bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Dachuan Lin[a], Ruifeng Sun[b], Yaoju Tan[c], Jing Wang[d], Xinchun Chen[a*]

[a]Guangdong Key Laboratory of Regional Immunity and Diseases, Department of Pathogen Biology, Shenzhen University School of Medicine, Shenzhen 518054, China

[b] Office of Student Resources and Residential Life, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

[c] State Key Laboratory of Respiratory Disease, Department of Clinical Laboratory, Guangzhou Chest Hospital, Guangzhou 510095, China

[d]

Key Laboratory of Prevention and Control of Biological Hazard Factors (Animal Origin) for Agrifood Safety and Quality, Ministry of Agriculture of China, Yangzhou University, Yangzhou 225009, China

*Corresponce to Xinchun Chen (chenxinchun@szu.edu.cn)

Mycobacterium abscessus complex, which is frequently reported causing 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 depend on single gene identification methods like the genes namely *hsp65*, *16s* with a limited accuracy. This study confirmed the limitations of the single gene based method in the subspecies identification. We performed a comprehensive analysis of MABC genomes in the NCBI database and tried to build an accurate and user-friendly identify method. Here, we describe an improved assay for *Mycobacterium abscessus* complex fast 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 result showed that the method built-in this study could 100% identification the subspecies for the *Mycobacterium abscessus* complex in the public genome database (893 genomes from NCBI database and 6 clinical isolates from this study). Because this software can be

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

34 easily integrated into a routine workflow to quickly and precisely provide
35 subspecies-level identification and discrimination MABC different subspecies in
36 clinical isolates by WGS. This assay will facilitate accurate molecular identification of
37 species from the MABC complex in a variety of clinical specimens and diagnostic
38 contexts.

39 **Key words:**
40 *Mycobacterium abscessus*, taxogenomics, whole genome sequence

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

64 Introduction

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

78 Despite the high genome similarity, the members of the MABC usually have
79 distinct phenotypes in culture, antibiotic resistance pattern, particularly the critical
80 first line antibiotic treatment-clarithromycin[10, 11], and especially cause differing
81 treatment outcomes for patients infected with *M. abscessus subsp. abscessus* versus
82 *M. abscessus subsp. massiliense*[12]. Different treatment requirements and outcomes
83 are thought to vary among the different subspecies[13], thus it is clinically
84 significant to differentiate MABC. So the main purpose of this study was to
85 construct an accuracy and user-friendly method for MABC subspecies identification.

86 The MABC represent a diverse and clinically important family of bacteria.
87 Although controversy still exists about the taxonomy and nomenclature of *M.*
88 *abscessus* subspecies, most of the researchers believed that MABC comprised three
89 subspecies: *M. abscessus subsp. abscessus*, *M. abscessus subsp. massiliense* and *M.*
90 *abscessus subsp. bolletii*[14, 15].With the technical improvement, the debates
91 (re-classification, elevated to species level) of classification of *Mycobacterium*
92 *abscessus* had companied the discovery since 1953[16], even during the past 30
93 years the MABC taxonomic classification still have a serial changes from single

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

94 species to different species then to three subspecies[17]. And the debate still last
95 nowadays[18], so one of the purposes of this study was to further clarify the debate
96 by the genome comparison.

97 Prior work had demonstrated the commonly used 16S rRNA was not good
98 enough to identify MABC[19], and only using *rpoB* also could lead misidentifying
99 for Mycobacterium[23]. Thus, in this study we evaluated the current taxonomic
100 methods based on 16S rRNA, *rpoB*, *erm*(41) and *erm*(42) and in order to find a good
101 MABC taxonomic classication method by the combination of those marker genes
102 together. Another thing is as we all known assessing species boundaries among
103 different species or subspecies is critical for taxonomy, thus special cut-off value is
104 needed for different species, even for the traditional identification genes. So this
105 study comprehensive compared the genome in database and selected the cut-off for
106 the MABC taxonomic calssication. Additionally, to our knowledge, no specific genes
107 were identified to discriminate *M. abscessus subsp. bolletii* (The traditional *rpoB*
108 gene is not good enough for subspecies identification). To our knowledge,
109 WGS-based protocols have so far not been developed for subspecies level like
110 MABC identification. Therefore, one of the purposes of this study was to identify
111 specific gene for *M. abscessus subsp. bolletii* identification by the combination of
112 several genes, and to build a powerful and reliable taxonomical tool for the MABC
113 based on minimum and maximum identity values.
114

115 MATERIALS AND METHODS

116 **Bacterial isolates** A total of 6 MABC isolates were obtained from different
117 patients from the same hospital at the same time period (2009). Of them, 2 strains
118 were isolated from sputum samples and 4 were obtained from Bronchoalveolar
119 Lavage Fluid (BALF). All of the isolates had been classified as *M. abscessus* based
120 on the results of biochemistry followed the Leao's method[20].

121 **DNA Extraction** Bacterial DNA extraction was performed as described
122 previously[21]. In brief, harvested the bacterium after growing in Middlebrook 7H9
123 liquid medium for 5 days, and then the samples were crushed with zirconia beads (1

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

124 mm in diameter) in a tissue disintegrator instrument. Total genomic DNA was
125 extracted from the crushed suspension using a commercial ethanol precipitation kit
126 according to the manufacturer's instructions and stored at -20°C. Samples were
127 sequenced on the Illumina HiSeq X10 sequencer, using 150 base-pair paired-end
128 reads.

129 **Genome assemble** After checking the length and quality of the reads with
130 FastQC (Version 0.11.8)[22]. The reads were de novo assembled with
131 SPAdes(v3.13.0), using the '-careful' setting and k-mers 21, 33, 55, 77, and 99[23,
132 24], and the assembled genome was manually trimmed the Short contigs(length than
133 3,000bp) and very low coverage contigs(less than 10).The result of the assembly was
134 evaluated using QUAST (Version v.5.0.2, http://quast.bioinf.spbau.ru/;)[25].

135 **Vitro drug susceptibility testing** Using the microdilution method following
136 the recommendations of the Clinical and Laboratory Standards Institute [Clinical
137 and Laboratory Standards Institute (CLSI) of the rapidly growing mycobacteria with
138 microdilution method][26]. The susceptibility results to ciprofloxacin, moxifloxacin
139 and clarithromycin were judged by the established breakpoints from CLSI document
140 (M24-A2-2011).

141 To evaluate the traditional identification taxonomy method, 16S rRNA
142 (1,468-bp), *rpoB* (409-bp), *erm*(41)(GenBank accession number:CU458896.1),
143 *erm*(42)(GenBank accession number: FJ358487.1) were selected for comparison[27, 28].
144 Nucleotide sequence were extracted from the reference sequences *M. abscessus*
145 *subsp. abscessus* ATCC 19977 and *Mycobacterium_abscessus_subsp_bolletii* CIP
146 108541 using the traditional primers(The primer used in this study was listed in
147 supplemental table 1). For comparision, the *erm*(41) fragment was selected with the
148 same beginning and end with *erm*(42) fragment. The identities among the fragments
149 were got by BLASTN searches with 1E-5 as a cut-off value [29]. And the fragments
150 from the genomes were gotten by TBtools[30].

