Metabolic phenotype of clinical and environmental *Mycobacterium avium* subsp. *hominissuis* isolates

Andrea Sanchini [1] , **Flavia Dematheis** [2] , **Torsten Semmler** [3] , **Astrid Lewin** [Corresp. 1]

[1] Division 16, Mycotic and Parasitic Agents and Mycobacteria, Robert Koch Institute, Berlin, Germany

[2] Institute of Microbiology and Epizootics, Free University Berlin, Berlin, Germany

[3] NG 1 Microbial Genomics, Robert Koch Institute, Berlin, Germany

Corresponding Author: Astrid Lewin
Email address: LewinA@rki.de

Background. *Mycobacterium avium* subsp. *hominissuis* (MAH) is an emerging opportunistic human pathogen. It can cause pulmonary infections, lymphadenitis and disseminated infections in immuno-compromised patients. In addition, MAH is widespread in the environment, since it has been isolated from water, soil or dust. In the recent years, knowledge on MAH at molecular level substantially increased. On the contrary, knowledge of the MAH metabolic phenotypes remains limited.

Methods. In this study for the first time we analyzed the metabolic substrate utilization of ten MAH isolates, five from clinical and five from environmental source. We used the BIOLOG Phenotype Microarray™ technology. This technology permits the rapid and global analysis of metabolic phenotypes.

Results. The ten MAH isolates tested showed different metabolic patterns pointing to high intra-species diversity. Our MAH isolates preferred to use fatty acids such as Tween, caproic, butyric and propionic acid as a carbon source, and L-cysteine as a nitrogen source. Environmental MAH isolates resulted to be more metabolically active than clinical isolates, since the former metabolized more strongly butyric acid ($p = 0.0209$) and propionic acid ($p = 0.00307$).

Discussion. Our study provides new insight into the metabolism of MAH. Understanding how bacteria utilize substrates during infection might help the developing of strategies to fight such infections.

1   **Article Title**

2   Metabolic phenotype of clinical and environmental *Mycobacterium avium* subsp. *hominissuis*

3   isolates

5   **Authors**

6   Andrea Sanchini[1], Flavia Dematheis[2], Torsten Semmler[3], Astrid Lewin[1,*]

8   **Author affiliations**

9   [1] Division 16, Mycotic and Parasitic Agents and Mycobacteria, Robert Koch Institute, Berlin,

10 Germany

11 [2] Institute of Microbiology and Epizootics, Free University Berlin, Berlin, Germany

12 [3] NG 1 Microbial Genomics, Robert Koch Institute, Berlin, Germany

14 **Corresponding author**

15 Dr Astrid Lewin

16 Division 16, Mycotic and Parasitic Agents and Mycobacteria, Robert Koch Institute, Seestraße

17 10, 13353, Berlin, Germany.

18 Telephone: +4930187542112

19   E-mail: LewinA@rki.de

21   **Abstract**

22   **Background.** *Mycobacterium avium* subsp. *hominissuis* (MAH) is an emerging opportunistic

23   human pathogen. It can cause pulmonary infections, lymphadenitis and disseminated infections

24   in immuno-compromised patients. In addition, MAH is widespread in the environment, since it

25   has been isolated from water, soil or dust. In the recent years, knowledge on MAH at molecular

26   level substantially increased. On the contrary, knowledge of the MAH metabolic phenotypes

27   remains limited.

28   **Methods.** In this study for the first time we analyzed the metabolic substrate utilization of ten

29   MAH isolates, five from clinical and five from environmental source. We used the BIOLOG

30   Phenotype Microarray™ technology. This technology permits the rapid and global analysis of

31   metabolic phenotypes.

32   **Results.** The ten MAH isolates tested showed different metabolic patterns pointing to high intra-

33   species diversity. Our MAH isolates preferred to use fatty acids such as Tween, caproic, butyric

34   and propionic acid as a carbon source, and L-cysteine as a nitrogen source. Environmental MAH

35   isolates resulted to be more metabolically active than clinical isolates, since the former

36   metabolized more strongly butyric acid ($p = 0.0209$) and propionic acid ($p = 0.00307$).

37   **Discussion.** Our study provides new insight into the metabolism of MAH. Understanding how

38   bacteria utilize substrates during infection might help the developing of strategies to fight such

39   infections.

40

41

42 **Introduction**

43 *Mycobacterium avium* subsp. *hominissuis* (MAH) is clinically one of the most relevant non-

44 tuberculous mycobacteria (Tortoli 2014). MAH is an opportunistic human pathogen causing

45 pulmonary infections, lymphadenitis in small children and disseminated infections (Despierres et

46 al. 2012; Rindi & Garzelli 2014). It is of increasing public health relevance, with reports of

47 MAH infections increasing worldwide (Hoefsloot et al. 2013). Moreover, MAH is widespread in

48 the environment (Falkinham 2013; Lahiri et al. 2014). In recent years, there have been

49 substantial advances in the analysis of bacteria at the molecular level. Indeed, several whole

50 genome sequences are now available for many mycobacterial species, including MAH

51 (Bannantine et al. 2014; Kim et al. 2012; Uchiya et al. 2013; Wynne et al. 2010). In contrast,

52 there has been little concomitant advance in knowledge at the phenotypic level. Phenotype

53 analysis deserves greater attention, as it is the phenotype that selection pressure acts upon to

54 confer evolutionary advantages to the bacterial species (Plata et al. 2015). In order to address this

55 knowledge gap for bacterial phenotypes, BIOLOG Inc. developed the Phenotype MicroArray™

56 (PM) (BIOLOG, Hayward CA), a high throughput method for the rapid and global analysis of

57 microbial metabolic phenotypes (Bochner 2003; Bochner 2009; Bochner et al. 2001; Bochner et

58 al. 2008). The PM technology consists of several commercially available 96-well plates in which

59 every well has a different substrate, allowing nearly 2000 different microbial metabolic

60 phenotypes to be tested (Bochner 2003; Bochner 2009; Bochner et al. 2001; Bochner et al.

61 2008). PM technology has been applied to several microorganisms, including mycobacteria

62 (Baloni et al. 2014; Bochner et al. 2008; Borglin et al. 2012; Chen et al. 2012; Gupta et al. 2015;

63 Johnson et al. 2008; Khatri et al. 2013; Lofthouse et al. 2013; Mackie et al. 2014; Mishra &

64 Daniels 2013; Nai et al. 2013; Omsland et al. 2009; Tohsato & Mori 2008). One possible

65   application of PM is the detection of phenotype changes due to gene knock-out. For example,

66   Chen and co-authors showed that a *leuD* mutant of *M. avium* subsp *paratuberculosis* lost the

67   ability to use several carbon, nitrogen, sulfur and phosphorous substrates (Chen et al. 2012).

