1 **Six species of nontuberculous mycobacteria carry non-identical 16S rRNA gene copies**

2

3 Keita Takeda[a,b,c,*] , Kinuyo Chikamatsu[a],Yuriko Igarashi[a], Yuta Morishige[a], Yoshiro Murase[a],

4 Akio Aono[a], Hiroyuki Yamada[a], Akiko Takaki[a], Satoshi Mitarai[a,c]

5

6 [a] Department of Mycobacterium Reference and Research, the Research Institute of

7 Tuberculosis, 3-1-24 Matsuyama, Kiyose, Tokyo, 204-0022, Japan

8 [b] Center for Pulmonary Diseases, National Hospital Organization Tokyo National Hospital, 3-

9 1-1 Takeoka, Kiyose, Tokyo 204-8585, Japan

10 [c] Department of Basic Mycobacteriology, Graduate School of Biomedical Science, Nagasaki

11 University, 1-12-4 Sakamoto, Nagasaki, 852-8523, Japan

12

13 **E-mail addresses:** takedak@tokyo-hosp.jp (K. Takeda), chikamatsu@jata.or.jp (K.

14 Chikamatsu), igarashi@jata.or.jp (Y. Igarashi), ymorishige@jata.or.jp (Y. Morishige),

15 ymurase@jata.or.jp (Y. Murase), a.aono1967@gmail.com (A. Aono), hyamada@jata.or.jp (H.

16 Yamada), takaki@jata.or.jp (A. Takaki), mitarai@jata.or.jp (S. Mitarai)

17

18 *Corresponding author at: Department of Mycobacterium Reference and Research, the

19 Research Institute of Tuberculosis, 3-1-24 Matsuyama, Kiyose, Tokyo, 204-0022, Japan

20 E-mail: takedak@tokyo-hosp.jp (K. Takeda)

21 **Abstract**

22 Nontuberculous mycobacteria (NTM) can carry two or more 16S rRNA gene copies
23 that are, in some instances, non-identical. In this study, we used a combined cloning and
24 sequencing approach to analyze 16S rRNA gene sequences of six NTM species,
25 *Mycobacterium cosmeticum*, *M. pallens*, *M. hodleri*, *M. crocinum*, *M. flavescens*, and *M.*
26 *xenopi*. Our approach facilitated the identification of two distinct gene copies in each species.
27 The two *M. cosmeticum* genes had a single nucleotide difference, whereas two nucleotide
28 polymorphisms were identified in *M. hodleri*, *M. flavescens,* and *M. xenopi*. *M. pallens* had a
29 difference in four nucleotides and *M. crocinum* — in 23 nucleotides. Thus, we showed that
30 the six NTM species possess at least two non-identical 16S rRNA gene copies. The full-
31 length sequences of the intraspecies 16S rRNA variants will facilitate NTM identification and
32 sequence analysis of specimens or other samples.

33

34 **Keywords:** 16S rRNA, gene copy, *Mycobacterium*, nontuberculous mycobacteria

35

36 **Abbreviations**: LB, Luria-Bertani; NTM, nontuberculous mycobacteria; WGS, whole-
37 genome sequencing

38

39 **1. Introduction**

40 Nontuberculous mycobacteria (NTM) comprise over 190 species of the genus
41 *Mycobacterium* (Euzéby, 1997). NTM species are classified based on sequence comparisons
42 of select housekeeping genes, such as 16S rRNA, *rpoB*, and *hsp65*, as well as the 16S–23S
43 rRNA internal transcribed spacer region (Domenech et al., 1994; Kim et al., 2015; Kirschner
44 et al., 1993; Lee et al., 2000; Rogall et al., 1990; Stahl and Urbance, 1990; Tagliazucchi et al.,
45 2017; Telenti et al., 1993; Tortoli, 2012; Tortoli et al., 2000). Several commercial tests are
46 available that use these genomic sequences for species identification (Bergmann and Woods,
47 1996; Helb et al., 2010; Huh et al., 2015; Lee et al., 2009; Tanaka et al., 2010). However,
48 some NTM species cannot be identified using these tests because of insufficient genome
49 sequence information.