151 In order to test method got in this study, all of the genomes (complete,
152 chromosome, scaffold) were downloaded from NCBI genome
153 (https://www.ncbi.nlm.nih.gov/genome/genomes/1360?). In order to evaluate the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

154 subspecies of MABC, average nucleotide identity (ANI), average amino acid
155 identity (AAI) and genome to genome distance (GGDC) were calculated among the
156 typical strains using KostasLab two-way average AAI calculator
157 (http://enve-omics.ce.gatech.edu/aai/)[31], ANI
158 (http://enve-omics.ce.gatech.edu/)[32], OrthoANI calculator[33],
159 Genome-to-Genome Distance Calculator (http://ggdc.dsmz.de/ggdc.php)[34],
160 fastANI[40] was also used to estimate the ANI among the genome datasets
161 constructed in this study. In order to get the most effective gene to distinguish
162 *M.bolleti* from other MABC, Up to date Bacterial Core Gene (UBCG) tool was
163 chose to select gene from the 92 bacterial core genes[35]. The Phylogenetic tree was
164 visualized using iTOL[36].

165 # Result

166 Patients were primarily male (4/6 patients) and greater than 65 years of age
167 (4/6patients). Antimicrobial resistance patterns of *M. abscessus subsp. abscessus* and
168 *M. abscessus subsp. massiliense* isolated from respiratory specimens were shown in
169 supplemental table 2. The result confirmed the innate resistance of MABC to several
170 drug classes, such as the cefoxitin. Fortunately, those isolates are still sensitive to
171 clarithromycin in 3 days. And the isolate 1 was the only one with observed resistance
172 to clarithromycin at 7 days. Genomes analyze result also showed that no known
173 resistance-conferring mutation in the 23S rRNA was observed in those isolates.

174 Then study first verified that whether MABC belonged to the same specie or
175 different species. The AAI, ANI, GGDC results of the typical isolates of the MABC
176 were shown in table 1. Among the representative genome dataset as previously
177 reported (unpublished result), using the full length *ropB* obtained from *M. abscessus*
178 *subsp. abscessus* ATCC 19977 as reference, the similarities among MABC complex
179 represent isolates were from 97.78%-100% (table 1 and figure 1). All the results
180 showed MABC complex belonged to the same specie.

181 Then this study tested whether WGS comparison could further distinguish
182 subspecies among the MABC complex. The ANI values for *Mycobacterium*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

183 *abscessus subsp massiliense* were from 96.72%-100% (Supplem table 6). And the
184 ANI for *Mycobacterium abscessus subsp bolletii* were from 96.1263%-100%
185 (Supplemental table 7). Because the overlap value of the ANI, the ANI value could
186 not be used for subspecies distinguishing for MABC complex. The supplemental
187 figure 1 also shows although the based on the ANI value could distinguish the
188 species level but could not distinguish the sub species that are within the same
189 species. Then another WGS based method-UBCG was used for subspecies
190 verification. In accordance with our findings, based on the phylogenomic analyses of
191 the UBCG (92 genes) of isolates from the public genome database showed the clear
192 distance separating the three subsp of *M.abscessus* complex (figure 2 A). The
193 genome which had been labeled the subsp names were with logo in figure. From the
194 figure 2 A, we can see although many genomes were not labeled with supcies in the
195 NCBI dataset, all the genomes with clear subspecies could be separate from other
196 subspecies by different branches. So the UBCG is a good standard for subspecies
197 verification. But using UBCG method for the subspecies classification took long
198 time and with ambiguous boundaries (there is no clear cut-off value for UBCG
199 method). So this study further tested the representative genes for species
200 identification.

201 Firstly, in order to get the most accurate gene for identifying the MABC
202 complex from other Mycobacterium spp, this study first evaluated the current used
203 the gene fragment (Supplemental table 1). When we test the traditional marker for
204 MABC identification, the *hsp* 65 and 16S–23S ITS primers could not be found in the
205 genomes for the representative strains (Supplemental table 1). So this study only
206 compared the 16s and *rpoB* genes. Then this study found that when using the
207 fragment gotten by the primers rpoB-mycoF and rpoB-mycoR, the identity between
208 the fragments from *M. abscessus subsp. abscessus* ATCC 19977 and *M. abscessus*
209 *subsp. bolletii* CIP 108541 was 96.81. As the identity was below 97% which was
210 settled by the traditional cut-off value, this study then used the full length rpoB genes
211 for comparison. The similarities of the 16s and *rpoB* genes were shown in Figure 1.
212 As shown in Figure 1, compared with 16s identity the rpoB genes identity could

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

213 better distinguish the MABC with other Mycobcacterium spp. While 16s identities
214 could not distinguish the MABC complex with similar species such as
215 *Mycobacterium_salmoniphilum*, *Mycobacterium_immunogenum*,
216 *Mycobacterium_stephanolepidis*, *Mycobacterium_franklinii*, *Mycobacterium_chelonae*,
217 *Mycobacterium_saopaulense*. To verified whether the cut-off value could suit for all the
218 MABC complex isolates, the genome dataset download from NCBI was used for
219 testing. As the Supplemental_table 4 shows the *rpoB* identities were among
220 97.697%-100%. (893 isolates from 910 totals isolates, 7 of the genomes in the public
221 databases were labeled contaminated were excluded from the dataset. Another 10
222 isolates in the genome database were shown in supplemental Supplemental_table 4,
223 and all genome based identify method showed that they were very far from MABC
224 complex). So the full length of *rpoB* gene and the cut-off value 97% could be used
225 for MABC complex identification. Then this study tested whether *rpoB* could
226 distinguish the subspecies. This study first utilized dataset of *M. abscessus subsp.*
227 *Bolletii* from NCBI to test whether the full length *rpoB* gene could be used as
228 subspecies marker. But when using the *rpoB* gene from *M. abscessus subsp.*
229 *abscessus* ATCC 19977, the identities are ranging from 97. 697−97. 896%, and using the *rpoB*
230 gene from *M. abscessus subsp. massiliense* the identities are from 98.052-99.742%. The
231 *M. abscessus subsp. bolletii* CIP 108541 identities are ranged from 97.811 to 100%. As the
232 ranges are overlapping with each other, using different rpoB genes could not distinguish other.
233 After using *rpoB* for MABC species identification, this study then verified the
234 current used method for subspecies identificaiton. This study first selected the
235 erythromycin ribosomal methylase gene, *erm*(41) for the further distinguishing the
236 subspecies. The identities of *erm*(41) and *erm*(42) were shown in supplemental table
237 8. When using *erm*(41) fragment from *Mycobacterium abscessus subsp abscessu*
238 ATCC 19977 as reference, the identities were from 97.893%-100% (586/893); and
239 *erm*(42) fragment *Mycobacterium abscessus sub sp massiliense* (GenBank accession
240 number: FJ358487.1) were 99.58%-100% (307/893). Because there were no genomes
241 could share the identities with both *erm*(41) and *erm*(42), this study used identity
242 99.58% for *erm*(42) as the cut-off value for selecting *Mycobacterium abscessus sub*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

