study demonstrated that the rpls fragment with the identity cut-off value of 99.53% could serve as a unique M. abscessus. subsp. bolletii identification marker.

The ANI has been validated as a prokaryotic species taxonomy tool, where ANI values higher than 95–96% are consistent with strains belonging to the same species. To our knowledge, no cut-off has been proposed to define the boundary of subspecies. Although most of the M. abscessus subsp. abscessus genomes were not labeled in the genome database, the ANI value could not distinguish M. abscessus subsp. bolletii from M. abscessus subsp. massiliense (attachment table 6 and 7). This study also showed that the ANI value was not suitable to distinguish the subspecies of the MABC complex. However, when using WGS assemble and then analyzing with our software, this rapid approach achieves accurate, user-friendly MABC complex subspecies identification using rpoB, erm(42) and rpls that is sufficiently reliable to serve as the routine methodology in the clinical field. Thus, our method has high clinical meaning.

Conclusion

This study reported genome sequences and genomic features of 6 MABC isolates derived from one hospital in China. The results of this study showed that the distance among the MABC complex was insufficient to warrant distinction at the species level, as there are subspecies in the MABC complex. Additionally, after verifying the taxonomy of the MABC complex, we developed a user-friendly accurate method based on a set of genes for differentiation at the subspecies level for M. abscessus subsp. massiliense, M. abscessus subsp. bolletti and M. abscessus subsp. abscess. Using all available sequences to test the method, the sequence analysis of public databases indicated that the combination of this design could provide high discrimination among the MABC complex.

Whole-genome sequencing technology has become more widely used in clinical diagnosis, and developing more user-friendly software would greatly facilitate the acquisition of more precise information about pathogens to aid in the choice of more discriminative therapies.

Declarations

Abbreviations

MABC *Mycobacterium abscessus* complex

CF cystic fibrosis

CLSI Clinical and Laboratory Standards Institute

ANI average nucleotide identity

AAI average amino acid identity

GGDC genome to genome distance

Declarations

Not applicable

Consent to publish

Not applicable

Competing Interests

None of the authors have any conflict of interest to declare.

Authors' contributions

DC conceived and designed the work; RF designed the software; LY collected the clinical isolates; JW revised the manuscript; YJ proved the material and partially wrote the manuscript; XC instructed and supervised all work; all authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Funds of China (81802060) and the start-up funding of Shenzhen University.

Acknowledgements

The authors thank Dr. Dong Yun for critical reading of the manuscript and members of the Xinchun lab for useful discussion.

References

1. RM, D., et al., *Genome sequencing of Mycobacterium abscessus isolates from patients in the united states and comparisons to globally diverse clinical strains.* Journal of Clinical Microbiology, 2014. **52**(10): p. 3573-82.
2. KK, C., et al., *First United States Outbreak of Mycobacterium abscessus Hand and Foot Disease Among Children Associated With a Wading Pool.* Journal of the Pediatric Infectious Diseases Society, 2018. **undefined**(undefined): p. undefined.