50 Specifically, Chikamatsu et al. (2018) reported failed attempts to identify several
51 NTM species using the PyroMark Q24 test kit (Qiagen, Tokyo, Japan), which is based on
52 pyrosequencing. Ambiguous bases were found within the 16S rRNA gene in six NTM
53 species, *Mycobacterium cosmeticum*, *M. pallens*, *M. hodleri*, *M. crocinum*, *M. flavescens*, and
54 *M. xenopi* (Chikamatsu et al., 2018). Direct sequencing using the Sanger method suggested
55 that these species carry two non-identical 16S rRNA gene copies. In general, rapid-growing
56 mycobacteria carry two copies of the 16S rRNA gene (Domenech et al., 1994), whereas, with

57 a few exceptions, slow-growing mycobacteria possess only one copy (Ji et al., 1994). For
58 example, it has been reported that isolates of the slow-growing *M. terrae* complex (Ninet et
59 al., 1996) and *M. celatum* (Reischl et al., 1998) harbor two non-identical copies of the 16S
60 rRNA gene. Here, we applied a combined cloning and sequencing approach to unequivocally
61 determine the copy numbers and complete sequences of all 16S rRNA genes of the six NTM
62 species investigated earlier (Chikamatsu et al., 2018): *M. cosmeticum*, *M. pallens*, *M. hodleri*,
63 *M. crocinum*, *M. flavescens*, and *M. xenopi*.
64

65 **2. Materials and Methods**

66 *2.1. Bacterial strains*

67 *M. cosmeticum* JCM14739, *M. pallens* JCM16370, *M. hodleri* JCM12141, and *M.*
68 *crocinum* JCM16369 were obtained from the Japan Collection of Microorganisms (JCM,
69 Ibaraki, Japan). *M. flavescens* ATCC14474 and *M. xenopi* ATCC19250 were acquired from
70 the American Type Culture Collection (ATCC, Manassas, VA). All strains were initially
71 grown on 7H10 agar and then cloned from single colonies. The isolates were sub-cultured in
72 2% Ogawa medium at 37 °C.
73

74 *2.2. DNA extraction*

75 Bacterial DNA was extracted using the Isoplant Kit (Nippon Gene Co., Ltd, Toyama,

76 Japan). Briefly, one inoculation loop (approximately 10 µL) of fresh colonies grown on
77 Ogawa medium were suspended in 300 µL of extraction buffer and again suspended in 150
78 µL of lysis buffer for 15-min incubation at 50 °C. Genomic DNA was extracted with sodium
79 acetate (pH 5.2) on ice for 15 min. After centrifugation (12,000 × *g*, 15 min at 4 °C), the
80 upper phase was transferred to a new tube, and the genomic DNA was precipitated with 70%
81 ethanol. The DNA pellet was dissolved in 50 µL of TE buffer (10 mM Tris-HCl, 1 mM
82 EDTA).
83
84 *2.3. Cloning*
85 The target 16S rRNA genes from each bacterial DNA preparation were amplified with
86 primers 285 (5′ GAG AGT TTG ATC CTG GCT CAG 3′) and rp2 (5′ ACG GCT ACC TTG
87 TTA CGA CTT 3′) yielding the almost complete 16S rRNA gene (Adekambi and Drancourt,
88 2004; Domenech et al., 1994). In brief, 25 µL of a mixture containing ExTaq HS (TaKaRa
89 Bio Inc., Shiga, Japan), 2.5 mM dNTP mixture, 10 µM of each primer, and 5 µL template
90 DNA was used for PCR. Amplification was performed in a GeneAmp PCR System 9700
91 (Applied Biosystems, Foster City, CA) using 30 cycles of 30 s at 94 °C, 30 s at 60 °C, and 90
92 s at 72 °C. Then, the PCR products were purified and cloned using a TOPO TA Cloning Kit
93 (Invitrogen, USA). In brief, the PCR products, salt solution, water, and TOPO® vector using
94 vaccinia topoisomerase I were mixed at room temperature (22–23 °C) and incubated for 30

