Title:

Comparative Genomics Shows Differences in the Electron Transport and Carbon Metabolic Pathways of *Mycobacterium africanum* relative to *Mycobacterium tuberculosis* and suggests an adaptation to low oxygen tension

Authors and Affiliations:

Boatema Ofori-Anyinam[1,2,3], Abi Janet Riley[2], Tijan Jobarteh[2], Ensa Gitteh[2], Binta Sarr[2], Tutty Isatou Faal-Jawara[2], Leen Rigouts[1,4], Madikay Senghore[2], Aderemi Kehinde[5,6], Nneka Onyejepu[7], Martin Antonio[2,8,9], Bouke C. de Jong[1], Florian Gehre[1,2,10], Conor J. Meehan[1,11]*

[1]Mycobacteriology Unit, Institute of Tropical Medicine, Antwerp, Belgium

[2]Vaccines and Immunity Theme, Medical Research Council Unit, The Gambia, Banjul, The Gambia

[3]Center for Global Health Security and Diplomacy, Ottawa, Canada

[4]Department of Biomedical Sciences, Antwerp University, Antwerp, Belgium

[5]Department of Medical Microbiology & Parasitology, University College Hospital, Ibadan, Nigeria

[6]Department of Medical Microbiology & Parasitology, College of Medicine, University of Ibadan, Ibadan, Nigeria

[7]Center for Tuberculosis Research, Nigeria Institute of Medical Research, Lagos, Nigeria

[8]Division of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom

[9]Medical School, University of Warwick, Coventry, United Kingdom

[10] Bernhard-Nocht Institute for Tropical Medicine, Hamburg, Germany

[11]School of Chemistry and Biosciences, University of Bradford, Bradford, United Kingdom

***Corresponding Author:** Conor J. Meehan, University of Bradford, Richmond Building, Richmond Road, BD7 1DP, UK, c.meehan2@bradford.ac.uk

Summary:

The geographically restricted *M. africanum* lineages (MAF) are primarily found in West Africa, where they account for a significant proportion of tuberculosis. Despite this phenomenon, little is known about the co-evolution of these ancient lineages with West Africans. MAF and *M. tuberculosis* sensu stricto lineages (MTB) differ in their clinical, *in vitro* and *in vivo* characteristics for reasons not fully understood. Therefore, we compared genomes of 289 MAF and 205 MTB clinical isolates from the 6 main human- adapted *M. tuberculosis* complex lineages, for mutations in their Electron Transport Chain and Central Carbon Metabolic pathway in an attempt to explain these metabolic differences. Furthermore, we determined, *in silico*, whether each mutation could affect the function of genes encoding enzymes in these pathways.

We found more potentially deleterious mutations in these pathways in MAF lineages compared to MTB lineages. Additionally, we found that convergent evolution in these pathways occurred between MAF and some MTB lineages.

Generally, our findings show further differences between MAF and MTB lineages that may have contributed to the MAF clinical- and growth phenotype and may indicate potential adaptation of MAF lineages to a distinct ecological niche, that we suggest may include areas characterized by low oxygen tension.

Running title: The *M. africanum* respiratory chain and carbon metabolic pathway

Keywords: Tuberculosis, *Mycobacterium africanum*, *Mycobacterium tuberculosis*, genome, Electron Transport, Carbon Metabolism

52 **1 Introduction**

53 The *Mycobacterium tuberculosis* complex (MTBC) consists of a group of human–adapted ecotypes-
54 *Mycobacterium tuberculosis* sensu stricto, *Mycobacterium africanum*, *Mycobacterium canetti* and
55 animal-adapted ecotypes (1-4). There are seven known MTBC lineages (L) associated with particular
56 geographic regions and adapted to specific human populations. These are the five lineages that
57 make up *M. tuberculosis* sensu stricto (lineages 1-4 and lineage 7) and the two *M. africanum* lineages
58 (lineages 5 and 6). Africa uniquely has a representation of all seven lineages.

59 MTBC strains from the seven lineages differ on average by about 1200 single nucleotide
60 polymorphisms (5), with clear distinction between MAF and MTB lineages (6-8).