243 *sp massiliense* from the MABC complex. Then using already existed labeled
244 *Mycobacterium abscessus subsp* tested the classification, and the results were shown
245 in figure 2 B. From the figure 2 B we can see there were no cross between *erm*(41)
246 and *erm*(42) (The genomes which contained both *erm*(41) and *erm*(42) are located in
247 different branches). The figure showed that the *erm*(42) were exactly matched the
248 *Mycobacterium abscessus subsp massiliense.* So does the *erm*(41) with
249 *Mycobacterium abscessus subsp bolletii* and the isolates without special labeled. So
250 the fragment of *erm*(42) with the cut-off value 99.58% was used for the
251 identification of *Mycobacterium abscessus sub sp massiliense.*

252 In order to further distinguish the *Mycobacterium abscessus subsp abscessus*
253 and *M. abscessus subsp. Bolletii,* this study first tested the traditional marker the
254 fragment of *rpoB*. According to the traditional method *M. abscessus subsp. Bolletii*
255 genomes should be group together according to identities of *rpoB* fragment. This
256 study tested the identies of *rpoB* fragment from the *M. abscessus subsp. Bolletii* with
257 the *rpoB* fragment from all the three species of MABC complex. The *rpoB* identities
258 among the *M. abscessus subsp. Bolletii were* from 95.88% to 99.73 %, when using
259 the fragment from *Mycobacterium abscessus subsp massiliense* GO 06. The identities
260 were from 94.16% to 100%, when using the fragment from *Mycobacterium abscessus*
261 *subsp abscessu* ATCC 19977, and the identies were from 95.88% to 100% when
262 using the fragement from *M. abscessus subsp. bolletii* CIP 108541 (Supplemental table
263 9). As the *ropB* gene could not distinguish the *M. abscessus subsp. Bolletii,* this
264 study tried to find a marker gene for *M. abscessus subsp. Bolletii* from the UBCG
265 selected genes. All the 92 genes tested in UBCG were evaluated in this study. From
266 all those individual gene trees, when using *rpls* gene for drawing the Phylogenetic
267 tree, fewest genomes were involved with the genomes which had been labeled with
268 *Mycobacterium abscessus subsp bolletii*(Figure 3 A). Then in order to get the cut-off
269 value for the *rpls* gene for the *Mycobacterium abscessus subsp bolletii* identification,
270 this study first used full length *rpls* gene identities(Supplemental table 10 A).
271 Because two genomes' identities were little far from other genomes, this study
272 selected partial common *rpls* gene frgement share by *Mycobacterium abscessus*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

273 *subsp bolletii* for reference (Supplemental file 11). Although *Mycobacteroides*
274 *abscessus subsp. bolletii* 50594 was labeled as *Mycobacteroides abscessus subsp.*
275 *bolletii,* it carried *erm*(42) gene instead of *erm*(41). What is more this strain was also
276 labeled as *Mycobacterium massiliense* 50594 (heterotypic synonym). So this study
277 excluded this genome from *Mycobacteroides abscessus subsp. bolletii* dataset, and
278 the cut-off value for *rpls* fragment identity was 99.53% (Supplemental table 10 B).
279 Then study used the fragement and the cut-off value to test all the MABC complex
280 genomes from the NCBI. The genomes with the identity of *rpls* fragement > 99.53%
281 and identity of *erm*(42) fragement < 99.58% were exactly the branch of the
282 *Mycobacteroides abscessus subsp. bolletii* (Figure 2 C, A and Supplemental table 10
283 C). So *rpls* gene fragment and the cut-off value 99.53% was selected for the
284 *Mycobacterium abscessus subsp bolletii* identification.

285 The clinical isolates involved in this study were submitted to NCBI with the Accession
286 number PRJNA594106. Last but not least, this study summary the analyzing data (Figure 4)
287 and then contrsusted the software NucleotideQuery for the MABC complex classification
288 (Attachment file 12) which is accurate and user-friendly (No need for installation, and could
289 detect the genomes directly)..

290 # Discussion

291 All of the three *M.abscessus* subsp had been reported could cause serious
292 infections such as pulmonary [37], and the controversies about MABC's taxonomic
293 status have never stopped. The proposed and generally accepted species boundary
294 for AAI, ANI and dDDH values are 95%, 95~96 and 70 %, respectively [34, 38, 39].
295 As the result showed in Table 1, the AAI values among the MABC typical strains
296 were 97.61%-97.95%, and all larger than the cut-off value 95%. The ANI values
297 among the MABC typical strains were 96.94%-97.39% (the results gotten by
298 different softwares were similar as the Table 1 shown); and they were all larger than
299 cut-off value 96%. the GGDC value among the MABC typical strains were
300 73.4%-77.2%. The above results of this study supported the current classification of
301 *M. abscessus subsp. abscessus, M. abscessus subsp. massiliense* and *M. abscessus*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

302 *subsp as sub species* are belonged to the same specie.

303 The most common method to identify the MABC complex isolates were based
304 on the *rpoB*, but this method had been questioned[40]. Andt he cut-off value of *rpoB*
305 gene for MABC complex and even for the subsp in MABC complex had been
306 studied for a long time[28, 41-43]. To our knowledge this study first time using the
307 full length *rpoB* gene as the marker gene. What is more this study involved the
308 largest number of MABC complex (all the available genomes in the NCBI genome).
309 With the cut-off value of this study of *rpoB* to identifythe MABC complex was
310 97.697% that could distinguish the MABC complex with other *Mycoabcterium* spp.
311 As showed in supplemental table 5, the three subspecies in MABC could not be
312 distinguished with each other by *rpoB* only. This study then further evaluated the
313 current used marker, and selected the 97.893 and 99.5588 for *erm*(41) and *erm*(42)
314 fragement, respectively, which would distinguish *M.abscessus subsp. Massiliense*
315 with *M.abscessus subsp.abscessus*. To our knowledge, this study is the first time to
316 set the clearly cut-off value of the *erm* gene during the distinction of MABC
317 complex. The Figure 2 showed that these cut-off values were suitable for
318 *Mycobacterium abscessus subsp massiliense* identification.