95 min. The recombinant TA cloning mixes were incubated with *E. coli* competent cells (DH5

96 alpha) for 30 min on ice to perform the transformation. The process was stopped by

97 incubating the samples at 42 °C for 60 s (heat shock), immediately followed by incubation on

98 ice. Super optimal broth with catabolite repression was added to the samples and incubated at

99 37 °C for 1 h.

100 The competent cells were cultured on Luria-Bertani (LB) agar supplemented with 2

101 mg of X-gal. Ten white colonies of each transformation were picked from the LB agar and

102 individually cultured in LB broth. Plasmid DNA was isolated and purified using the a

103 minipreparation Flexiprep Kit (Amersham Biosciences, Little Chalfont, Buckinghamshire,

104 UK) and a column method with a FastGene Gel/PCR Extraction Kit (Nippon Genetics Co.,

105 Ltd, Tokyo, Japan).

106

107 *2.4. Sequence analysis*

108 Sequencing of each 16S rRNA clone was performed using the primers M13 Forward

109 (5′ GTA AAA CGA CGG CCA GT 3′), M13 Reverse (5′ CAG GAA ACA GCT ATG AC

110 3′) and 264 (5′ TGC ACA CAG GCC ACA AGG GA 3′) with a BigDye Terminator Cycle

111 sequencing kit ver. 3.1 (Applied Biosystems) in an ABI 3500 Genetic Analyzer (Applied

112 Biosystems). Finally, the sequences (approximately 1,500 bp each) of the 10 clones of each

113 species (approximately 1,500 bp each) were aligned and further analyzed using Molecular

114 Evolutionary Genetics Analysis software package ver. 7 (Kumar et al., 2016).

115

116 **3. Results**

117 The sequence alignments led to the identification of two non-identical 16S rRNA

118 copies for each of the six NTM species. The results have been deposited with GenBank under

119 accession numbers as follows: *M. cosmeticum* (MH169224 and MH169226), *M. pallens*

120 (MH169208 and MH169209), *M. hodleri* (MH169216 and MH169217), *M. crocinum*

121 (MH169218 and MH169219), *M. flavescens* (MH169220 and MH169222), and *M. xenopi*

122 (MH169221 and MH169241). Nucleotide polymorphisms are shown in Fig. 1. *M.*

123 *cosmeticum* had a single nucleotide difference between the two sequences. Two-nucleotide

124 differences were found in *M. hodleri*, *M. flavescens,* and *M. xenopi*. *M. pallens* had a

125 difference in four nucleotides and *M. crocinum* — in 23 nucleotides.

126

127 **4. Discussion**

128 It is well documented that some species of the genus *Mycobacterium* harbor multiple

129 16S rRNA gene copies with distinct sequences (Chikamatsu et al., 2018; Cilia et al., 1996;

130 Conville et al., 2005; Menendex et al., 2002; Ninet et al., 1996; Reischl et al., 1998; Viezens

131 and Arvand, 2008). In this study, the cloning experiments targeting 16S rRNA genes

132 facilitated the identification of two distinct copies in six NTM species: *M. cosmeticum*, *M.*

133 *pallens*, *M. hodleri*, *M. crocinum*, *M. flavescens*, and *M. xenopi*. Each strain was re-isolated
134 from a single colony, and each of the two 16S rRNA gene copies was reproducibly obtained
135 from multiple clones derived from these isolates. In addition, nucleotide polymorphisms
136 observed for the species-specific gene copies were supported by earlier findings obtained by
137 direct sequencing, e.g., the nucleotide position 185 of the 16S rRNA copies in *M. cosmeticum*
138 was A or G by our cloning-sequencing experiment (Fig. 1), but a mixture of A and G at this
139 position had been indicated earlier by direct Sanger sequencing (Chikamatsu et al., 2018).
140 Our new data were confirmed by multiple clones per strain to minimize the possible impact
141 of technical sequencing errors.