3. Daniel-Wayman, S., et al., *Antibiotic Susceptibility of Mycobacterium Abscessus Patients with Long Term Antibiotic Treatment*, in *A106. NTM CLINICAL AND TRANSLATIONAL SCIENCE*2018, American Thoracic Society. p. A2593-A2593.
4. R, N., et al., *Mycobacterium abscessus: a new antibiotic nightmare.* The Journal of antimicrobial chemotherapy, 2012. **67**(4): p. 810-8.
5. Stout, J.E., W.J. Koh, and W.W. Yew, *Update on pulmonary disease due to non-tuberculous mycobacteria.* International Journal of Infectious Diseases, 2016. **45**(C): p. 123-134.
6. Kwon, Y.S. and W.J. Koh, *Diagnosis and Treatment of Nontuberculous Mycobacterial Lung Disease.* Journal of Korean Medical Science, 2016. **31**(5): p. 649-659.
7. A, K., et al., *Activity of the New β-Lactamase Inhibitors Relebactam and Vaborbactam in Combination with β-Lactams against Mycobacterium abscessus Complex Clinical Isolates.* Antimicrob Agents Chemother, 2019. **63**(3): p. undefined.
8. Bryant, J.M., et al., *Emergence and spread of a human-transmissible multidrug-resistant nontuberculous mycobacterium.* Science, 2016. **354**(6313): p. 751-757.
9. Boeck, L., et al., *Late Breaking Abstract-Multidimensional genomic mapping reveals the pathophysiology of Mycobacterium abscessus*, 2018, Eur Respiratory Soc.
10. Meng-Rui, L., et al., *Mycobacterium abscessus Complex Infections in Humans.* Emerging infectious diseases, 2015. **21**(9): p. 1638-1646.
11. Toshiyuki, H., et al., *Clinical and microbiological differences between Mycobacterium abscessus and Mycobacterium massiliense lung diseases.* Journal of Clinical Microbiology, 2012. **50**(11): p. 3556-61.
12. Won-Jung, K., et al., *Clinical significance of differentiation of Mycobacterium massiliense from Mycobacterium abscessus.* American Journal of Respiratory & Critical Care Medicine, 2011. **183**(3): p. 405.
13. Lipworth, S., et al., *Whole-Genome Sequencing for Predicting Clarithromycin Resistance in Mycobacterium abscessus.* Antimicrob Agents Chemother, 2019. **63**(1): p. e01204-18.
14. DE, G., *Mycobacterium abscessus subsp abscessus lung disease: 'trouble ahead, trouble behind…'.* F1000prime reports, 2014. **6**(undefined): p. 107.
15. K, N., et al., *Discrimination of Mycobacterium abscessus subsp. massiliense from Mycobacterium abscessus subsp. abscessus in clinical isolates by multiplex PCR.* Journal of Clinical Microbiology, 2014. **52**(1): p. 251-9.
16. E, T., et al., *Emended description of Mycobacterium abscessus, Mycobacterium abscessus subsp. abscessus and Mycobacteriumabscessus subsp. bolletii and designation of Mycobacteriumabscessus subsp. massiliense comb. nov.* International journal of systematic and evolutionary microbiology, 2016. **66**(11): p. 4471-4479.
17. Lee, M.-R., et al., *Mycobacterium abscessus complex infections in humans.* Emerging infectious diseases, 2015. **21**(9): p. 1638.

18. Adekambi, T., et al., *Reinstating Mycobacterium massiliense and Mycobacterium bolletii as species of the Mycobacterium abscessus complex.* International journal of systematic and evolutionary microbiology, 2017. **67**(8): p. 2726-2730.
19. M, B., et al., *Careful use of 16S rRNA gene sequence similarity values for the identification of species.* New microbes and new infections, 2018. **22**(undefined): p. 24-29.
20. Leao, S.C., et al., *Practical handbook for the phenotypic and genotypic identification of mycobacteria.* Vanden Broele, Brugges, Belgium, 2004.
21. Votintseva, A.A., et al., *Mycobacterial DNA Extraction for Whole-Genome Sequencing from Early Positive Liquid (MGIT) Cultures.* Journal of Clinical Microbiology, 2015. **53**(4): p. 1137-43.
22. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*, 2010, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
23. S, N., et al., *Assembling single-cell genomes and mini-metagenomes from chimeric MDA products.* Journal of computational biology : a journal of computational molecular cell biology, 2013. **20**(10): p. 714-37.
24. A, B., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* Journal of computational biology : a journal of computational molecular cell biology, 2012. **19**(5): p. 455-77.
25. Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies.* Bioinformatics, 2013. **29**(8): p. 1072-1075.
26. Woods, G., et al., *Susceptibility testing of mycobacteria, nocardiae, and other aerobic actinomycetes*2011: Clinical and Laboratory Standards Institute.
27. Simmon, K.E., et al., *Simultaneous sequence analysis of the 16S rRNA and rpoB genes by use of RipSeq software to identify Mycobacterium species.* Journal of Clinical Microbiology, 2010. **48**(9): p. 3231-3235.
28. T, A., C. P, and D. M, *rpoB-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria.* Journal of Clinical Microbiology, 2003. **41**(12): p. 5699-708.
29. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**(1): p. 421.
30. Chen, C., et al., *TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface.* BioRxiv, 2018: p. 289660.
31. Rodriguez-R, L. and K. Konstantinidis, *Bypassing cultivation to identify bacterial species. Microbe 9: 111–118*, 2014.
32. Rodriguez-R, L.M. and K.T. Konstantinidis, *The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes*, 2016, PeerJ Preprints.
33. I, L., et al., *OrthoANI: An improved algorithm and software for calculating average nucleotide identity.* International journal of systematic and evolutionary microbiology, 2016. **66**(2): p. 1100-1103.
34. JP, M.-K., et al., *Genome sequence-based species delimitation with confidence intervals and improved distance functions.* BMC Bioinformatics, 2013. **14**(undefined): p. 60.