319 Due to the clinical significance, all of the three subsp should be differentiated
320 with each other rapidly. Differed with *M.abscesusssubsp.massiliense, M.abscessus*
321 *subsp.bolletii* harbored a inducible functional erythromycin ribosome
322 methyltransferase *erm*(41)[17]. Thus, this study further designed new target genes to
323 distinguish *M.abscessus subsp.bolletii* from *M.abscessus subsp.abscessus*.
324 Scientists had noticed the importance of identification of the subsp of MABC
325 complex for a long time, but the previous reports for taxonomy of *Mycobacterium*
326 *abscessus subsp massiliense* were ambiguous, especially for the cut-off value[44].
327 The same situation also occurs for the *Mycobacterium abscessus subsp bolletii,*
328 previous reports used *rpoB*, *hsp65* or 16S–23S ITS for identification [44-48], and no
329 cut-off value was set for each gene. Furthermore, the previous reported primers
330 *hsp*65 and 16S–23S ITS[37]could not be found in the genomes for the two
331 representative strains. So this study did not involved these genes for testing. In

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

332 contast with the tradiotinal marker – the fragement of *rpoB* frangment[49, 50], this
333 study also had showed that the *rpoB* gene could not distinguish *Mycobacterium*
334 *abscessus subsp massiliense* with *Mycobacterium abscessus subsp bolletii*as shown
335 in the attachment table 9. When using *rpoB* fragment from *M.abscessu.subsp.bolletii*
336 as reference, the identities among the labeled *M.abscessu.subsp.bolletii* genomes
337 were from 95.878% to 100%; when using *rpoB* fragment from
338 *M.abscessu.subsp.massiliense* as reference, the identity among the labelled
339 *M.abscessu.subsp.bolletii* genomes were from 95.878%-99.734%. As the ranges
340 were over cross with each other, so this study believed the *rpoB* fragments could not
341 be used for *M.abscessu.subsp.bolletii* identification. Previous study also confirmed
342 that single gene was not good enough for subsp taxonomy[51]. Other methods like
343 the pulsed-field gel electrophoresis are complex and costly that are not suitable for
344 clinical field[52]. In order to find out the most effective gene for distinguishing
345 *M.abscessus.subsp. bolletii*, all 92 genes selected by the UBCG were evaluated by
346 the number of the smallest group including *M.abscessus.subsp. bolletii*. And then
347 this study got the cut-off value by the *M.abscessus.subsp. bolletii.* After verified the
348 selected gene *rpls* and cut-off value using the MABC complex dataset, this study
349 demonstrated the *rpls* fragement with the identity cut-off value 99.53% could be
350 served as the unique *M.abscessus.subsp. bolletii* identication marker.

351 The ANI has been validated as prokaryotic species taxonomy tool, where ANI
352 values higher than 95–96% are consistent with strains belonging to the same species.
353 To our knowledge, no cut-off has been proposed yet to define the boundary of
354 subspecies. Although most of the *M.abscessus.subsp.abscessus* genomes were not
355 labeled in the genome database, the ANI value could not distinguish
356 *M.abscessus.subsp.bolletii* with *M.abscessus.subsp. massiliense*(attachment table 6
357 and 7). This study also showed that the ANI value was not suitable to distinguish
358 the sub species of MABC complex. But when using WGS assemble and then
359 analyzed by our software; this rapid approach offers the accurate, user-friendly
360 MABC complex subspecies identification using *rpoB, erm(42)* and *rpls* are
361 sufficiently reliable to serve as the routine methodology in clinical field. So, our

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

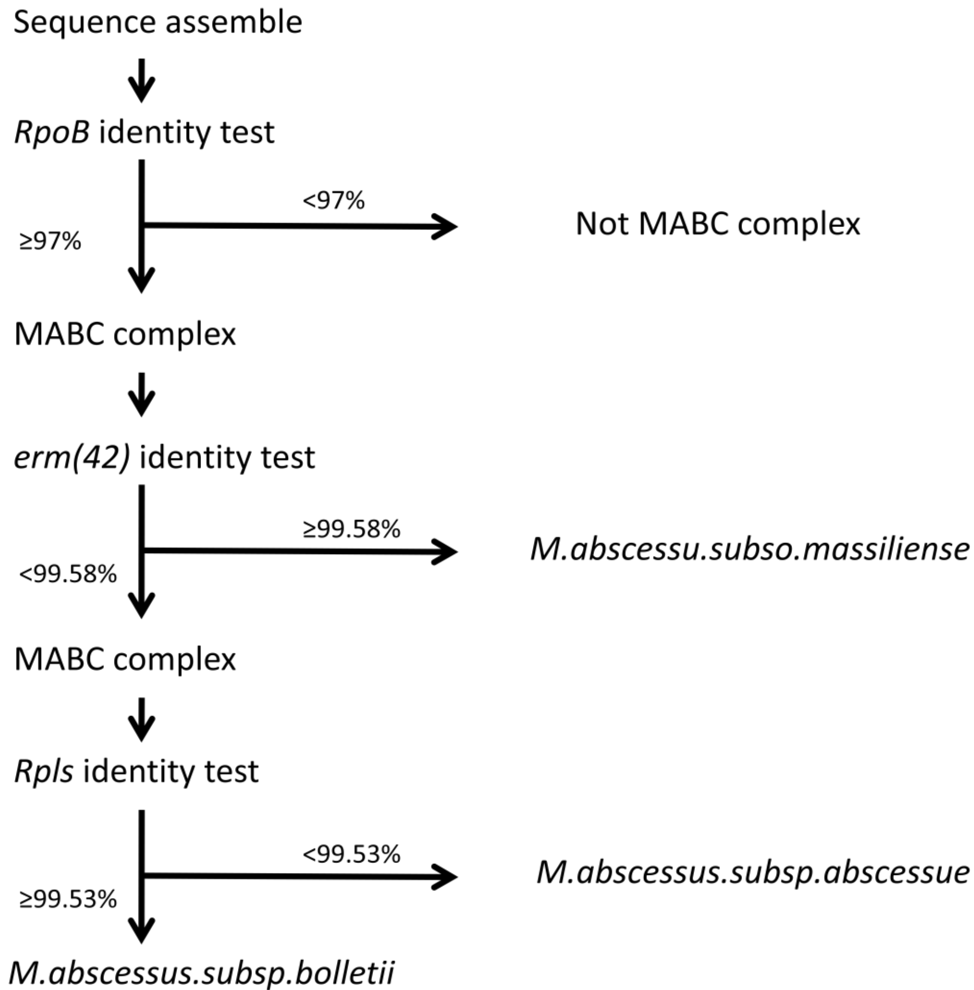
362 method was with high clinical meaning.

363 Conclusion

364 This study reported genome sequences and genomic features of 6 MABC
365 isolates derived from one hospital in China. The results of this study showed that the
366 distance among MABC complex was insufficient to warrant distinction at the species
367 level, so there are subspecies in MABC complex. And after verified the taxonomy of
368 MABC complex, we had developed an user-friendly accurate method based on a set
369 of genes for differentiation at the subspecies level *M.abscessus subsp. massiliense,*
370 *M.abscessu subsp.bolletti* and *M.abscessus subsp.abscess*. Using all available
371 sequences to test the method, and the sequence analysis of public databases indicated
372 that the combination by this design could be high discrimination among MABC
373 complex.

