JCM
Journals.ASM.org

Support from Phylogenomic Networks and Subspecies Signatures for Separation of *Mycobacterium massiliense* from *Mycobacterium bolletii*

Joon Liang Tan,[a] Yun Fong Ngeow,[a] Siew Woh Choo[b]

Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia[a]; Department of Oral Biology and Biomedical Sciences, Faculty of Dentistry, University of Malaya, Kuala Lumpur, Malaysia[b]

***Mycobacterium abscessus* subspecies classification has important clinical implications. We used phylogenomic network and amino acid analyses to provide evidence for the separation of *Mycobacterium bolletii* and *Mycobacterium massiliense* into two distinct subspecies which can potentially be differentiated rapidly by their protein signatures.**

Mycobacterium abscessus has become one of the most frequently isolated nontuberculous mycobacterium (NTM) in clinical laboratories. It is associated with chronic, recurrent infections that are difficult to treat, partly because of its resistance to many of the usual medications for NTM infections. This species was previously divided into three subspecies (*M. abscessus*, *M. massiliense*, and *M. bolletii*) based on biological and genetic differences (1–3). Currently, however, only two subspecies are recognized; while *M. abscessus* is retained as *Mycobacterium abscessus* subsp. *abscessus*, *M. massiliense* and *M. bolletii* are placed in the same subspecies designated *Mycobacterium abscessus* subsp. *bolletii* (4). This tenuous merging of *M. massiliense* and *M. bolletii* is still being debated as recent publications support the previous three-subspecies classification (5). Here, we present more evidence for the retention of the former three-subspecies taxonomic division, which correlates better with the expected treatment outcomes in infected patients (6).

(This research was conducted by J. L. Tan in partial fulfillment of the requirements for a Ph.D. from University of Malaya, Kuala Lumpur, Malaysia.)

For our genomic and amino acid analyses, we used 12 genomes from strains isolated in the Diagnostic Microbiology Laboratory

Received 26 February 2015 **Returned for modification** 3 May 2015
Accepted 9 June 2015

Accepted manuscript posted online 8 July 2015

Citation Tan JL, Ngeow YF, Choo SW. 2015. Support from phylogenomic networks and subspecies signatures for separation of *Mycobacterium massiliense* from *Mycobacterium bolletii*. J Clin Microbiol 53:3042–3046. doi:10.1128/JCM.00541-15.

Editor: G. A. Land

Address correspondence to Yun Fong Ngeow, yunngeow@yahoo.com.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.00541-15.