68   Other researchers showed that the use of 12 carbon substrates differentiated *M. tuberculosis* from

69   *M. bovis* (Khatri et al. 2013; Lofthouse et al. 2013).

70   In this study we tested clinical and environmental isolates of MAH using the PM technology.

71   Our aim was to describe the metabolic substrates utilized by MAH isolates and to identify any

72   metabolic differences between clinical and environmental MAH isolates.

74   **Materials and Methods**

75   **Bacterial isolates and BIOLOG phenotype microarray**

76   We analyzed five clinical and five environmental MAH isolates (Table 1).

77   We performed the BIOLOG Phenotype Microarray ™ (BIOLOG, Hayward, CA) according to

78   the manufacturer´s recommendations. The technology is based on the measurement of bacterial

79   respiration, which produces NADH (Bochner et al. 2001). If bacteria are able to metabolize a

80   specific substrate, electrons from NADH reduce a tetrazolium dye in an irreversible reaction

81   generating a purple color in the PM plate wells. This color change is measured and recorded

82   every 15 minutes by the reporter instrument OmniLog™ (BIOLOG, Hayward, CA), generating a

83   kinetic response curve for each well (Bochner 2003; Bochner 2009).

84   The ten MAH isolates were tested with the 96-wells plates PM1 to PM4, containing 190 carbon

85   (PM1 and PM2), 95 nitrogen (PM3), 59 phosphorous (PM4) and 35 sulfur (PM4) substrates. The

86   PM plates 1, 2 and 3 include one negative control well, in which bacteria are tested without any

87   substrate. The PM4 plate includes two negative control wells, one for the phosphorus and one for

88   the sulfur substrates. All isolates were tested three times. Briefly, we cultivated each MAH

89   isolate in 30 ml of 7H10 Middlebrook medium supplemented with 10% modified ADC-

90   enrichment (2% of glucose, 5% of BSA, 0,85% of NaCl) until an $OD_{600\,nm}$ of 0.3-0.6 was

91   achieved (mid-logarithmic phase of growth). The use of liquid cultures in place of agar reduces

92   bacterial clumping. Bacterial cultures were harvested by centrifugation for 10 minutes at 4000g

93   and pellets were re-suspended in 10 ml of distilled water. Bacterial cells were starved for one

94   night in water at room temperature to minimize false positive reactions due to nutrient

95   accumulation in MAH cells and to ensure the use of the substrates provided by the PM plates.

96   The following day the cells were centrifuged and re-suspended using a sterile stick in tubes

97   containing 10 ml of GN/GP-IF-0a (BIOLOG inoculating fluid), 120 µl of 100× BIOLOG Redox

98   Dye Mix G and 1 ml of the appropriate additive (Table 2), until 85% transmittance was reached

99   as measured using the turbidimeter provided by BIOLOG. In order to reduce bacterial clumping,

100   the sterile stick used for inoculation was ground against the wall of the tube. A volume of 100 µl

101   of this final suspension was added to each of the 96 wells of the PM plates. The PM plates were

102   then sealed to avoid drying and incubated at 37°C in the OmniLog® (BIOLOG, Hayward, CA)

103   incubator reader for 8 days.

104   As recommended by BIOLOG, we tested plates PM1 to PM4 using the same assay protocol but

105   without addition of bacteria in order to identify wells with abiotic dye reduction, which can

106   generate false positive results.

107   **Analysis of BIOLOG phenotype microarray data**

108   The raw kinetic data were exported as CSV files using OmniLog PM file Management/kinetic

109   Analysis module (Bochner 2003; Khatri et al. 2013). Differences in the metabolization of the

110   different substrates by the ten MAH isolates were investigated by analyzing the maximum height

111   of the bacterial respiration curves (parameter A) using the R-package opm (Vaas et al. 2013). To

112   allow comparisons across plates processed in different experimental runs, the A parameters were

113   normalized by subtracting the well mean of the negative control (Vaas et al. 2013). Furthermore,

114   the A parameters of the triplicates were combined by calculating the mean and discretized into

115   "positive", "moderate" and "negative" metabolization using the method "discrete" within the R-

116   opm package. Substrates differentiating the isolates from each other were visualized as a

117   heatmap generated using the R-packages heatmap.plus with the Euclidean algorithm. The

118   heatmap displays the utilization of each substrate with a color key: yellow for strong positive

119   metabolization, green for moderate metabolization and blue for no metabolization.

120   **Analysis of metabolic pathways**

121   The metabolic pathways of the two substrates of interest butyric and propionic acid have been

122   further analyzed. Specifically, we extracted all sequences of the genes known to be associated

123   with the pathways related to butyric and propionic acid from the KEGG pathway database

124   (Kanehisa et al. 2016). We extracted the genes from all the *M. avium* subspecies (n=8) present in

125   the KEGG pathway database, namely: *M. avium* subsp. *paratuberculosis* K-10, *M. avium* subsp.

126   *paratuberculosis* MAP4, *M. avium* subsp. *paratuberculosis* E1, *M. avium* subsp.

127   *paratuberculosis* E93, *M. avium* subsp. *avium* DJO-44271, *M. avium* subsp. *avium* 2285 (R), *M.*

128   *avium* subsp. *avium* 2285 (S) and the *M. avium* 104. The redundant genes have been excluded.

129   Then we screened all such genes in genomes of our ten MAH isolates by performing a Custom

130   BLAST analysis using Geneious version 9 (Kearse et al. 2012). The parameters for the screening

131   that we used to determine if a gene was present or not were: sequence identity ≥ 90 %, sequence

132   coverage ≥ 90 %, e value ≤ 0.01.

133   In addition, we analyzed the number of Single Nucleotide polymorphisms (SNP)s (both

134   synonymous and nonsynonymous) in the sequence of the genes detected in our MAH isolates.

135   For each gene, we also constructed a phylogenetic tree using the nucleotide sequences to

136   determine whether any SNP was associated with clinical or environmental source of the isolates

137   based on the Tamura–Nei model using Geneious version 9.

138   **Statistical analyses**

139   We generated two groups, one with data from all clinical isolates and the other with data from all

140   environmental isolates. Statistical differences between clinical and environmental isolates in the

141   metabolization of butyric acid and propionic acid were evaluated by means of 95% family-wise

142   comparison of group means (Tukey contrast test) of the parameter A on specific wells using the

143   function "opm_mcp" within the opm R-package. A p value less than 0.05 was considered to be

144   statistically significant.