374 Whole-genome sequencing technology has been more and more widely used in
375 clinical diagnosis, developing more user-friendly software would greatly facilitate
376 the acquisition of more-precise information about pathogen, to aid in the choice of
377 more discriminate therapies.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



378

379 Figure 4 Flow chart of the study software

380

381

382 ACKNOWLEDGMENT

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

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

387

388 CONFLICTS OF INTEREST

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

390

391

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

392 1. RM, D., et al., *Genome sequencing of Mycobacterium abscessus isolates from patients in the*
393 *united states and comparisons to globally diverse clinical strains.* Journal of Clinical
394 Microbiology, 2014. **52**(10): p. 3573-82.
395 2. KK, C., et al., *First United States Outbreak of Mycobacterium abscessus Hand and Foot*
396 *Disease Among Children Associated With a Wading Pool.* Journal of the Pediatric Infectious
397 Diseases Society, 2018. **undefined**(undefined): p. undefined.
398 3. Daniel-Wayman, S., et al., *Antibiotic Susceptibility of Mycobacterium Abscessus Patients*
399 *with Long Term Antibiotic Treatment*, in *A106. NTM CLINICAL AND TRANSLATIONAL*
400 *SCIENCE*2018, American Thoracic Society. p. A2593-A2593.
401 4. R, N., et al., *Mycobacterium abscessus: a new antibiotic nightmare.* The Journal of
402 antimicrobial chemotherapy, 2012. **67**(4): p. 810-8.
403 5. Stout, J.E., W.J. Koh, and W.W. Yew, *Update on pulmonary disease due to non-tuberculous*
404 *mycobacteria.* International Journal of Infectious Diseases, 2016. **45**(C): p. 123-134.
405 6. Kwon, Y.S. and W.J. Koh, *Diagnosis and Treatment of Nontuberculous Mycobacterial Lung*
406 *Disease.* Journal of Korean Medical Science, 2016. **31**(5): p. 649-659.
407 7. A, K., et al., *Activity of the New β-Lactamase Inhibitors Relebactam and Vaborbactam in*
408 *Combination with β-Lactams against Mycobacterium abscessus Complex Clinical Isolates.*
409 Antimicrob Agents Chemother, 2019. **63**(3): p. undefined.
410 8. Bryant, J.M., et al., *Emergence and spread of a human-transmissible multidrug-resistant*
411 *nontuberculous mycobacterium.* Science, 2016. **354**(6313): p. 751-757.
412 9. Boeck, L., et al., *Late Breaking Abstract-Multidimensional genomic mapping reveals the*
413 *pathophysiology of Mycobacterium abscessus*, 2018, Eur Respiratory Soc.
414 10. Meng-Rui, L., et al., *Mycobacterium abscessus Complex Infections in Humans.* Emerging
415 infectious diseases, 2015. **21**(9): p. 1638-1646.
416 11. Toshiyuki, H., et al., *Clinical and microbiological differences between Mycobacterium*
417 *abscessus and Mycobacterium massiliense lung diseases.* Journal of Clinical Microbiology,
418 2012. **50**(11): p. 3556-61.
419 12. Won-Jung, K., et al., *Clinical significance of differentiation of Mycobacterium massiliense*
420 *from Mycobacterium abscessus.* American Journal of Respiratory & Critical Care Medicine,
421 2011. **183**(3): p. 405.
422 13. Lipworth, S., et al., *Whole-Genome Sequencing for Predicting Clarithromycin Resistance in*
423 *Mycobacterium abscessus.* Antimicrob Agents Chemother, 2019. **63**(1): p. e01204-18.
424 14. DE, G., *Mycobacterium abscessus subsp abscessus lung disease: 'trouble ahead, trouble*
425 *behind…'.* F1000prime reports, 2014. **6**(undefined): p. 107.
426 15. K, N., et al., *Discrimination of Mycobacterium abscessus subsp. massiliense from*
427 *Mycobacterium abscessus subsp. abscessus in clinical isolates by multiplex PCR.* Journal of
428 Clinical Microbiology, 2014. **52**(1): p. 251-9.
429 16. E, T., et al., *Emended description of Mycobacterium abscessus, Mycobacterium abscessus*
430 *subsp. abscessus and Mycobacteriumabscessus subsp. bolletii and designation of*
431 *Mycobacteriumabscessus subsp. massiliense comb. nov.* International journal of systematic
432 and evolutionary microbiology, 2016. **66**(11): p. 4471-4479.
433 17. Lee, M.-R., et al., *Mycobacterium abscessus complex infections in humans.* Emerging
434 infectious diseases, 2015. **21**(9): p. 1638.
435 18. Adekambi, T., et al., *Reinstating Mycobacterium massiliense and Mycobacterium bolletii as*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

436 *species of the Mycobacterium abscessus complex.* International journal of systematic and
437 evolutionary microbiology, 2017. **67**(8): p. 2726-2730.
438 19. M, B., et al., *Careful use of 16S rRNA gene sequence similarity values for the identification of*
439 *species.* New microbes and new infections, 2018. **22**(undefined): p. 24-29.
440 20. Leao, S.C., et al., *Practical handbook for the phenotypic and genotypic identification of*
441 *mycobacteria.* Vanden Broele, Brugges, Belgium, 2004.
442 21. Votintseva, A.A., et al., *Mycobacterial DNA Extraction for Whole-Genome Sequencing from*
443 *Early Positive Liquid (MGIT) Cultures.* Journal of Clinical Microbiology, 2015. **53**(4): p.
444 1137-43.
445 22. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*, 2010,
446 Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
447 23. S, N., et al., *Assembling single-cell genomes and mini-metagenomes from chimeric MDA*
448 *products.* Journal of computational biology : a journal of computational molecular cell
449 biology, 2013. **20**(10): p. 714-37.
450 24. A, B., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell*
451 *sequencing.* Journal of computational biology : a journal of computational molecular cell
452 biology, 2012. **19**(5): p. 455-77.
453 25. Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies.* Bioinformatics,
454 2013. **29**(8): p. 1072-1075.
455 26. Woods, G., et al., *Susceptibility testing of mycobacteria, nocardiae, and other aerobic*
456 *actinomycetes*2011: Clinical and Laboratory Standards Institute.
457 27. Simmon, K.E., et al., *Simultaneous sequence analysis of the 16S rRNA and rpoB genes by use*
458 *of RipSeq software to identify Mycobacterium species.* Journal of Clinical Microbiology, 2010.
459 **48**(9): p. 3231-3235.
460 28. T, A., C. P, and D. M, *rpoB-based identification of nonpigmented and late-pigmenting rapidly*
461 *growing mycobacteria.* Journal of Clinical Microbiology, 2003. **41**(12): p. 5699-708.
462 29. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**(1):
463 p. 421.
464 30. Chen, C., et al., *TBtools, a Toolkit for Biologists integrating various biological data handling*
465 *tools with a user-friendly interface.* BioRxiv, 2018: p. 289660.
466 31. Rodriguez-R, L. and K. Konstantinidis, *Bypassing cultivation to identify bacterial species.*
467 *Microbe 9: 111–118*, 2014.
468 32. Rodriguez-R, L.M. and K.T. Konstantinidis, *The enveomics collection: a toolbox for specialized*
469 *analyses of microbial genomes and metagenomes*, 2016, PeerJ Preprints.
470 33. I, L., et al., *OrthoANI: An improved algorithm and software for calculating average*
471 *nucleotide identity.* International journal of systematic and evolutionary microbiology, 2016.
472 **66**(2): p. 1100-1103.
473 34. JP, M.-K., et al., *Genome sequence-based species delimitation with confidence intervals and*
474 *improved distance functions.* BMC Bioinformatics, 2013. **14**(undefined): p. 60.
475 35. Na, S.I., et al., *UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree*
476 *reconstruction.* Journal of Microbiology, 2018. **56**(4): p. 1-6.
477 36. Letunic, I. and P. Bork, *Interactive tree of life (iTOL) v3: an online tool for the display and*
478 *annotation of phylogenetic and other trees.* Nucleic Acids Research, 2016. **44**(Web Server
479 issue): p. gkw290.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.02.893412; this version posted January 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