35. Na, S.I., et al., *UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction.* Journal of Microbiology, 2018. **56**(4): p. 1-6.
36. Letunic, I. and P. Bork, *Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees.* Nucleic Acids Research, 2016. **44**(Web Server issue): p. gkw290.
37. Kazue, N., et al., *Multiple cases of cutaneous Mycobacterium massiliense infection in a "hot spa" in Japan.* Journal of Clinical Microbiology, 2011. **49**(2): p. 613.
38. Goris, J., et al., *DNA-DNA hybridization values and their relationship to whole-genome sequence similarities.* Int J Syst Evol Microbiol, 2007. **57**(1): p. 81-91.
39. KT, K., R.-M. R, and A. R, *Uncultivated microbes in need of their own taxonomy.* The ISME journal, 2017. **11**(11): p. 2399-2406.
40. M, S. and D. M, *Genome analysis reveals three genomospecies in Mycobacterium abscessus.* BMC genomics, 2014. **15**(undefined): p. 359.
41. R, d.Z., v.I. J, and v.S. D, *Utility of rpoB gene sequencing for identification of nontuberculous mycobacteria in the Netherlands.* Journal of Clinical Microbiology, 2014. **52**(7): p. 2544-51.
42. A, M., et al., *Assessment of partial sequencing of the 65-kilodalton heat shock protein gene (hsp65) for routine identification of Mycobacterium species isolated from clinical sources.* Journal of Clinical Microbiology, 2004. **42**(7): p. 3000-11.
43. Edouard, M., et al., *Multilocus sequence analysis and rpoB sequencing of Mycobacterium abscessus (sensu lato) strains.* Journal of Clinical Microbiology, 2011. **49**(2): p. 491.
44. S, Y., et al., *Further isolation of Mycobacterium abscessus subsp. abscessus and subsp. bolletii in different regions of Japan and susceptibility of these isolates to antimicrobial agents.* International journal of antimicrobial agents, 2013. **42**(3): p. 226-31.
45. C, V.-N., et al., *Molecular characterization of Mycobacterium massiliense and Mycobacterium bolletii in isolates collected from outbreaks of infections after laparoscopic surgeries and cosmetic procedures.* Journal of Clinical Microbiology, 2008. **46**(3): p. 850-5.
46. Duarte, R.S., et al., *Epidemic of postsurgical infections caused by Mycobacterium massiliense.* Journal of Clinical Microbiology, 2009. **47**(7): p. 2149-2155.
47. Nie, W., et al., *Species identification of Mycobacterium abscessus subsp. abscessus and Mycobacterium abscessus subsp. bolletii using rpoB and hsp65, and susceptibility testing to eight antibiotics.* International Journal of Infectious Diseases, 2014. **25**: p. 170-174.
48. Lee, M.-R., et al., *Bacteraemia caused by Mycobacterium abscessus subsp. abscessus and M. abscessus subsp. bolletii: clinical features and susceptibilities of the isolates.* International journal of antimicrobial agents, 2014. **43**(5): p. 438-441.
49. Adékambi, T. and M. Drancourt, *Mycobacterium bolletii respiratory infections.* Emerging infectious diseases, 2009. **15**(2): p. 302.
50. Rollet-Cohen, V., et al., *Mycobacterium bolletii Lung Disease in Cystic Fibrosis.* Chest, 2019.

51. E, M., et al., *Inaccuracy of single-target sequencing for discriminating species of the Mycobacterium abscessus group.* Journal of Clinical Microbiology, 2009. **47**(8): p. 2596-600.
52. Zelazny, A.M., et al., *Cohort study of molecular identification and typing of Mycobacterium abscessus, Mycobacterium massiliense, and Mycobacterium bolletii.* Journal of Clinical Microbiology, 2009. **47**(7): p. 1985-1995.

Table

Due to technical limitations, Table 1 is only available for download from the Supplementary Files section.