145   **Whole genome sequencing of MAH isolates**

146   Genomic DNAs were extracted from the MAH isolates as described previously (Lewin et al.

147   2003). Whole genome sequencing (WGS) was performed using Illumina MiSeq 300 bp paired-

148   end sequencing, yielding a coverage that exceeded 100x. The NGS QC tool kit was used to

149   assess the quality of the data reads, which was set as reads with a minimum of 70 % of bases

150   having a phred score greater than 20 (Patel & Jain 2012). De novo assembly of the resulting

151   reads into multiple contigs was performed using CLC Genomics Workbench 8.0 (CLC bio,

152   Aarhus, Denmark) and contigs annotation was done using RAST (Aziz et al. 2008).

153   **Determination of the maximum common genome and of the accessory genome**

154   We determined the maximum common genome (MCG), comprising those genes present in all of

155   the ten MAH genomes, as reported previously (von Mentzer et al. 2014). All these genes were

156   then extracted from all genomes, concatenated and aligned. The resulting alignment was used to

157   generate a clustering tree using RAxML 8.1 (Stamatakis 2014).

158   For determination of the accessory genome we applied the PanGenome Pipeline – Roary. After

159   determination of the accessory genome of the ten MAH genomes and its distribution within

160   them, we separated those genes that are exclusively present only in either the environmental

161   strains or the clinical strains (Page et al. 2015).

162   **Results**

163   **Substrate utilization of the ten MAH isolates**

164   We tested the capability of our ten MAH isolates to metabolize 379 different substrates. In total,

165   334/379 (88.1 %) substrates were negative for all of the isolates (see Supplemental Table S1). A

166   total of 23/379 (6.1 %) substrates caused abiotic reactions and were excluded from further

167   analysis. A list of false-positive substrates is shown in the Supplemental Table S2. The kinetic

168   curves corresponding to the control plates PM1 to PM4 tested without bacteria are presented in

169   the Supplemental Fig. S1.

170   Only two carbon substrates, the fatty acid derivatives Tween 20 and Tween 40 were strongly

171   positive for all of the ten MAH isolates. The kinetic curves for these substrates reached 250

172   Omnilog units, amongst the highest values recorded in our analysis (see Supplemental Fig. S2

173   for all kinetic curves of the ten MAH isolates). The opm analysis revealed that a total of 20/379

174   (5.3 %) substrates were metabolized differently among the MAH isolates (Table 3). We therefore

175   carried out further analysis using only these substrates. The majority of these 20 substrates were

176   carbon substrates, 15/20 (75.0 %), followed by 3 nitrogen and 2 phosphorous substrates. The

177   heatmap in Fig. 1 shows the utilization of these 20 substrates among the ten MAH isolates. The

178   isolates are grouped according to their substrate utilization. Isolates utilizing similar substrates

179   appear to cluster together.

180   Two major clusters, each composed of five isolates, could be observed. One was rich in

181   environmental isolates (4/5) and the other was rich in clinical isolates (4/5). The substrates

182   predominantly contributing to this clustering were butyric acid and propionic acid and indeed,

183   the Tukey's test revealed that environmental isolates metabolized more strongly butyric acid ($p =$

184   0.0209) and propionic acid ($p = 0.00307$) than clinical isolates with statistical significance.

185   **Metabolic pathways analysis**

186   The propionic and butyric acid are involved in three and one pathway, respectively (Table 3). A

187   total of 151 genes have been identified in the KEGG database associated with all these pathways

188   (Kanehisa et al. 2016). In the supplemental Table S3 we reported the distribution and SNPs

189   analysis of those genes in the MAH genomes. Of the 151 genes, 134 (88.7 %) are present in all

190   the ten MAH. The median gene length was 1099 bp (range 318-2253), whereas the median

191   number of SNPs per gene is 17 (range 1-147). The phylogenetic analysis revealed that none of

192   the SNPs could be associated with the group of the clinical or the group of environmental MAH

193   isolates (see supplemental Table S4). In the propanoate pathway four operons have been

194   identified: *fadAB* associated with the β-oxidation of several fatty acids (DiRusso 1990), *ech8-9*

195   encoding for hydrogenases that play a role in energy conversion (Sant'Anna et al. 2015), *sucCD*

196   responsible for the succinate metabolism (Cerdeno-Tarraga et al. 2003) and *mutAB* involved in

197   the methylmalonate pathway (Schoenwolf & Alvarez 1989). In the nicotinate pathway there are

198   two operons: the *pntAA-AB-B* responsible for the transhydrogenation between NADH and NADP

199   (Anderlund et al. 1999) and *nadABC* involved in the biosynthesis of NAD+ (Vilcheze et al.

200   2010). In the degradation of aromatic compounds pathway we identified the *pcaHGB* operon

201   involved in the β-ketoadipate pathway (Harwood & Parales 1996). In the butanoate pathway we

202   identified the *fadAB* and *ech8-9operons*, the *sdhCDAB* encoding for the succinate dehydrogenase

203   complex involved in the fatty acid metabolism (Nam et al. 2005) and the *ilvBN* responsible for

204   the acetolactate synthesis, a precursor of several amino acids (Keilhauer et al. 1993).

Clustering analysis and determination of the accessory genome

The WGS of the two reference strains MAH 104 and MAH TH135 were already in the GenBank

database and we submitted the remaining genomes at DDBJ/EMBL/GenBank under the

BioProject Number PRJNA299461. The MCG, the maximum number of genes shared by all ten

MAH isolates was 1,658, the alignment of which spanned 1.378 Mbp. The clustering analysis of

the ten MAH isolates is shown in Fig. 2. By comparing the genetic clustering obtained by WGS

with the phenotypic clustering obtained through BIOLOG PM we observed slight differences.

For examples, the isolates MAH E-96-2 and MAH E-82-7, which share identical metabolic

profiles, were genetically more distant from each other. Interestingly, at the genetic level there

was no obvious clustering between the group of clinical and the group of environmental isolates.