doi:10.1128/JCM.00541-15

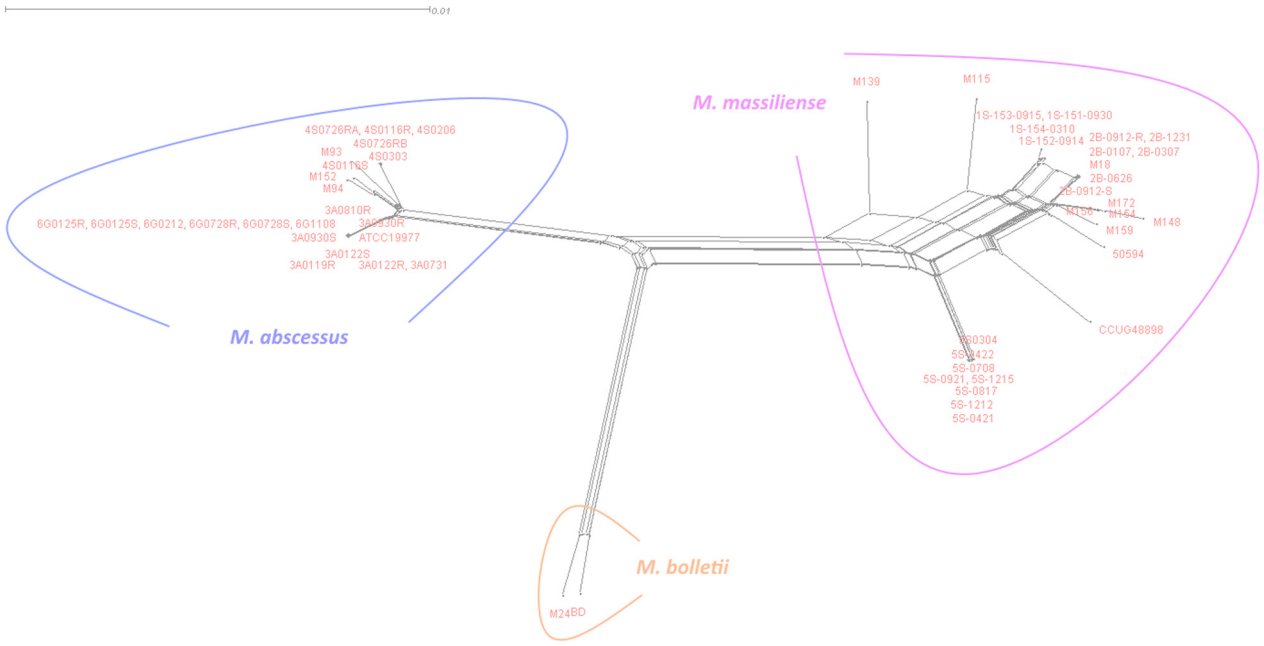

FIG 1 Phylogenomic split network tree obtained from the concatenation of single-copy core genes from *M. abscessus* subspecies. *M. massiliense* (right), *M. bolletii* (center), and *M. abscessus* (left) can be seen clearly as distinct groups.