480 37. Kazue, N., et al., *Multiple cases of cutaneous Mycobacterium massiliense infection in a "hot*
481 *spa" in Japan.* Journal of Clinical Microbiology, 2011. **49**(2): p. 613.
482 38. Goris, J., et al., *DNA-DNA hybridization values and their relationship to whole-genome*
483 *sequence similarities.* Int J Syst Evol Microbiol, 2007. **57**(1): p. 81-91.
484 39. KT, K., R.-M. R, and A. R, *Uncultivated microbes in need of their own taxonomy.* The ISME
485 journal, 2017. **11**(11): p. 2399-2406.
486 40. M, S. and D. M, *Genome analysis reveals three genomospecies in Mycobacterium abscessus.*
487 BMC genomics, 2014. **15**(undefined): p. 359.
488 41. R, d.Z., v.I. J, and v.S. D, *Utility of rpoB gene sequencing for identification of nontuberculous*
489 *mycobacteria in the Netherlands.* Journal of Clinical Microbiology, 2014. **52**(7): p. 2544-51.
490 42. A, M., et al., *Assessment of partial sequencing of the 65-kilodalton heat shock protein gene*
491 *(hsp65) for routine identification of Mycobacterium species isolated from clinical sources.*
492 Journal of Clinical Microbiology, 2004. **42**(7): p. 3000-11.
493 43. Edouard, M., et al., *Multilocus sequence analysis and rpoB sequencing of Mycobacterium*
494 *abscessus (sensu lato) strains.* Journal of Clinical Microbiology, 2011. **49**(2): p. 491.
495 44. S, Y., et al., *Further isolation of Mycobacterium abscessus subsp. abscessus and subsp.*
496 *bolletii in different regions of Japan and susceptibility of these isolates to antimicrobial*
497 *agents.* International journal of antimicrobial agents, 2013. **42**(3): p. 226-31.
498 45. C, V.-N., et al., *Molecular characterization of Mycobacterium massiliense and*
499 *Mycobacterium bolletii in isolates collected from outbreaks of infections after laparoscopic*
500 *surgeries and cosmetic procedures.* Journal of Clinical Microbiology, 2008. **46**(3): p. 850-5.
501 46. Duarte, R.S., et al., *Epidemic of postsurgical infections caused by Mycobacterium massiliense.*
502 Journal of Clinical Microbiology, 2009. **47**(7): p. 2149-2155.
503 47. Nie, W., et al., *Species identification of Mycobacterium abscessus subsp. abscessus and*
504 *Mycobacterium abscessus subsp. bolletii using rpoB and hsp65, and susceptibility testing to*
505 *eight antibiotics.* International Journal of Infectious Diseases, 2014. **25**: p. 170-174.
506 48. Lee, M.-R., et al., *Bacteraemia caused by Mycobacterium abscessus subsp. abscessus and M.*
507 *abscessus subsp. bolletii: clinical features and susceptibilities of the isolates.* International
508 journal of antimicrobial agents, 2014. **43**(5): p. 438-441.
509 49. Adékambi, T. and M. Drancourt, *Mycobacterium bolletii respiratory infections.* Emerging
510 infectious diseases, 2009. **15**(2): p. 302.
511 50. Rollet-Cohen, V., et al., *Mycobacterium bolletii Lung Disease in Cystic Fibrosis.* Chest, 2019.
512 51. E, M., et al., *Inaccuracy of single-target sequencing for discriminating species of the*
513 *Mycobacterium abscessus group.* Journal of Clinical Microbiology, 2009. **47**(8): p. 2596-600.
514 52. Zelazny, A.M., et al., *Cohort study of molecular identification and typing of Mycobacterium*
515 *abscessus, Mycobacterium massiliense, and Mycobacterium bolletii.* Journal of Clinical
516 Microbiology, 2009. **47**(7): p. 1985-1995.
517
518
519
520
521
522
523

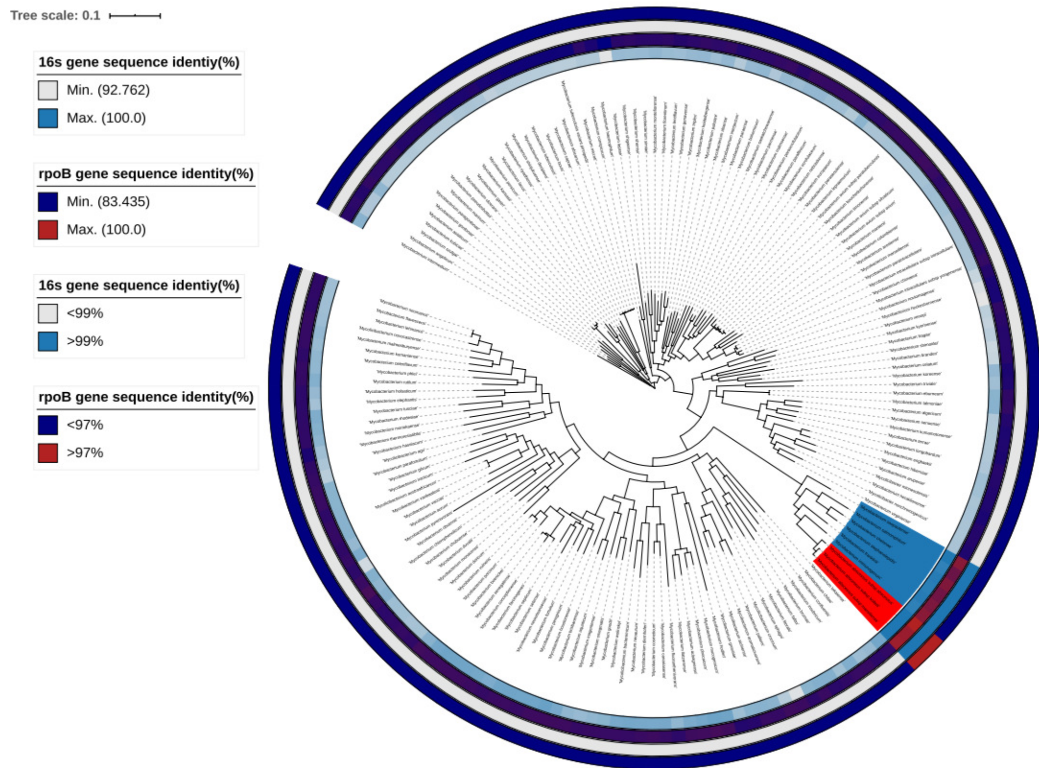
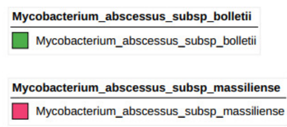
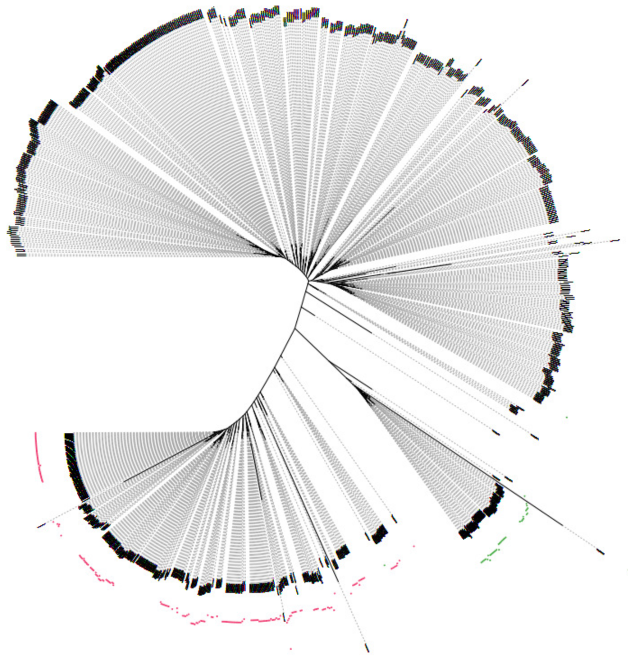