The accessory genome is constituted by 4,067 genes. A total of 1688 genes were specific for the

group of clinical isolates (supplemental Table S5). On the other hand, 698 genes were specific

for the group of environmental isolates (supplemental Table S6). We found no genes that were

present in all the clinical and absent in all the environmental isolates, and vice-versa. Among the

most abundant specific genes of the two groups of isolates, there were no known genes

associated with the pathways which involved butyric and propionic acid. However, genes

annotated as hypothetical proteins represented the most abundant specific genes of the two

groups

223   **Discussion**

224   This study represents the first phenotypic analysis of a collection of clinical and environmental

225   MAH isolates using the Biolog PM technology. We showed that the PM technology works well

226   and can be performed with MAH isolates. Strong positive reactions with several substrates were

227   observed with kinetic curves exceeding 200 Omnilog dye units. Although some substrates were

228   metabolized only moderately by our MAH isolates (green in Fig. 1), this might be due to the fact

229   that the use of such substrates by bacteria has a time lag.

230   The ten MAH isolates showed different metabolic patterns pointing to high intra-species

231   diversity. Only two out of the ten isolates had identical heatmap profiles (MAH E-96-2 and

232   MAH E-82-7).

233   Our study showed that MAH isolates prefer to metabolize fatty acids as a carbon source. Indeed,

234   the Tween substrates were strongly metabolized by all MAH isolates tested. This is in agreement

235   with prior studies, showing that Tween substrates were widely used by different mycobacterial

236   species (Baloni et al. 2014; Chen et al. 2012; Hayashi et al. 2010; Khatri et al. 2013; Lofthouse et

237   al. 2013; Wang et al. 2011). It has been reported that mycobacteria hydrolyze Tween 80 to

238   generate the fatty acid oleic acid, which can enter the Tricarboxylic acid (TCA) cycle or can be

239   used as a substrate for energy production (Lofthouse et al. 2013; Vandal et al. 2009). Other fatty

240   acids used by the majority of our MAH isolates are represented by two short fatty acids, caproic

241   acid and butyric acid (Kanehisa & Goto 2000; Kanehisa et al. 2016; Khatri et al. 2013). Caproic

242   acid and its derivatives are involved in several mycobacterial pathways such as the degradation

243   of aromatic compounds, oxocarboxylic acid metabolism or lysine degradation (Kanehisa & Goto

244   2000; Kanehisa et al. 2016). The butyric acid is the final product of butanoate metabolism.

245   Propionic acid is another fatty acid used by our MAH isolates and this represents the terminal

246   product of propanoate metabolism (Kanehisa & Goto 2000; Kanehisa et al. 2016). The nitrogen

247   source L-cysteine, used by six of our MAH isolates, is the final product of cysteine metabolism

248   and is involved in the biosynthesis of other amino acids such as methionine and histidine (Baloni

249   et al. 2014; Kanehisa & Goto 2000; Kanehisa et al. 2016).

250   The question of whether bacteria of the same species originating from either clinical or

251   environmental sources differ from each other is still a matter of discussion. Li and co-authors

252   showed that comparative genome analysis clearly distinguished clinical and environmental

253   *Vibrio parahaemolyticus* isolates from each other (Li et al. 2014). In contrast, other researchers

254   have reported no difference between clinical and environmental *Pseudomonas aeruginosa*

255   isolates with regard to virulence and metabolic properties (Alonso et al. 1999; Vives-Florez &

256   Garnica 2006). Although our study did not reveal any clear distinction between clinical or

257   environmental MAH isolates at the level of the whole genome, we observed differences between

258   clinical and environmental isolates with regard to substrate utilization. The most intriguing

259   difference is that the two fatty acids butyric acid and propionic acid are metabolized more by the

260   environmental than by clinical isolates.

261   We observed no difference in the presence / absence of genes associated with butyric or

262   propionic acid pathways among the group of clinical and the group of environmental MAH

263   isolates. The SNPs analysis of the genes involved in the pathways revealed that no SNPs were

264   associated with clinical or environmental origin of the MAH isolates. These evidences suggest

265   that the metabolic differences observed among clinical and environmental MAH isolates might

266   be due to difference in gene regulation. However, we screened all the genes that up to now have

267   been associated with the pathways of interest. We can speculate that there might be additional

268   genes, of unknown function, that might play a role in the above pathways.

269   The analysis of the accessory genome revealed that none of the genes specific for the clinical or

270   for the environmental isolates could be associated with the pathways of interest. However, future

271   studies on the high number of hypothetical proteins might clarify whether they have a role in the

272   pathways which involved the butyric and propionic acid.

273   The higher metabolic activity observed among environmental MAH isolates might be

274   advantageous for survival in an environment presenting a wider range of nutritional conditions

275   than the host cells alone. Further studies testing a larger number of isolates from different origins

276   might clarify this. In addition, it has been showed that in bacteria the fatty acids have a role in

277   adaptation to different environmental conditions (de Sarrau et al. 2012; de Sarrau et al. 2013;

278   Diomande et al. 2015).

279   **Conclusions**

280   Our study contributes to the understanding of the emerging pathogen MAH at the phenotypic and

281   metabolic level. Understanding how bacteria utilize their own or host-derived substrates during

282   infection might help the development of strategies to fight such infections. We encourage

283   phenotypic testing of microbial isolates from different ecological niches to identify key

284   substrates or pathways that can be used as targets for drug development or for selective growth

285   media development.

286

287 **Acknowledgments**

288 We would like to thank Barry Bochner (President of BIOLOG) for his invaluable input to this

289 study and Brian Weinrick (Albert Einstein College of Medicine, New York City) for his support

290 in developing the BIOLOG laboratory protocol. Elvira Richter (National Reference Center for

291 Mycobacteria, Borstel, Germany) and Roland Schulze-Röbbecke (University Hospital

292 Düsseldorf) provided a number of MAH isolates, Carsten Schwarz (Christiane Herzog Zentrum,

293 Charité, Berlin) provided respiratory samples from cystic fibrosis patients and Kei-ichi Uchiya

294 provided the reference strain MAH TH135. We thank Katharina Schaufler (Free University

295 Berlin) for her support with the BIOLOG data analysis and Inga Eichorn (Free University Berlin)

296 for her help with the whole genome sequencing data. We thank Steve Norley (Robert Koch