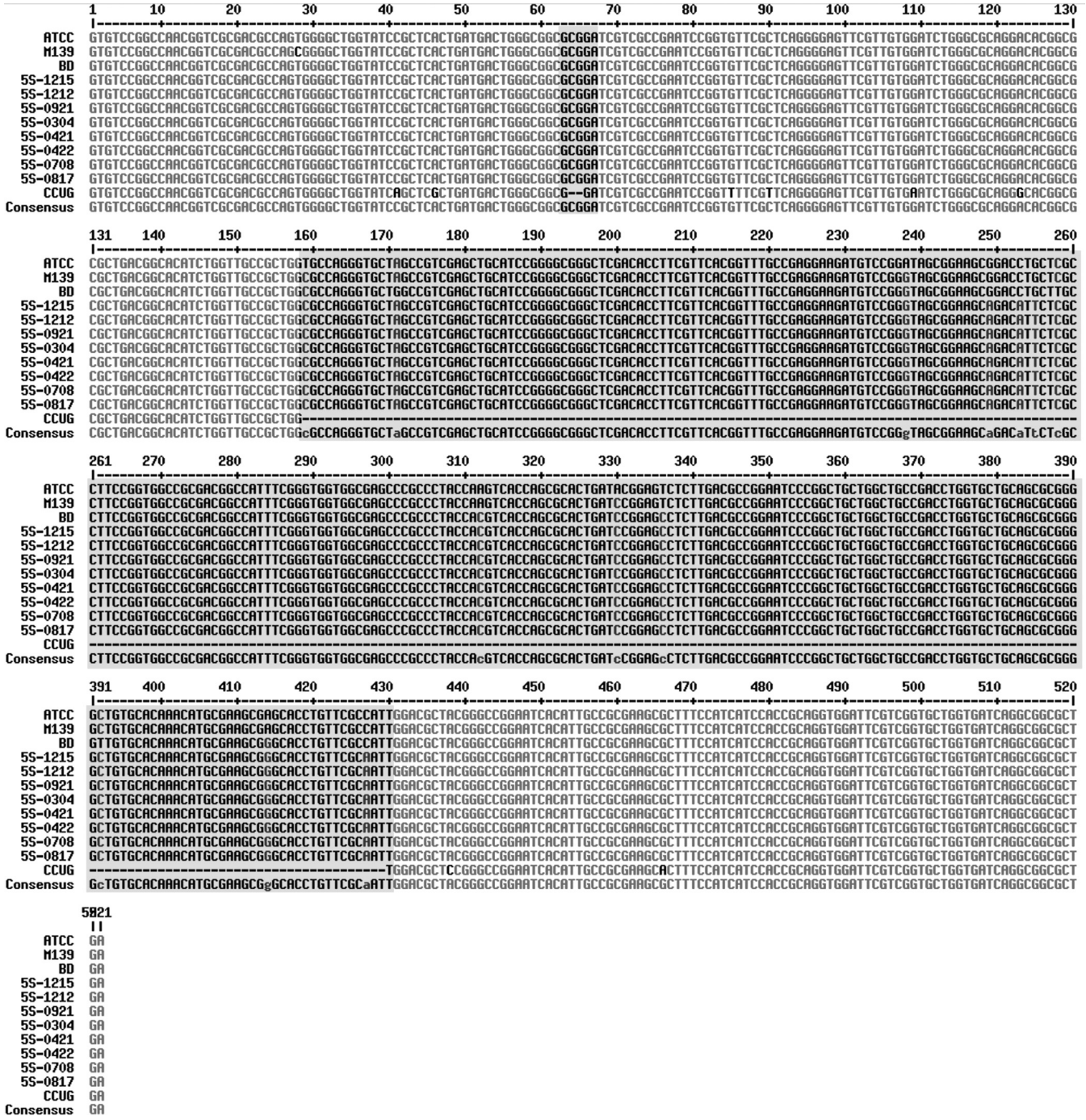

FIG 2 Multiple sequence alignments of *erm41* showing features of *M. massiliense* M139 and 5S strains compared to those of the type strains of *M. abscessus* ATCC 19977^T, *M. massiliense* CCUG 48898^T, and *M. bolletii* BD^T. The *M. massiliense* signatures are (i) deletions at positions 64 and 65 and (ii) a 274-bp deletion after position 159.

of the University of Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia, and 41 downloaded from the NCBI Genome database on July 2014 (see Table S1 in the supplemental material). Eleven of the UMMC strains have been previously reported to be *M. abscessus* (M93, M94, and M152), *M. bolletii* (M24), and *M. massiliense* (M18, M115, M152, M172, M159, M156, and M148). One strain, M139, was shown to have an ambiguous taxonomic position in a number of studies (7, 8).

The protein sequences for all strains were retrieved using the self-training structural annotation algorithm of GeneMarkS (9). To define orthologous sequences, we used the CD-HIT program (10) with the following criteria: word length of 2, local sequence identity threshold of 0.4, alignment coverage for both sequences of 0.4, and greedy algorithm off. We also used the BLASTClust program with the following parameters: reference and query sequences must cover at least 40% of the aligned sequence and reference and query sequences must have a minimum identity of 40% (11). To reduce false-positive results due to algorithmic er-