142 Cilia et al. (1996) reported earlier that sequences obtained from clones may be more
143 definitive than sequence data obtained from direct sequencing. Indeed, our sequence data
144 obtained from the cloning experiment unequivocally established the existence of two non-
145 identical gene copies per species, whereas the earlier direct sequencing only suggested a
146 polymorphism based on sequence ambiguities. Hence, it is recommended to avoid direct
147 sequencing for species identification if there are non-identical genomic copies of the target
148 sequence.

149 However, our study could not reveal whether the NTM species had more than two
150 16S rRNA gene copies per genome. It is possible that one genome carries several 16S rRNA
151 copies with identical sequences. This problem might be resolved by whole-genome

152 sequencing (WGS). A database search revealed that *M. cosmeticum* DSM 44829 has two 16S
153 rRNA genes (GenBank accession: NZ_CCBB010000003.1), *M. flavescens* strain M6 has
154 three genes (GenBank accession: NZ_MIHA00000000.1), and *M. xenopi* has one (Strain
155 DSM 43995, GenBank accession: LQQB01000023.1) or two (Strain RIVM700367, GenBank
156 accession: NZ_AJFI01000116.1) genes. These WGS data were obtained by shotgun
157 sequencing, which also has limitations regarding the identification of identical or almost
158 identical gene copies (Goodwin et al., 2016; Nakano et al., 2017; Schadt et al., 2010). Hence,
159 cloning along with sequencing is still required, but improvements in WGS data accuracy will
160 be obtained in the near future by implementing long-read sequencing using the single-
161 molecule real-time sequencing technology (Nakano et al., 2017; Schadt et al., 2010).
162 Currently, the exact 16S rRNA gene copy number is not yet known for *M. pallens*, *M.*
163 *hodleri,* and *M. crocinum*, which requires further analysis.

164 In this study, we established the existence of two 16S rRNA gene copies for each of
165 the six NTM species examined. However, species identification using 16S rRNA sequencing
166 can be challenging because of the detected nucleotide polymorphisms. The identification of
167 two non-identical 16S rRNA copies in the six NTM species will be helpful for sequence
168 analyses of specimens or other samples and sequencing efforts.

169

170 **Declaration of interest**

171 The authors declare no conflict of interest.

172

173 **References**

174 Adekambi, T. and Drancourt, M., 2004. Dissection of phylogenetic relationships among 19
175 rapidly growing *Mycobacterium* species by 16S rRNA, *hsp65*, *sodA*, *recA* and *rpoB*
176 gene sequencing. Int. J. Syst. Evol. Microbiol. 54, 2095–2105.
177 Bergmann, J.S. and Woods, G.L., 1996. Clinical evaluation of the Roche AMPLICOR PCR
178 *Mycobacterium tuberculosi*s test for detection of *M. tuberculosis* in respiratory
179 specimens. J. Clin. Microbiol. 34, 1083–1085.
180 Chikamatsu, K., Aono, A., Hata, H., et al., 2018. Evaluation of PyroMark Q24
181 pyrosequencing as a method for the identification of mycobacteria. Diagn. Microbiol.
182 Infect. Dis. 90, 35–39.
183 Cilia, V., Lafay, B., Christen, R., 1996. Sequence heterogeneities among 16S ribosomal RNA
184 sequences, and their effect on phylogenetic analyses at the species level. Mol. Biol.
185 Evol. 13, 451–461.
186 Conville, P.S. and Witebsky, F.G., 2005. Multiple copies of the 16S rRNA gene in *Nocardia*
187 *nova* isolates and implications for sequence-based identification procedures. J. Clin.
188 Microbiol. 43, 2881–2885.
189 Domenech, P., Menendez, M.C., Garcia, M.J., 1994. Restriction fragment length