297 Institute, Berlin) for the English revision of the manuscript.

298

299

300 **References**

301 Alonso A, Rojo F, and Martinez JL. 1999. Environmental and clinical isolates of Pseudomonas
302   aeruginosa show pathogenic and biodegradative properties irrespective of their origin.
303   *Environ Microbiol* 1:421-430.
304 Anderlund M, Nissen TL, Nielsen J, Villadsen J, Rydstrom J, Hahn-Hagerdal B, and Kielland-
305   Brandt MC. 1999. Expression of the Escherichia coli pntA and pntB genes, encoding
306   nicotinamide nucleotide transhydrogenase, in Saccharomyces cerevisiae and its effect on
307   product formation during anaerobic glucose fermentation. *Appl Environ Microbiol*
308   65:2333-2340.
309 Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass
310   EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK,
311   Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O,
312   Vonstein V, Wilke A, and Zagnitko O. 2008. The RAST Server: rapid annotations using
313   subsystems technology. *BMC Genomics* 9:75. 10.1186/1471-2164-9-75
314 Baloni P, Padiadpu J, Singh A, Gupta KR, and Chandra N. 2014. Identifying feasible metabolic
315   routes in Mycobacterium smegmatis and possible alterations under diverse nutrient
316   conditions. *BMC Microbiol* 14:276. 10.1186/s12866-014-0276-5
317 Bannantine JP, Bayles DO, Robbe-Austerman S, Burrell AM, and Stabel JR. 2014. Draft
318   Genome Sequence of a Mycobacterium avium Complex Isolate from a Broadbill Bird.
319   *Genome Announc* 2. 10.1128/genomeA.01268-13
320 Bochner BR. 2003. New technologies to assess genotype-phenotype relationships. *Nat Rev Genet*
321   4:309-314. 10.1038/nrg1046
322 Bochner BR. 2009. Global phenotypic characterization of bacteria. *FEMS Microbiol Rev* 33:191-
323   205. 10.1111/j.1574-6976.2008.00149.x
324 Bochner BR, Gadzinski P, and Panomitros E. 2001. Phenotype microarrays for high-throughput
325   phenotypic testing and assay of gene function. *Genome Res* 11:1246-1255.
326   10.1101/gr.186501
327 Bochner BR, Giovannetti L, and Viti C. 2008. Important discoveries from analysing bacterial
328   phenotypes. *Mol Microbiol* 70:274-280. 10.1111/j.1365-2958.2008.06383.x
329 Borglin S, Joyner D, DeAngelis KM, Khudyakov J, D'Haeseleer P, Joachimiak MP, and Hazen
330   T. 2012. Application of phenotypic microarrays to environmental microbiology. *Curr*
331   *Opin Biotechnol* 23:41-48. 10.1016/j.copbio.2011.12.006
332 Cerdeno-Tarraga AM, Efstratiou A, Dover LG, Holden MT, Pallen M, Bentley SD, Besra GS,
333   Churcher C, James KD, De Zoysa A, Chillingworth T, Cronin A, Dowd L, Feltwell T,
334   Hamlin N, Holroyd S, Jagels K, Moule S, Quail MA, Rabbinowitsch E, Rutherford KM,
335   Thomson NR, Unwin L, Whitehead S, Barrell BG, and Parkhill J. 2003. The complete
336   genome sequence and analysis of Corynebacterium diphtheriae NCTC13129. *Nucleic*
337   *Acids Res* 31:6516-6523.
338 Chen JW, Scaria J, and Chang YF. 2012. Phenotypic and transcriptomic response of auxotrophic
339   Mycobacterium avium subsp. paratuberculosis leuD mutant under environmental stress.
340   *PLoS One* 7:e37884. 10.1371/journal.pone.0037884
341 de Sarrau B, Clavel T, Clerte C, Carlin F, Ginies C, and Nguyen-The C. 2012. Influence of
342   anaerobiosis and low temperature on Bacillus cereus growth, metabolism, and membrane
343   properties. *Appl Environ Microbiol* 78:1715-1723. 10.1128/AEM.06410-11

344   de Sarrau B, Clavel T, Zwickel N, Despres J, Dupont S, Beney L, Tourdot-Marechal R, and
345   Nguyen-The C. 2013. Unsaturated fatty acids from food and in the growth medium
346   improve growth of Bacillus cereus under cold and anaerobic conditions. *Food Microbiol*
347   36:113-122. 10.1016/j.fm.2013.04.008
348   Despierres L, Cohen-Bacrie S, Richet H, and Drancourt M. 2012. Diversity of Mycobacterium
349   avium subsp. hominissuis mycobacteria causing lymphadenitis, France. *Eur J Clin*
350   *Microbiol Infect Dis* 31:1373-1379. 10.1007/s10096-011-1452-2
351   Diomande SE, Nguyen-The C, Guinebretiere MH, Broussolle V, and Brillard J. 2015. Role of
352   fatty acids in Bacillus environmental adaptation. *Front Microbiol* 6:813.
353   10.3389/fmicb.2015.00813
354   DiRusso CC. 1990. Primary sequence of the Escherichia coli fadBA operon, encoding the fatty
355   acid-oxidizing multienzyme complex, indicates a high degree of homology to eucaryotic
356   enzymes. *J Bacteriol* 172:6459-6468.
357   Falkinham JO, 3rd. 2013. Ecology of nontuberculous mycobacteria--where do human infections
358   come from? *Semin Respir Crit Care Med* 34:95-102. 10.1055/s-0033-1333568
359   Gupta KR, Kasetty S, and Chatterji D. 2015. Novel functions of (p)ppGpp and Cyclic di-GMP in
360   mycobacterial physiology revealed by phenotype microarray analysis of wild-type and
361   isogenic strains of Mycobacterium smegmatis. *Appl Environ Microbiol* 81:2571-2578.
362   10.1128/AEM.03999-14
363   Harwood CS, and Parales RE. 1996. The beta-ketoadipate pathway and the biology of self-
364   identity. *Annu Rev Microbiol* 50:553-590. 10.1146/annurev.micro.50.1.553
365   Hayashi D, Takii T, Mukai T, Makino M, Yasuda E, Horita Y, Yamamoto R, Fujiwara A, Kanai
366   K, Kondo M, Kawarazaki A, Yano I, Yamamoto S, and Onozaki K. 2010. Biochemical
367   characteristics among Mycobacterium bovis BCG substrains. *FEMS Microbiol Lett*
368   306:103-109. 10.1111/j.1574-6968.2010.01947.x
369   Hoefsloot W, van Ingen J, Andrejak C, Angeby K, Bauriaud R, Bemer P, Beylis N, Boeree MJ,
370   Cacho J, Chihota V, Chimara E, Churchyard G, Cias R, Daza R, Daley CL, Dekhuijzen
371   PN, Domingo D, Drobniewski F, Esteban J, Fauville-Dufaux M, Folkvardsen DB,
372   Gibbons N, Gomez-Mampaso E, Gonzalez R, Hoffmann H, Hsueh PR, Indra A, Jagielski
373   T, Jamieson F, Jankovic M, Jong E, Keane J, Koh WJ, Lange B, Leao S, Macedo R,
374   Mannsaker T, Marras TK, Maugein J, Milburn HJ, Mlinko T, Morcillo N, Morimoto K,
375   Papaventsis D, Palenque E, Paez-Pena M, Piersimoni C, Polanova M, Rastogi N, Richter
376   E, Ruiz-Serrano MJ, Silva A, da Silva MP, Simsek H, van Soolingen D, Szabo N,
377   Thomson R, Tortola Fernandez T, Tortoli E, Totten SE, Tyrrell G, Vasankari T, Villar M,
378   Walkiewicz R, Winthrop KL, Wagner D, and Nontuberculous Mycobacteria Network
379   European Trials G. 2013. The geographic diversity of nontuberculous mycobacteria
380   isolated from pulmonary samples: an NTM-NET collaborative study. *Eur Respir J*
381   42:1604-1613. 10.1183/09031936.00149212
382   Johnson DA, Tetu SG, Phillippy K, Chen J, Ren Q, and Paulsen IT. 2008. High-throughput
383   phenotypic characterization of Pseudomonas aeruginosa membrane transport genes. *PLoS*
384   *Genet* 4:e1000211. 10.1371/journal.pgen.1000211
385   Kanehisa M, and Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic*
386   *Acids Res* 28:27-30.
387   Kanehisa M, Sato Y, Kawashima M, Furumichi M, and Tanabe M. 2016. KEGG as a reference
388   resource for gene and protein annotation. *Nucleic Acids Res* 44:D457-462.
389   10.1093/nar/gkv1070