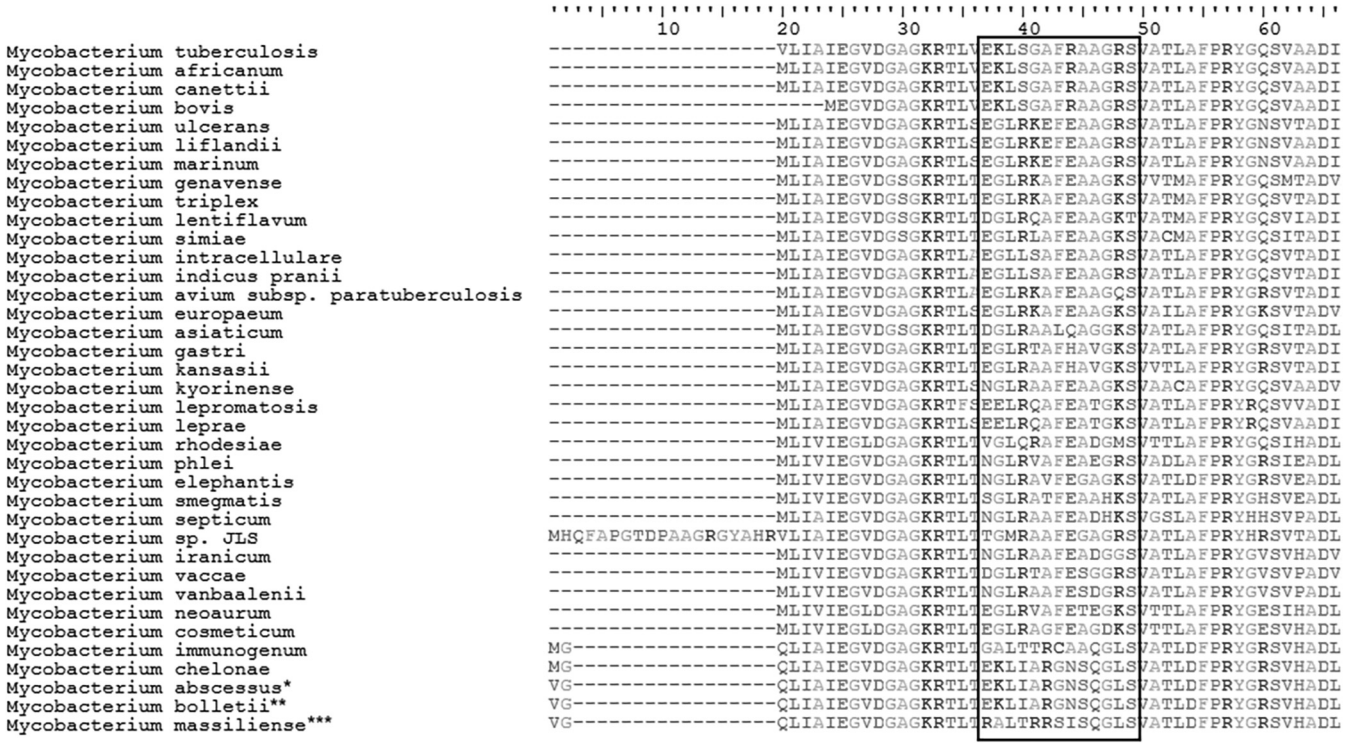

FIG 3 Consistent protein signatures in *M. massiliense* identified in multiple alignments of thymidylate kinase from *M. abscessus* and other selected mycobacteria: *23 strains of *M. abscessus* subsp. *abscessus*; ** 2 strains of *M. bolletii*; ***28 strains of *M. massiliense*.

rors, only the consensus sequences from both programs were extracted and used as the final list of orthologs. Nonduplicated conserved protein orthologs were aligned in MAFFT (12).

The protein sequence alignments were used as the reference for codon alignments in PAL2NAL (13). The aligned nucleotide sequences were concatenated into supersequences for phylogenomic analysis using the Neighbor-Net algorithm implemented in SplitsTree4 (14). This algorithm was considered the best for the resolution of complex taxonomy (15). To assess the subspecies classification derived from our network tree, we looked for subspecies-specific polymorphisms previously described for the erythromycin ribosome methyltransferase (*erm41*) and 16S to 23S internal transcribed spacer (ITS) genes.

Our network-based phylogenomic tree showed reticulated branches leading to three clearly distinctive monophyletic groups representing the three subspecies of *M. abscessus* (Fig. 1). The M139 and the 5S strains (5S-0421, 5S-0422, 5S-0708, 5S-0817, 5S-0921, 5S-1212, 5S-1215, and 5S-0304) clustered with the other *M. massiliense* strains. None of the branches in any of the three major clusters bifurcated to the other two clusters. The presence of 3-dimensional-like splits within the branches indicated incompatible phylogenetic signals that are likely to be the result of recombination following the horizontal transfer of genetic material among strains. Indeed, the recombination among our *M. abscessus* strains is statistically supported by the pairwise homoplasy index (PHI) ($P = 0$) (16). The incompatible signals occurred at random points in the tree, suggesting that recombination has occurred in ancestral states and within the respective subspecies. We also noticed unusual conflicting signals within the *M. massiliense* cluster, appearing as a major reticulation connecting the *M. massiliense* strains and suggesting a higher degree of genetic recombination in *M. massiliense* compared to that in the other two subspecies. To test the validity of this network phylogenomics approach, we used it on three members of the *M. avium* complex and found a clear separation of *Mycobacterium avium* subsp. *paratuberculosis*, *Mycobacterium avium* subsp. *hominissuis*, and *Mycobacterium avium* subsp. *avium* into three distinctive monophyletic groups, as observed with the *M. abscessus* complex (see Fig. S1 in the supplemental material).