Figures

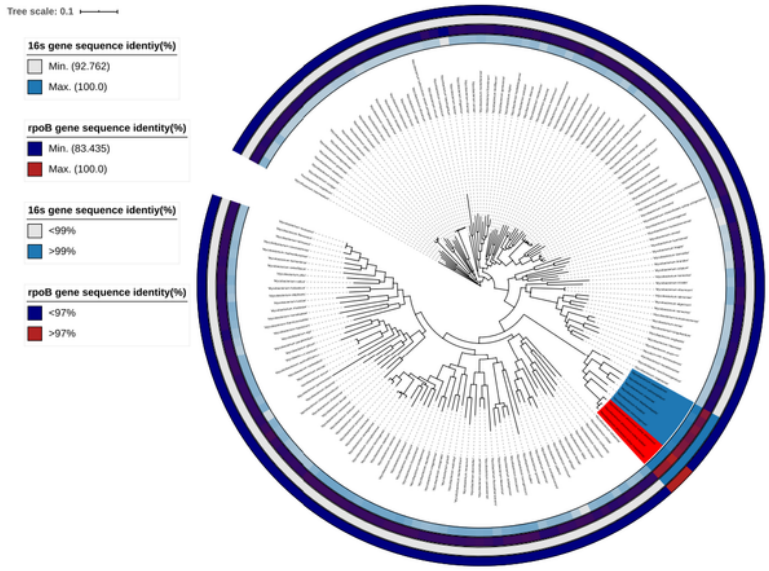

Figure 1

Phylogenetic relatedness of Mycobacterium spp and the identities of the 16s and rpoB genes The inner range is labeled with red and blue branches; red branches correspond to the MABC, while the blue branches correspond to the neighbor species that were incorrectly identified as MABC by 16s identity. The gray and light blue ranges show the identities of 16s rRNA, and the dark blue and red ranges show the

identity of rpoB. The outside two ranges show the identities of the 16s and rpoB genes with the cut-off values.

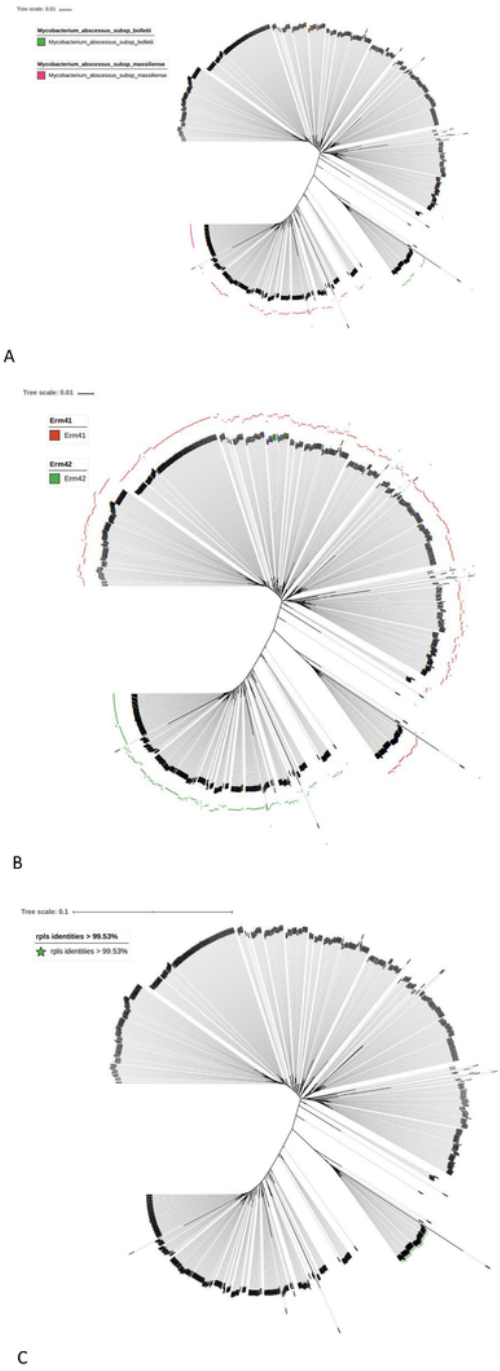

Figure 2

Phylogenetic tree for the genomes of the MABC complex. Phylogenetic tree inferred using UBCGs (concatenated alignment of 92 core genes). A Phylogenetic trees for the MABC complex from NCBI with all clear labels of M. abscessus subsp. massiliense and M. abscessus subsp. Bolletii. B Phylogenetic

trees for the MABC complex from NCBI with all clear labels of the identities of erm(41) and erm(42). C Phylogenetic trees for the MABC complex from NCBI with rpls gene identities > 99.53% and erm(42) gene identities < 99.58%.