390 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
391   Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, and Drummond A. 2012.
392   Geneious Basic: an integrated and extendable desktop software platform for the
393   organization and analysis of sequence data. *Bioinformatics* 28:1647-1649.
394   10.1093/bioinformatics/bts199
395 Keilhauer C, Eggeling L, and Sahm H. 1993. Isoleucine synthesis in Corynebacterium
396   glutamicum: molecular analysis of the ilvB-ilvN-ilvC operon. *J Bacteriol* 175:5595-
397   5603.
398 Khatri B, Fielder M, Jones G, Newell W, Abu-Oun M, and Wheeler PR. 2013. High throughput
399   phenotypic analysis of Mycobacterium tuberculosis and Mycobacterium bovis strains'
400   metabolism using biolog phenotype microarrays. *PLoS One* 8:e52673.
401   10.1371/journal.pone.0052673
402 Kim BJ, Choi BS, Lim JS, Choi IY, Lee JH, Chun J, Kook YH, and Kim BJ. 2012. Complete
403   genome sequence of Mycobacterium intracellulare strain ATCC 13950(T). *J Bacteriol*
404   194:2750. 10.1128/JB.00295-12
405 Lahiri A, Kneisel J, Kloster I, Kamal E, and Lewin A. 2014. Abundance of Mycobacterium
406   avium ssp. hominissuis in soil and dust in Germany - implications for the infection route.
407   *Lett Appl Microbiol* 59:65-70. 10.1111/lam.12243
408 Lewin A, Freytag B, Meister B, Sharbati-Tehrani S, Schafer H, and Appel B. 2003. Use of a
409   quantitative TaqMan-PCR for the fast quantification of mycobacteria in broth culture,
410   eukaryotic cell culture and tissue. *J Vet Med B Infect Dis Vet Public Health* 50:505-509.
411 Li L, Wong HC, Nong W, Cheung MK, Law PT, Kam KM, and Kwan HS. 2014. Comparative
412   genomic analysis of clinical and environmental strains provides insight into the
413   pathogenicity and evolution of Vibrio parahaemolyticus. *BMC Genomics* 15:1135.
414   10.1186/1471-2164-15-1135
415 Lofthouse EK, Wheeler PR, Beste DJ, Khatri BL, Wu H, Mendum TA, Kierzek AM, and
416   McFadden J. 2013. Systems-based approaches to probing metabolic variation within the
417   Mycobacterium tuberculosis complex. *PLoS One* 8:e75913.
418   10.1371/journal.pone.0075913
419 Mackie AM, Hassan KA, Paulsen IT, and Tetu SG. 2014. Biolog Phenotype Microarrays for
420   phenotypic characterization of microbial cells. *Methods Mol Biol* 1096:123-130.
421   10.1007/978-1-62703-712-9_10
422 Mishra MN, and Daniels L. 2013. Characterization of the MSMEG_2631 gene (mmp) encoding
423   a multidrug and toxic compound extrusion (MATE) family protein in Mycobacterium
424   smegmatis and exploration of its polyspecific nature using biolog phenotype microarray.
425   *J Bacteriol* 195:1610-1621. 10.1128/JB.01724-12
426 Nai C, Wong HY, Pannenbecker A, Broughton WJ, Benoit I, de Vries RP, Gueidan C, and
427   Gorbushina AA. 2013. Nutritional physiology of a rock-inhabiting, model microcolonial
428   fungus from an ancestral lineage of the Chaetothyriales (Ascomycetes). *Fungal Genet
429   Biol* 56:54-66. 10.1016/j.fgb.2013.04.001
430 Nam TW, Park YH, Jeong HJ, Ryu S, and Seok YJ. 2005. Glucose repression of the Escherichia
431   coli sdhCDAB operon, revisited: regulation by the CRP*cAMP complex. *Nucleic Acids
432   Res* 33:6712-6722. 10.1093/nar/gki978
433 Omsland A, Cockrell DC, Howe D, Fischer ER, Virtaneva K, Sturdevant DE, Porcella SF, and
434   Heinzen RA. 2009. Host cell-free growth of the Q fever bacterium Coxiella burnetii.
435   *Proc Natl Acad Sci U S A* 106:4430-4434. 10.1073/pnas.0812074106