M. massiliense is known to be different from the other two subspecies in having a truncated *erm41* with nucleotide deletions at the 64th to 65th and 159th to 432nd positions, as well as mutations in the ITS (a A to G substitution at the 60th position and a C insertion at the 102nd position) (2). M139 and the eight 5S strains previously classified as *M. massiliense* and appearing as *M. massiliense* in our phylogenomic tree did not show the *erm41* features associated with *M. massiliense* (Fig. 2). M139 additionally lacked the ITS mutations characteristic of *M. massiliense* and did not show inducible resistance to macrolides (17). Overall, however, there was good concordance (83%) between the subspecies classifications by *erm41* signatures and by the network tree.

In the multiple sequence alignment of the orthologous proteins from our 53 strains, we noted 46 proteins with at least one amino acid that can be used to differentiate the three subspecies (see Table S2 in the supplemental material) and another two proteins (thymidylate kinase [tk] and 30S ribosomal protein S3 [S3]) that can differentiate *M. massiliense* from the other two subspecies (Fig. 3 and 4). We used BLAST to search the amino acid sequences of tk3 and S3 against all *Mycobacterium* genomes in the NCBI database and found them in 37 and 44 species, respectively. After realigning against these mycobacterial species, we confirmed the amino acid signatures of tk (RALTRRSISQGLS at position 20 to