190 polymorphism of 16S rRNA genes in the differentiation of fast-growing
191 mycobacterial species. FEMS Microbiol. Lett. 116, 19–24.
192 Euzéby, J.P., 1997. List of prokaryotic names with standing in nomenclature: a folder
193 available on the Internet. Int. J. Syst. Bacteriol. 47, 590–592.
194 Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-
195 generation sequencing technologies. Nat. Rev. Genet. 17, 333–351.
196 Helb, D., Jones, M., Story, E., et al., 2010. Rapid detection of *Mycobacterium tuberculosis*
197 and rifampin resistance by use of on-demand, near-patient technology. J. Clin.
198 Microbiol. 48, 229–237.
199 Huh, H.J., Kwon, H.J., Ki, C.S., et al., 2015. Comparison of the genedia MTB detection kit
200 and the cobas TaqMan MTB assay for detection of *Mycobacterium tuberculosis* in
201 respiratory specimens. J. Clin. Microbiol. 53, 1012–1014.
202 Ji, Y.E., Colston, M.J., Cox, R.A., 1994. The ribosomal RNA (*rrn*) operons of fast-growing
203 mycobacteria: primary and secondary structure and their relation to *rrn* operons of
204 pathogenic slow-growers. Microbiology 140, 2829–2840.
205 Kim, S.Y., Park, H.Y., Jeong, B.H., et al., 2015. Molecular analysis of clinical isolates
206 previously diagnosed as *Mycobacterium intracellulare* reveals incidental findings of
207 "*Mycobacterium indicus pranii*" genotypes in human lung infection. BMC Infect. Dis.
208 15, 406.

209 Kirschner, P., Springer, B., Vogel, U., et al., 1993. Genotypic identification of mycobacteria
210 by nucleic acid sequence determination: report of a 2-year experience in a clinical
211 laboratory. J. Clin. Microbiol. 31, 2882–2889.
212 Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics
213 Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874.
214 Lee, H., Park, H.J., Cho, S., et al., 2000. Species identification of mycobacteria by PCR-
215 restriction fragment length polymorphism of the *rpoB* gene. J. Clin. Microbiol. 38,
216 2966–2971.
217 Lee, A.S., Jelfs, P., Sintchenko, V., et al., 2009. Identification of non-tuberculous
218 mycobacteria: utility of the GenoType Mycobacterium CM/AS assay compared with
219 HPLC and 16S rRNA gene sequencing. J. Med. Microbiol. 58, 900–904.
220 Menendez, M.C., Garcia, M.J., Navarro, M.C., et al., 2002. Characterization of an rRNA
221 operon (*rrnB*) of *Mycobacterium fortuitum* and other mycobacterial species:
222 implications for the classification of mycobacteria. J. Bacteriol. 184, 1078–1088.
223 Nakano, K., Shiroma, A., Shimoji, M., et al., 2017. Advantages of genome sequencing by
224 long-read sequencer using SMRT technology in medical area. Hum. Cell 30, 149–
225 161.
226 Ninet, B., Monod, M., Emler, S., et al., 1996. Two different 16S rRNA genes in a
227 mycobacterial strain. J. Clin. Microbiol. 34, 2531–2536.