436 Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane
437   JA, and Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis.
438   *Bioinformatics* 31:3691-3693. 10.1093/bioinformatics/btv421
439 Patel RK, and Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation
440   sequencing data. *PLoS One* 7:e30619. 10.1371/journal.pone.0030619
441 Plata G, Henry CS, and Vitkup D. 2015. Long-term phenotypic evolution of bacteria. *Nature*
442   517:369-372. 10.1038/nature13827
443 Rindi L, and Garzelli C. 2014. Genetic diversity and phylogeny of Mycobacterium avium. *Infect*
444   *Genet Evol* 21:375-383. 10.1016/j.meegid.2013.12.007
445 Sant'Anna FH, Lebedinsky AV, Sokolova TG, Robb FT, and Gonzalez JM. 2015. Analysis of
446   three genomes within the thermophilic bacterial species Caldanaerobacter subterraneus
447   with a focus on carbon monoxide dehydrogenase evolution and hydrolase diversity. *BMC*
448   *Genomics* 16:757. 10.1186/s12864-015-1955-9
449 Schoenwolf GC, and Alvarez IS. 1989. Roles of neuroepithelial cell rearrangement and division
450   in shaping of the avian neural plate. *Development* 106:427-439.
451 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
452   large phylogenies. *Bioinformatics* 30:1312-1313. 10.1093/bioinformatics/btu033
453 Tohsato Y, and Mori H. 2008. Phenotype profiling of single gene deletion mutants of E. coli
454   using Biolog technology. *Genome Inform* 21:42-52.
455 Tortoli E. 2014. Microbiological features and clinical relevance of new species of the genus
456   Mycobacterium. *Clin Microbiol Rev* 27:727-752. 10.1128/CMR.00035-14
457 Uchiya K, Takahashi H, Yagi T, Moriyama M, Inagaki T, Ichikawa K, Nakagawa T, Nikai T,
458   and Ogawa K. 2013. Comparative genome analysis of Mycobacterium avium revealed
459   genetic diversity in strains that cause pulmonary and disseminated disease. *PLoS One*
460   8:e71831. 10.1371/journal.pone.0071831
461 Vaas LA, Sikorski J, Hofner B, Fiebig A, Buddruhs N, Klenk HP, and Goker M. 2013. opm: an
462   R package for analysing OmniLog(R) phenotype microarray data. *Bioinformatics*
463   29:1823-1824. 10.1093/bioinformatics/btt291
464 Vandal OH, Nathan CF, and Ehrt S. 2009. Acid resistance in Mycobacterium tuberculosis. *J*
465   *Bacteriol* 191:4714-4721. 10.1128/JB.00305-09
466 Vilcheze C, Weinrick B, Wong KW, Chen B, and Jacobs WR, Jr. 2010. NAD+ auxotrophy is
467   bacteriocidal for the tubercle bacilli. *Mol Microbiol* 76:365-377. 10.1111/j.1365-
468   2958.2010.07099.x
469 Vives-Florez M, and Garnica D. 2006. Comparison of virulence between clinical and
470   environmental Pseudomonas aeruginosa isolates. *Int Microbiol* 9:247-252.
471 von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, Rasko DA, Joffre
472   E, Corander J, Pickard D, Wiklund G, Svennerholm AM, Sjoling A, and Dougan G.
473   2014. Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term
474   global distribution. *Nat Genet* 46:1321-1326. 10.1038/ng.3145
475 Wang C, Mahrous EA, Lee RE, Vestling MM, and Takayama K. 2011. Novel Polyoxyethylene-
476   Containing Glycolipids Are Synthesized in Corynebacterium matruchotii and
477   Mycobacterium smegmatis Cultured in the Presence of Tween 80. *J Lipids* 2011:676535.
478   10.1155/2011/676535
479 Wynne JW, Seemann T, Bulach DM, Coutts SA, Talaat AM, and Michalski WP. 2010.
480   Resequencing the Mycobacterium avium subsp. paratuberculosis K10 genome: improved

481   annotation and revised genome sequence. *J Bacteriol* 192:6319-6320. 10.1128/JB.00972-
482   10

483

Characteristics of the ten MAH isolates analyzed in this study

1   **Table 1** Characteristics of the ten MAH isolates analyzed in this study.

MAH Isolate name	Year of isolation	Source	Provider or reference	Accession of whole genome sequence
P-10091-06	2006	Clinical - Child with lymphadenitis	NRC for Mycobacteria, Borstel, Germany	LNAV00000000
2721	2004	Clinical - Child with lymphadenitis	NRC for Mycobacteria, Borstel, Germany	AWXJ00000000
P-9-13	2013	Clinical - Adult pulmonary infection	Charité Hospital, Berlin, Germany	LNBB00000000
104	1983	Clinical - Adult pulmonary infection	Reference strain, USA	CP000479
TH135	2013	Clinical - Adult pulmonary infection	Reference strain, Japan	AP012555
E-128	2010	Environmental - Soil	Friedrich Löffler Institute, Jena, Germany	LVCS00000000
E-96-2	2010	Environmental - Soil	This study	LMVW00000000
E- 82-7	2010	Environmental - Dust	This study	LNAF00000000
27-1	2010	Environmental - Dust	This study	AWXK00000000
E-2514	na	Environmental - Water	University of Düsseldorf, Germany	LNBJ00000000

2   MAH: *Mycobacterium avium* subsp. *hominissuis*; NRC: National reference center; na: not available

Additives used for each PM plates

As additive are usually provided nutrient that are absent to the PM minimal media, but present in a standard MAH growth conditions. We used additives to make a complete minimal medium but omitted anything that could act as a source of the substrates of interest (for example, we did not include nitrate additives in the nitrogen source plates).

1 **Table 2** Additives used for each PM plates. As additive are usually provided nutrient that are absent to the PM minimal media, but present in a
2 standard MAH growth conditions. We used additives to make a complete minimal medium but omitted anything that could act as a source of the
3 substrates of interest (for example, we did not include nitrate additives in the nitrogen source plates).

	Additive a	Additive b
PM plate usage	**PM1, PM2, PM4**	**PM3**
Ingredients	24mM MgCl2	24mM MgCl2
	12mM CaCl2	12mM CaCl2
	0,0012% ZnSO4	0,0012% ZnSO4
	0,06% ferric ammonium citrate	0,01% tween 80
	1,2% NH4Cl	
	0,01% tween 80	

The 20 substrates differentiating the ten MAH isolates analyzed in this study.

1 **Table 3** The 20 substrates differentiating the ten MAH isolates analyzed in this study.