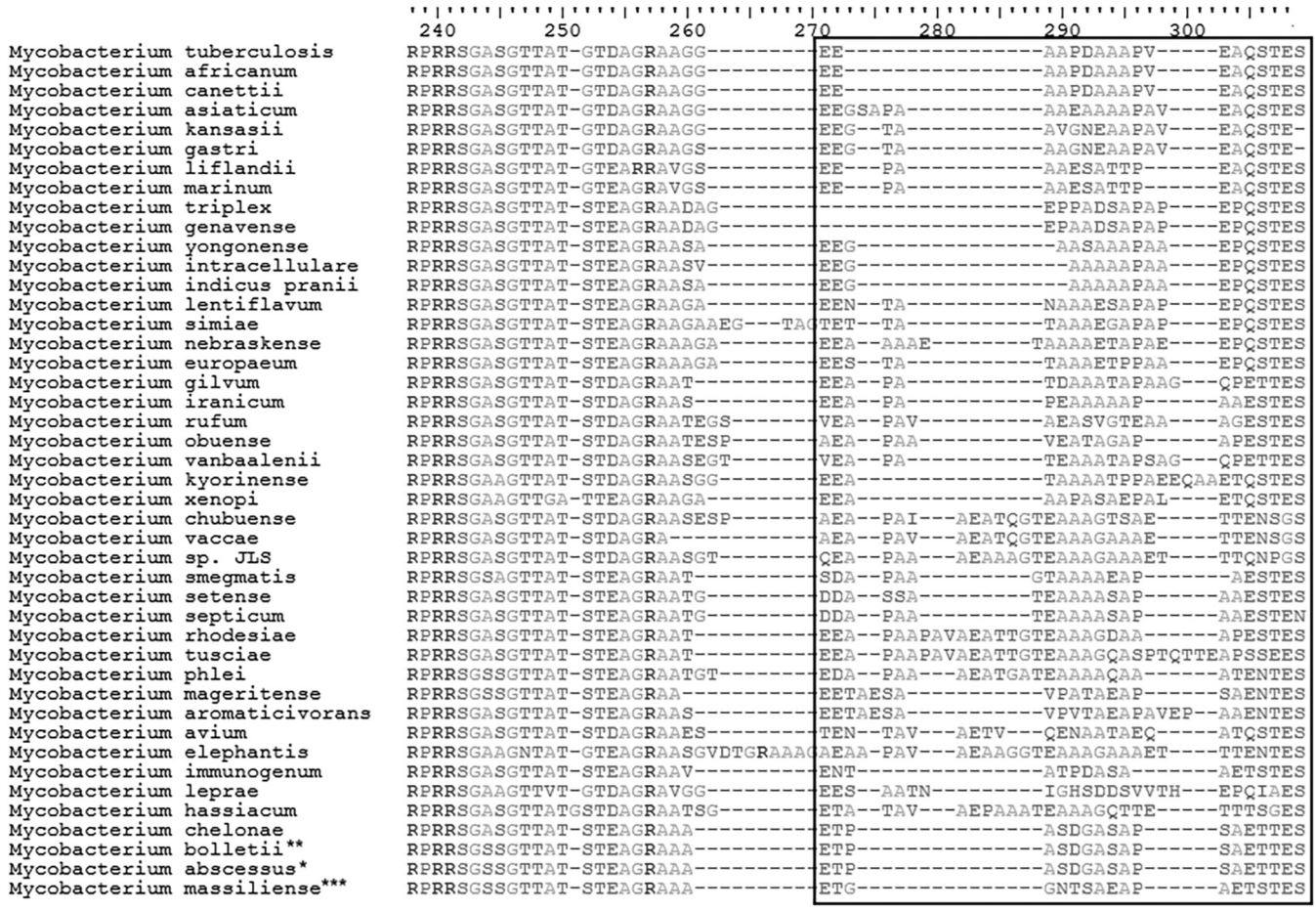

FIG 4 Consistent protein signatures in *M. massiliense* identified in multiple alignments of 30S ribosomal protein S3 from *M. abscessus* and other selected mycobacteria: *23 strains of *M. abscessus* subsp. *abscessus*; ** 2 strains of *M. bolletii*; ***28 strains of *M. massiliense*.

30) and S3 (ETGGNTSAEAPAETSTES at position 260 to 277) to be specific for *M. massiliense* (Fig. 3 and 4). The presence of these signatures in M139 and the 5S strains supported their classification as *M. massiliense*, in agreement with the classification by the phylogenomic network. They will need to be experimentally verified as suitable biomarkers for the identification of *M. massiliense* in clinical material.

It is well known that *M. abscessus* subspecies exhibit different clinical and epidemiological features (18, 19). *M. massiliense* is more susceptible to antibiotics but is also more often associated with clinical infections. *M. bolletii*, on the other hand, is rarely isolated from clinical material but is more highly antibiotic resistant. While the reasons behind these differences are still unclear, there is sufficient justification for subspecies identification in patient care. Our analyses support the division of *M. abscessus* into three subspecies and the reinstatement of *M. massiliense* as a taxon independent of *M. bolletii*. The specific identification of these two subspecies which show different antibiotic susceptibilities will enable the clinician to prescribe appropriate antibiotics for the effective treatment of infections.

ACKNOWLEDGMENTS

This study was supported by research grants UM.C/625/1/HIR/MOHE/CHAN/14/4 and UM.C/HIR/MOHE/08 from the University of Malaya.

REFERENCES

1. **Adékambi T, Berger P, Raoult D, Drancourt M.** 2006. *rpoB* gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov. Int J Syst Evol Microbiol **56:**133–143. http://dx.doi.org/10.1099/ijs.0.63969-0.
2. **Adékambi T, Reynaud-Gaubert M, Greub G, Gevaudan MJ, La Scola B, Raoult D, Drancourt M.** 2004. Amoebal coculture of "*Mycobacterium massiliense*" sp. nov. from the sputum of a patient with hemoptoic pneumonia. J Clin Microbiol **42:**5493–5501. http://dx.doi.org/10.1128/JCM.42.12.5493-5501.2004.
3. **Moore M, Frerichs JB.** 1953. An unusual acid-fast infection of the knee with subcutaneous, abscess-like lesions of the gluteal region; report of a case with a study of the organism, *Mycobacterium abscessus*, n. sp. J Investig Dermatol **20:**133–169.
4. **Leao SC, Tortoli E, Euzeby JP, Garcia MJ.** 2011. Proposal that *Mycobacterium massiliense* and *Mycobacterium bolletii* be united and reclassified as *Mycobacterium abscessus* subsp. *bolletii* comb. nov., designation of *Mycobacterium abscessus* subsp. *abscessus* subsp. nov. and emended description of *Mycobacterium abscessus*. Int J Syst Evol Microbiol **61:**2311–2313. http://dx.doi.org/10.1099/ijs.0.023770-0.
5. **Cho YJ, Yi H, Chun J, Cho SN, Daley CL, Koh WJ, Shin SJ.** 2013. The genome sequence of '*Mycobacterium massiliense*' strain CIP 108297 suggests the independent taxonomic status of the *Mycobacterium abscessus* complex at the subspecies level. PLoS One **8:**e81560. http://dx.doi.org/10.1371/journal.pone.0081560.
6. **Koh WJ, Jeon K, Lee NY, Kim BJ, Kook YH, Lee SH, Park YK, Kim CK, Shin SJ, Huitt GA, Daley CL, Kwon OJ.** 2011. Clinical significance of differentiation of *Mycobacterium massiliense* from *Mycobacterium absces-*