Figure 1: Phylogenetic relatedness of *Mycobacterium* spp and the identities of the 16s and *rpoB* genes

The inner range 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 identities. The grey and light blue range shows the identities of 16s rRNA and the the dark blue and red ranges show the identity of *rpoB*. The outside two ranges show the identities of 16s and rpoB genes with the cut-off values.

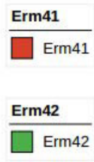
Tree scale: 0.01

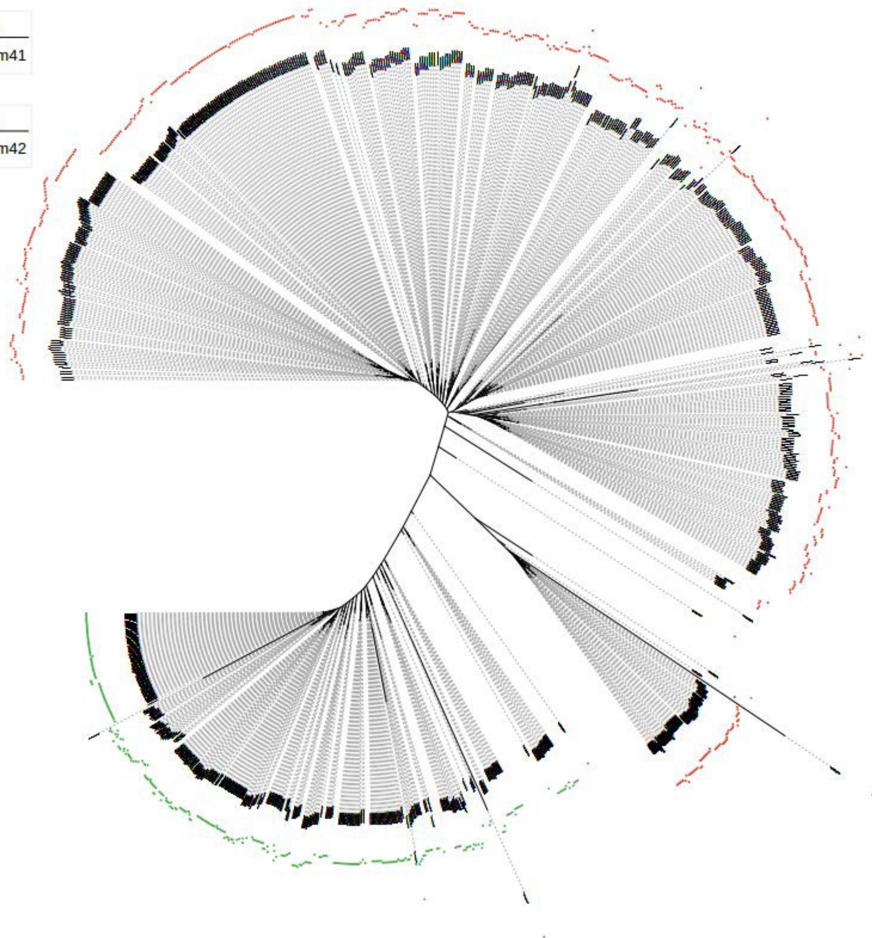




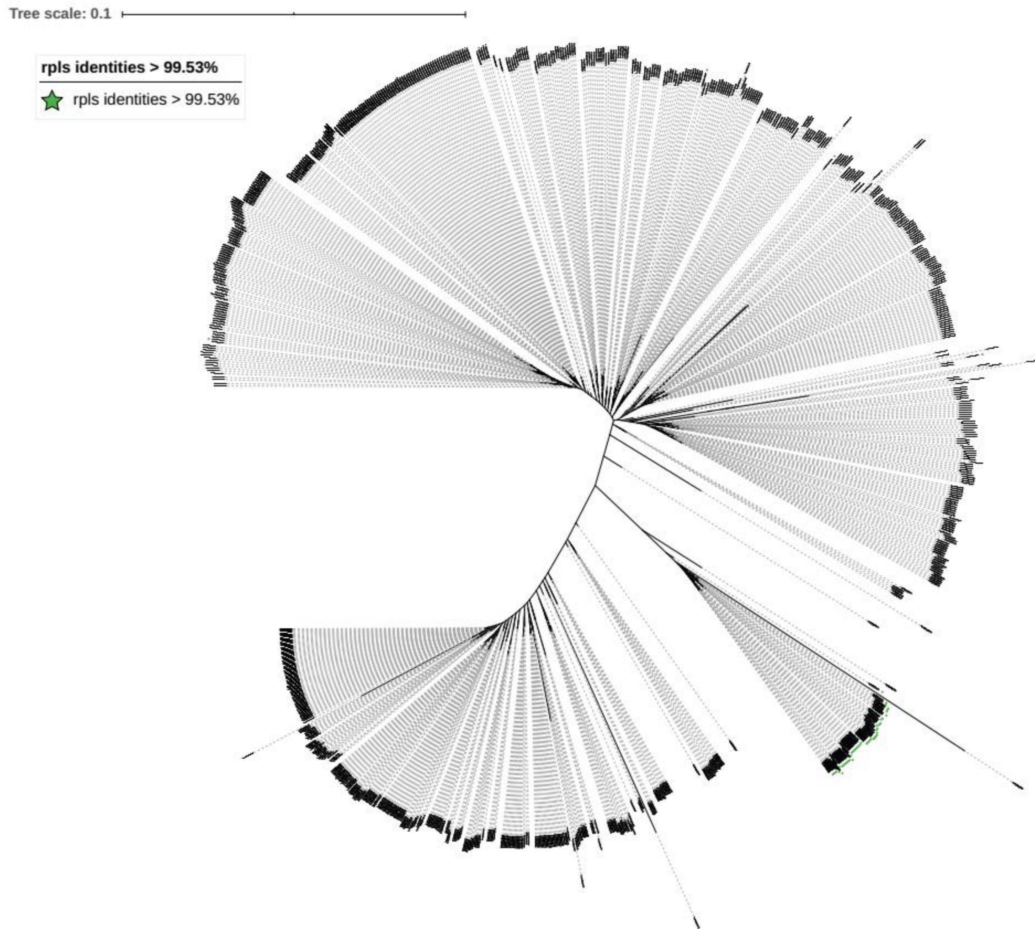
A

Tree scale: 0.01





B



C

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 MABC complex from the NCBI with all the clear label of *M. abscessus subsp. massiliense* and *M. abscessus subsp. Bolletii.* B Phylogenetic trees for MABC complex from the NCBI with all the clear label of the identities of *erm*(41) and *erm*(42). C Phylogenetic trees for MABC complex from the NCBI with *rpls* gene identities > 99.53% and *erm*(42) gene identities < 99.58% .

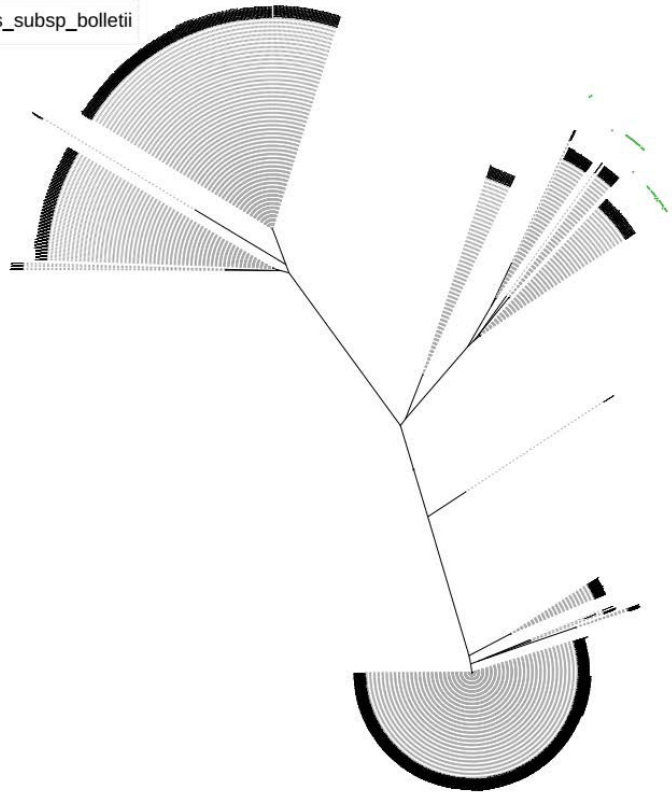


Figure 3 The phylogenomic tree for identification of *Mycobacterium abscessus subsp bolletii* A phylogenomic tree based on *rpls*. The green square labeled the *Mycobacterium abscessus subsp bolletii* genomes with clear labeling as *Mycobacterium abscessus subsp bolletii* in NCBI database.

Table 1 ANI, GGDC, DDH, *rpoB* and AAI genomic relatedness among the type strains *M. abscessus subsp. abscessus* ATCC 19977, *M. abscessus subsp. massiliense* GO 06 and *M. abscessus subsp. bolletii* CIP 108541

A: ANI, GGDC, DDH and *rpoB* similarity values among the MABC

Genome 1	Genome 2	OrthoANI value (%)	Orginal ANI value (%)	GGDC distance	Formula 2 Distance	Formula 2 DDH estimate	*rpoB* similarity
Mycobacterium_abscessus_subsp_abscessus	*Mycobacterium_abscessus_subsp_bolletii*	97.3906	97.1484	0.026672229	0.0267	77.20%	97.78