228 Reischl, U., Feldmann, K., Naumann, L., et al., 1998. 16S rRNA sequence diversity in
229 *Mycobacterium celatum* strains caused by presence of two different copies of 16S
230 rRNA gene. J. Clin. Microbiol. 36, 1761–1764.
231 Rogall, T., Wolters, J., Flohr, T., et al., 1990. Towards a phylogeny and definition of species at
232 the molecular level within the genus *Mycobacterium*. Int. J. Syst. Bacteriol. 40, 323–
233 330.
234 Schadt, E.E., Turner, S., Kasarskis, A., 2010. A window into third-generation sequencing.
235 Hum. Mol. Genet. 19, 227–240.
236 Stahl, D.A. and Urbance, J.W., 1990. The division between fast- and slow-growing species
237 corresponds to natural relationships among the mycobacteria. J. Bacteriol. 172, 116–
238 124.
239 Tagliazucchi, S., Fabio, A., Bettua, C., et al., 2017. The new phylogeny of the genus
240 *Mycobacterium*: The old and the news. Infect. Genet. Evol. 56, 19–25.
241 Tanaka, H., Hirose, H., Kato, Y., et al., 2010. Clinical evaluation of TRCRapid M.TB for
242 detection of *Mycobacterium tuberculosis* complex in respiratory and nonrespiratory
243 specimens. J. Clin. Microbiol. 48, 1536–1541.
244 Telenti, A., Marchesi, F., Balz, M., et al., 1993. Rapid identification of mycobacteria to the
245 species level by polymerase chain reaction and restriction enzyme analysis. J. Clin.
246 Microbiol. 31, 175–178.

247 Tortoli, E., Fedrizzi, T., Meehan, C.J., et al., 2000. Novel diagnostic algorithm for
248 identification of mycobacteria using genus-specific amplification of the 16S-23S
249 rRNA gene spacer and restriction endonucleases. J. Clin. Microbiol. 38, 1094–1104.

250 Tortoli, E., 2012. Phylogeny of the genus *Mycobacterium*: many doubts, few certainties.
251 Infect. Genet. Evol. 12, 827–831.

252 Viezens, J. and Arvand, M., 2008. Simultaneous presence of two different copies of the 16S
253 rRNA gene in *Bartonella henselae*. Microbiology 154, 2881–2886.

254

255

256

257 **Figure legend**

258 **Fig. 1.** Nucleotide polymorphisms between two non-identical 16S rRNA sequences of six
259 NTM species. Nucleotide positions were derived from an alignment with the 16S rRNA gene
260 of *M. tuberculosis* H37Rv ATCC272 (GenBank accession: NC_000962).

261

Figure 1

M. cosmeticum

	185
Variant 1 :	ATAGGACTCC**A**GCCTTCATGG
Variant 2 :	. . . .**G**. . . . .

M. pallens

	185—187	470
Variant 1 :	ATAGGACCAC**ATG**CTTCATGGTG	GACGGTACCT**A**TAGAAGAAGC
Variant 2 :	. . . .**GGC**.	. . . .**G**. . . . .

M. hodleri

	197|198
Variant 1 :	GATGCATGTC**TC**TTGGTGGAAA
Variant 2 :	. . . .**CT**. . . . .

M. crocinum

	73 77 80 87 90 93	183|185–187 196|197	
Variant 1 :	AACGG**T**AAG-GC**C**CTTCGG**G**GT-AC**A**CGAGT	GGACC**A**C**GGC**CTTCATGG**GT**TGTGG	
Variant 2 :	. . . . .**A**. . .**T**. .**T**. . . . . .**A**. .**G**. .**T**.	. . . . .**G**.**ATG**. . . . .**TG**.	
	441——445 467——471		1245
Variant 1 :	TTTCA**GTAGG**GACGAAGCGCAAGTGACGGTA**CCTAT**AGAAG		CCGGTACAAA**A**GGCTGCGATG
Variant 2 :	. . . . .**CCCAC**. . . . .**GTGGG**.		. . . .**G**. . . . .

M. flavescens

	184|185	
Variant 1 :	AATATTCCCT**AT**TGGTCGCATG	
Variant 2 :	..........**GC**..........	
M. xenopi		
	210	434
Variant 1 :	TGGTGGAAAG**T**GTTTGGTAGC	GTTGTAAACC**T**CTTTCAGCCT
Variant 2 :	..........**C**..........	..........**C**..........

Highlights

- 16S rRNA gene sequences of six nontuberculous mycobacterium species were obtained
- Two distinct 16S rRNA gene copies were obtained from each of the six species
- The two copies varied in 1–23 nucleotides, depending on the species