61 MTB Lineages 2 and 4 are more widespread geographically, more pathogenic and more
62 transmissible, while MAF Lineages are exclusively found in West Africa and less transmissible (5, 9,
63 10). Clinically, MTB L4 is relatively more virulent than MAF L6 as evidenced by significantly faster
64 progression, in contacts of infectious cases, to active disease (11). MAF lineages are associated with
65 extrapulmonary disease and MAF L6 more commonly causes disease in immunocompromised
66 persons and those with lower Body Mass Index, implying a more opportunistic pathogen (10, 12).
67 Furthermore, MAF L5 and L6 grow markedly slower than MTB and prefer microaerobic growth
68 conditions (13-15). Reasons for these differences are not completely known, although ours and
69 other's previous studies have attempted to explain some of the observations. A study documented
70 non-synonymous SNPs or frameshift mutations in some genes associated with growth attenuation
71 in MAF and higher mutation frequency in genes necessary for transport of sulphur, ions and
72 lipids/fatty acids across the cell membrane (14). Another reported under-expression of, and MAF L6
73 specific mutations in, dormancy regulon genes, a network of genes crucial for the survival of MTB
74 during hypoxia or anaerobiosis (15, 16).

75 Genes that encode proteins involved in nutrient metabolism and respiration, which are closely
76 linked and together govern bacterial growth and survival, and those of unknown function, the
77 conserved hypotheticals, are highest among the 4173 MTB H37Rv genes reported, emphasizing the
78 importance of the nutrient metabolic and respiratory pathways (17).

79 As obligate aerobes, the mycobacteria respire and produce energy from varied nutritional or energy
80 sources, such as carbon sources (18, 19). These enable the bacteria even to maintain metabolism
81 without growth. The nutritional demands of the mycobacteria have been a topic of interest for over

82 100 years. Pioneering work throughout the 20th century elegantly showed that mycobacteria had
83 unique nutritional requirements (20-29). Central to these findings was that different members of
84 the MTBC were supported by different nutritional sources and consistent results of multiple
85 phenotypic studies led to differentiating members of the MTBC based on their nutritional
86 requirements (20, 30). For instance, it was observed that MTB showed eugonic growth on glycerol,
87 while colonies of MAF and *M. bovis* were dysgonic, indicating an inability to properly utilize this
88 carbon source. MAF and *M. bovis* were only able to show luxurious growth in the presence of sodium
89 pyruvate (20, 30). Furthermore, MAF was found to grow small umbilicated colonies, which on
90 paraffin embedded thin sections revealed extension deep into the media, rather than the surface
91 (31).

92 Almost all the energy used by the bacteria is derived from the Central Carbon Metabolic Pathway
93 (32). After nutrients are metabolized through this pathway, reducing equivalents are generated that
94 eventually enter into the Electron Transport Chain for the generation of significant amounts of
95 Adenosine Triphosphate (ATP) (18, 33).

96 Mycobacteria generate ATP via substrate level phosphorylation and oxidative phosphorylation,
97 which produces more ATP through the activity of the F_1-F_0 ATP synthase in the Electron Transport
98 Chain. Substrate level phosphorylation alone is insufficient to support growth of these bacteria (18).

99 Genome sequencing analyses show that the mycobacteria possess a branched respiratory pathway
100 for electron transfer from electron donors to acceptors under different growth conditions.
101 However, it appears that the transfer of electrons to oxygen, which seems to be the most preferred
102 terminal electron acceptor, occurs with little plasticity, given that only two terminal oxidases have
103 been found in mycobacteria, the bioenergetically more efficient *aa3*-type cytochrome c and the less
104 efficient cytochrome *bd*-type menaquinol oxidases (34). During hypoxia or anoxia, mycobacterial
105 growth is inhibited, even in the presence of alternate terminal electron acceptors within the
106 branched respiratory chain such as fumarate and nitrate reductase. However, mycobacteria are still
107 able to adapt and maintain metabolic functions (34).

108 Therefore, an intact Central Carbon Metabolic pathway and Electron Transport Chain are essential
109 to mycobacterial growth and survival.

110 The importance of comparative genomics in unravelling the basis of distinct metabolic phenotypes
111 in the MTBC has been shown. Using molecular genomic approaches, previous authors found that

112 the inability of *M. bovis* to use glycerol and carbohydrates as sole carbon sources and its
113 requirement for pyruvate in growth media was caused by a single nucleotide polymorphism in the
114 *pykA* gene, encoding pyruvate kinase, resulting in a Glu220Asp amino acid substitution and causing
115 the disruption of sugar catabolism (35). This mutation was also found in 3 MAF strains tested in the
116 same analysis. Additionally, the authors showed that a frameshift at codon 191 of the *glpK* gene of
117 the same *M. bovis* strain led to an incomplete coding sequence and the inability to use glycerol,
118 although this *glpK* mutation was not present in all *M. bovis* strains.