Mycobacterium_abscessus_subsp_abscessus	*Mycobacterium_abscessus_subsp_massiliense*	97.3751	97.2282	0.027495489	0.0275	76.60%	98.067
Mycobacterium_abscessus_subsp_bolletii	*Mycobacterium_abscessus_subsp_massiliense*	96.94	96.7821	0.031488937	0.0315	73.40%	

Genome 1	Genome 2	OrthoANI value (%)	Orginal ANI value (%)	GGDC distance	Formula 2 Distance	Formula 2 DDH estimate
Mycobacterium_abscessus_subsp_abscessus	Mycobacterium_abscessus_subsp_bolletii	97.3906	97.1484	0.026672229	0.0267	77.20%
Mycobacterium_abscessus_subsp_abscessus	Mycobacterium_abscessus_subsp_massiliense	97.3751	97.2282	0.027495489	0.0275	76.60%
Mycobacterium_abscessus_subsp_bolletii	Mycobacterium_abscessus_subsp_massiliense	96.94	96.7821	0.031488937	0.0315	73.40%

The ANI values were gotten by OrthoANI calculator[32], and the GGDC and DDH values were gotten from http://ggdc.dsmz.de/ggdc.php#.

The *rpoB* similarity by pairwise alignment of sequences based on BLAST

B: AAI value values among the MABC

SeqA	SeqB	AAI	SD	N	Omega	Frx
Mycobacterium_abscessus_subsp.massiliense	Mycobacterium_abscessus_subsp.massiliense	100	0	4556	4558	99.95612
Mycobacterium_abscessus_subsp_abscessus	Mycobacterium_abscessus_subsp.massiliense	97.95052693	4.738468189	4270	4558	93.68144
Mycobacterium_abscessus_subsp_abscessus	Mycobacterium_abscessus_subsp_abscessus	100	0	4940	4942	99.95953
Mycobacterium_abscessus_subsp_bolletii	Mycobacterium_abscessus_subsp.massiliense	97.62332391	4.939236204	4245	4558	93.13295
Mycobacterium_abscessus_subsp_bolletii	Mycobacterium_abscessus_subsp_abscessus	97.6113439	6.814697402	4353	4898	88.87301
Mycobacterium_abscessus_subsp_bolletii	Mycobacterium_abscessus_subsp_bolletii	100	0	4892	4898	99.8775

The ANI values were gotten from http://enve-omics.ce.gatech.edu/g-matrix/.