PM Plate	Substrate and well number	Pathway involved	Reference
PM1 Carbon	Acetic acid – C08	Pyruvate metabolism	(Baloni *et al.* 2014; Nai *et al.* 2013)
	Acetoacetic acid – G07	Pyruvate metabolism	(Baloni *et al.* 2014; Nai *et al.* 2013)
	Methyl pyruvate – G10	Pyruvate metabolism	(Baloni *et al.* 2014; Nai *et al.* 2013)
	Mono−methyl Succinate – G09	Tricarboxylic acid cycle	(Baloni *et al.* 2014; Nai *et al.* 2013)
	Propionic acid – F07	Propanoate metabolism, Nicotinate and nicotinamide metabolism, Degradation of aromatic compounds	(Baloni, *et al.* 2014; Kanehisa and Goto 2000; Kanehisa *et al.* 2016; Nai *et al.* 2013)
	D−psicose – H05	Glycolysis and branches	(Baloni *et al.* 2014; Nai *et al.* 2013)
	Pyruvic acid – H08	Pyruvate metabolism	(Baloni *et al.* 2014; Nai *et al.* 2013)
	Tween 80 – E05	Fatty acid metabolism	(Baloni *et al.* 2014; Nai *et al.* 2013)
PM2 Carbon	L−alaninamide – G02	Amino acid metabolism	(Nai *et al.* 2013)
	Butyric acid – D12	Butanoate metabolism	(Baloni, *et al.* 2014; Kanehisa and Goto 2000; Kanehisa *et al.* 2016; Nai *et al.* 2013)
	Caproic acid – E02	Carboxylic acid metabolism	(Nai *et al.* 2013)
	L−histidine – G06	Amino acid metabolism	(Nai *et al.* 2013)
	γ−hydroxy−butyric acid – E09	Succinate metabolism	(Breitkreuz *et al.* 2003; Nai *et al.* 2013)
	β−methyl−D−galactoside – C07	Galactose Metabolism	(Nai *et al.* 2013)
	Sebacic acid – F08	Carboxylic acid metabolism	(Nai *et al.* 2013)
PM3 Nitrogen	D,L−α−amino−caprylic acid – G10	Amino acid metabolism	(Baloni *et al.* 2014)
	L−cysteine – A11	Amino acid metabolism	(Baloni *et al.* 2014)
	D−galactosamine – E09	Amino-sugar pathway	(Baloni *et al.* 2014)
PM4 Phosphorous and sulphur	Carbamyl phosphate – B05	Urea cycle and Pyrimidine synthesis	(Nelson 2004)
	Sodium pyrophosphate – A03	Phosphoric acid synthesis	(Nelson 2004)

Figure 1

Heatmap showing the 20 substrates that were differently metabolized by the ten MAH isolates analyzed in this study.

The color key scale for each substrate is based on dye reduction quantified by Omnilog units. A yellow color indicates strong positive substrate metabolization, a green color moderate metabolization and a blue color indicates no substrate metabolization. Regarding the MAH isolates, environmental isolates are marked in orange, while clinical isolates are marked in blue.
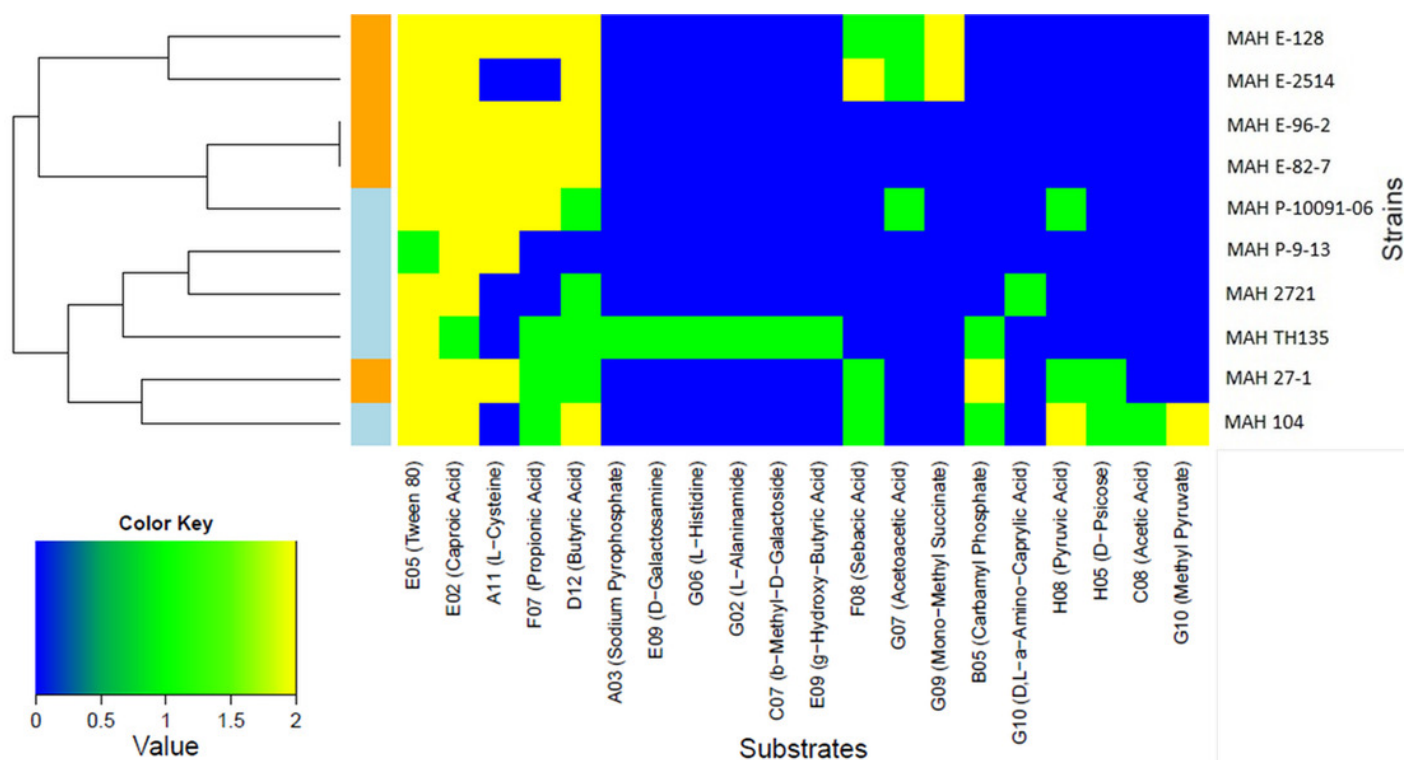
Figure 2

Clustering of the 10 MAH isolates.

The tree was generated using RAxML 8.1. The alignment comprised 1,658 genes constituting the maximum common genome of our ten MAH isolates. Two reference strains were also included (MAH 104 and MAH TH135). The genome sequence of *M. avium* subsp. *paratuberculosis* K10 (Accession Number: AE016958) was used as outgroup. Isolate origin is also represented by blue for clinical origin and orange for environmental origin. The percentage of trees in which the associated taxa clustered together is shown adjacent to the branches.