119 Therefore, we aimed to investigate the genes involved in central carbon metabolism and respiration
120 in the MTB and MAF lineages using a whole genome sequencing approach coupled with comparative
121 genomics. We find important differences between the MAF and MTB lineages in their energy and
122 nutrient metabolic pathways that likely contributed to the phenotypic differences observed
123 between these lineages.

124

125 **2 Materials and Methods**

126 **2.1 Ethical Statement**

127 The study was conducted within the framework of an intervention trial of Enhanced Case Finding
128 (ECF) in the Greater Banjul Area of The Gambia (Clinicaltrials.gov NCT01660646), piloted in 2012
129 and conducted between 2013 and 2017. This study was carried out in accordance with the
130 recommendations of the Joint Gambia Government/MRC Ethics Committee and the Institute of
131 Tropical Medicine, Antwerp Institutional Review Board. The protocol, including bacterial sub-
132 studies, was approved by the Joint Gambia Government/MRC Ethics Committee and the Institute of
133 Tropical Medicine, Antwerp Institutional Review Board. Nigerian isolates were collected from
134 Southwest Nigeria within the West African Node of Excellence for TB, AIDS and Malaria (WANETAM)
135 with the recommendations of the University of Ibadan and University College Hospital, Ibadan Joint
136 Ethical Review Committees and the Nigerian Institute of Medical Research, Institutional Board (36).
137 All subjects gave written informed consent in accordance with the Declaration of Helsinki and were
138 anonymized.

139

140 ## 2.2 Bacterial isolates

141 In The Gambia, MTB L4 followed by MAF L6 are the most isolated MTBC lineages. For almost a
142 decade, the prevalence of all the MTBC lineages isolated in The Gambia has remained constant at
143 4.3% (L1), 2.5% (L2), 0.8% (L3), 57.2% (L4), 1.0% (L5), and 35.4% (L6) (37).

144 Within the framework of the ECF study conducted in the Greater Banjul Area, we sequenced 280
145 MAF L6 (32%), 3 MAF L5 (0.3%), 19 MTB L1 (2.2%), 36 MTB L2 (4%)(Beijing), 10 MTB L3 (1%) and
146 534 (60.5%) MTB L4 consisting of 85 MTB L4 Cameroon (9.6%), 14 MTB L4 Ghana (1.6%), 224 MTB
147 L4 Haarlem (25.4%), and 211 MTB L4 LAM (23.9%). Given that only 3 MAF L5 were isolated from
148 The Gambia within the period of analysis, we included 6 MAF L5 from Nigeria (Eastern West
149 Africa), to improve the representativeness of this lineage.

150 Of this dataset, we analyzed the whole genome sequences of all 280 MAF L6 strains, the 3 MAF L5
151 isolated from The Gambia and the 6 MAF L5 from Nigeria, resulting in a total 289 MAF strains. For
152 the MTB lineages, we analyzed a total 205 strains consisting of all 19 MTB L1, all 10 MTB L3 and 14
153 MTB L4 Ghana, 35 MTB L2 and a random number of L4 Haarlem (44), Cameroon (36) and LAM
154 (46), while ensuring that all MTB L4 sublineages isolated were represented and isolates from each
155 year of isolation, 2012 to 2014, were included.

156 ## 2.3 DNA Extraction

157 Genomic DNA was extracted from loopfuls of pure MTBC colonies grown on Lowenstein-Jensen
158 media (38) using the Maxwell 16 DNA Purification Kit (Promega). DNA from Nigerian strains was
159 extracted using the Cetyl trimethylammonium bromide (CTAB) method (39).

160 ## 2.4 Whole-genome sequencing

161 Sequencing of MTBC isolates was performed at MicrobesNG, Birmingham; GenoScreen, France;
162 FISABIO, Valencia or the Beijing Genome Institute (BGI), Beijing. Sequencing reads were generated
163 on a HiSeq or Miseq platform (IIIumina). Quality control was performed for each provider to ensure
164 adequate sequencing depth (>30X) and genome coverage (>95% of the H37Rv reference strain).
165 Raw Illumina reads have been deposited in the ENA with accession <to be undertaken on
166 acceptance>.