sus. Am J Respir Crit Care Med **183:**405–410. http://dx.doi.org/10.1164/rccm.201003-0395OC.

7. **Tan JL, Khang TF, Ngeow YF, Choo SW.** 2013. A phylogenomic approach to bacterial subspecies classification: proof of concept in *Mycobacterium abscessus*. BMC Genomics **14:**879. http://dx.doi.org/10.1186/1471-2164-14-879.
8. **Wong YL, Ong CS, Ngeow YF.** 2012. Molecular typing of *Mycobacterium abscessus* based on tandem-repeat polymorphism. J Clin Microbiol **50:**3084–3088. http://dx.doi.org/10.1128/JCM.00753-12.
9. **Besemer J, Lomsadze A, Borodovsky M.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res **29:**2607–2618. http://dx.doi.org/10.1093/nar/29.12.2607.
10. **Fu L, Niu B, Zhu Z, Wu S, Li W.** 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics **28:**3150–3152. http://dx.doi.org/10.1093/bioinformatics/bts565.
11. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
12. **Katoh K, Standley DM.** 2014. MAFFT: iterative refinement and additional methods. Methods Mol Biol **1079:**131–146. http://dx.doi.org/10.1007/978-1-62703-646-7_8.
13. **Suyama M, Torrents D, Bork P.** 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res **34:**W609-612. http://dx.doi.org/10.1093/nar/gkl315.
14. **Huson DH, Bryant D.** 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **23:**254–267. http://dx.doi.org/10.1093/molbev/msj030.
15. **Morrison DA.** 2005. Networks in phylogenetic analysis: new tools for population biology. Int J Parasitol **35:**567–582. http://dx.doi.org/10.1016/j.ijpara.2005.02.007.
16. **Bruen TC, Philippe H, Bryant D.** 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics **172:**2665–2681. http://dx.doi.org/10.1534/genetics.105.048975.
17. **Griffith DE.** 2014. *Mycobacterium abscessus* subsp *abscessus* lung disease: 'trouble ahead, trouble behind. . . .' F1000Prime Rep **6:**107. http://dx.doi.org/10.12703/P6-107.
18. **Kim SY, Kim CK, Bae IK, Jeong SH, Yim JJ, Jung JY, Park MS, Kim YS, Kim SK, Chang J, Kang YA.** 2015. The drug susceptibility profile and inducible resistance to macrolides of *Mycobacterium abscessus* and *Mycobacterium massiliense* in Korea. Diagn Microbiol Infect Dis **81:**107–111. http://dx.doi.org/10.1016/j.diagmicrobio.2014.10.007.
19. **Nie W, Duan H, Huang H, Lu Y, Bi D, Chu N.** 2014. Species identification of *Mycobacterium abscessus* subsp. *abscessus* and *Mycobacterium abscessus* subsp. *bolletii* using *rpoB* and *hsp65*, and susceptibility testing to eight antibiotics. Int J Infect Dis **25:**170–174. http://dx.doi.org/10.1016/j.ijid.2014.02.014.