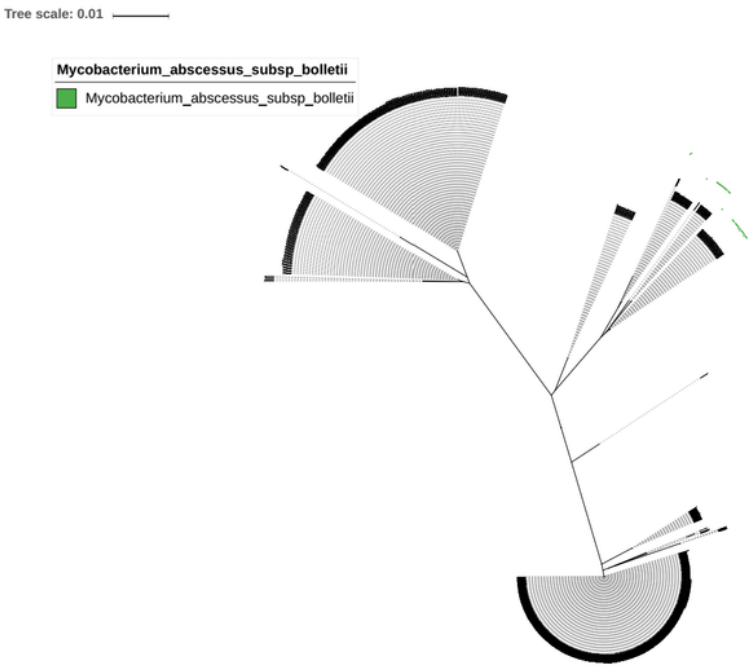

Figure 3

Phylogenomic tree for the identification of Mycobacterium abscessus subsp. bolletii A phylogenomic tree based on rpls. The green square is labeled for Mycobacterium abscessus subsp. bolletii genomes with

clear labeling as Mycobacterium abscessus subsp. bolletii in the NCBI database.

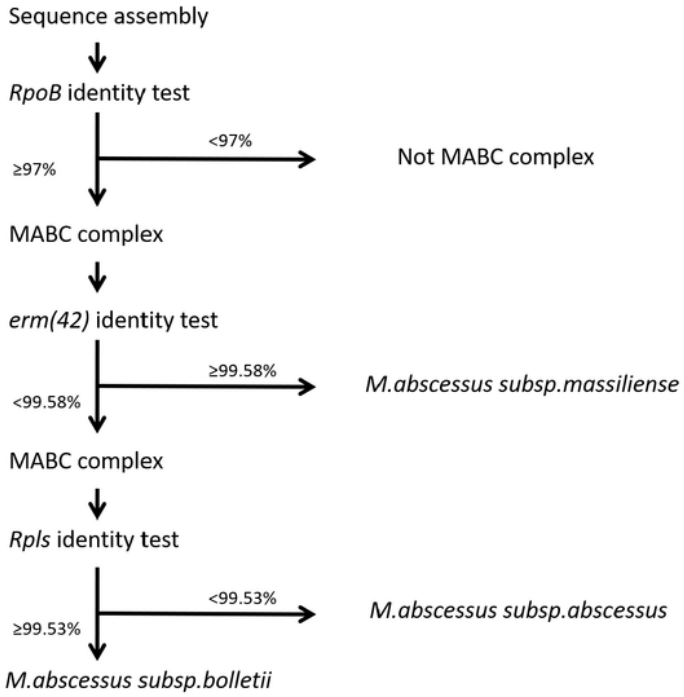

Figure 4

Flow chart of the study software

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementalfile11rplsfragement.txt
- Table1.pdf
- Supplementaltable10.xlsx
- Supplementaltable5abscessallabscess.xlsx
- Supplementaltable6.xlsx
- Supplementaltable4.xlsx
- Supplementaltable7.xlsx
- Supplementaltable123.xlsx
- Supplementaltable8.xlsx
- Supplementaltable9.xlsx