167

168 ## 2.5 Bioinformatics Analysis

169 ### 2.5.1 Mapping and Variant calling

170 We used Snippy version 3.1 for the analysis of genomes. Briefly, paired-end raw reads of each
171 sample were mapped to the *M. tuberculosis* H37Rv reference genome (GenBank accession number:
172 NC_000962.3) using BWA-MEM 0.7.12 (40). Mapped reads were converted to the SAM/BAM format
173 and sorted using Samtools 1.3 1 (41). Variant calling was done using Freebayes 0.9.20 (42). Variants
174 were called only if ≥10 reads covered variant positions and ≥90% of those reads differed from the
175 reference. Genes were annotated with SnpEff 4.1 (43). Samples were assigned to MTBC lineages
176 based on the classification of Coll and colleagues (44) using the VCF output of snippy and the
177 PhyResSE SNP list (45).

178 ### 2.5.2 Phylogenetic analysis

179 Using the Snippy output folders of all isolates and custom python scripts, a SNP alignment and a
180 count of all excluded invariant sites were created. A maximum-likelihood (ML) phylogeny was
181 inferred using RAxML version 8.2.9 (46) executing a thousand rapid bootstrap inferences under the
182 general time-reversible (GTR) model, with ascertainment bias correction using the Stamatakis
183 reconstituted DNA approach (47, 48). The resulting tree was exported to interactive Tree of Life
184 (iTOL) for visualization (49).

185 ### 2.5.3 Protein Predictions

186 The effect of amino acid substitutions were predicted using PROVEAN software tools at default
187 settings (PROVEAN protein) (50-52). The PROVEAN NCBI non redundant 2012 database is a newer
188 and larger sequence database than SIFT databases (53) but comparable to SIFT in prediction
189 accuracy. A PROVEAN score of ≤ -2.5 implies a potentially deleterious amino acid substitution.

190

191 # 3 Results and Discussion

192 As carbon sources within the niche of the mycobacterium are metabolized through the Central
193 Carbon Metabolic pathway, reducing equivalents in the form of NAD(P)H and $FADH_2$ are generated
194 and fed into the Electron Transport Chain, that turns over more moles of ATP for cellular processes
195 (34) (Figure 1).

196 Given that generalists and specialists differ in the vastness of their ecological niches, we compared
197 the specialist MAF lineages (289 strains) to the generalist MTB lineages (205 strains) for genomic
198 differences in their Electron Transport and Carbon metabolic pathways. We found a large number
199 of potentially deleterious mutations, particularly in MAF lineages (Figure 2 and Figure 4), however,
200 we also observed that convergent evolution in these pathways occurred between some MTB and
201 MAF lineages.

202 Understanding the differences in metabolism and respiration between the MTBC lineages is
203 important because in recent years, mycobacterial central metabolism and respiration have re-
204 emerged as potential targets for TB chemotherapy. The discovery and successful clinical
205 development of the new TB drug, Bedaquiline (TMC207) and the drug candidate, Telacebec (Q203),
206 both with novel targets in the respiratory chain, has renewed interest and fueled research in this
207 area. Interestingly, even the repurposed drug Clofazimine, and another new drug, Delamanid,
208 through their mechanisms of action reportedly interfere with redox cycling and cellular respiration
209 by generating Reactive Intermediates (54-58), strongly suggesting that respiratory inhibitors may
210 offer the next generation of core drugs against mycobacterial diseases (59).

211 The majority of work previously done on the Electron Transport Chain and oxidative
212 phosphorylation in mycobacteria were biochemical studies by Brodie and colleagues (60). However,
213 the use of molecular biology in establishing the genetic basis of enzymatic defects, following the
214 molecular Koch's postulate, revolutionized mycobacteriology. It is now increasingly apparent that
215 genetic and metabolic differences between the MTBC human-adapted lineages have the potential
216 to affect transmission, diagnostics and treatment (5, 61-66). Moreover, niche adaptation is
217 influenced by the metabolic requirements of an organism, a consequence of evolution.

218

219 **3.1 Mutations in genes encoding Central Carbon Metabolic pathway enzymes.**

220

221 We examined mutations in all genes in the glycolytic pathway, the Tricarboxylic Acid cycle and the
222 Methylcitrate Cycle. Genes and enzymes in these pathways contribute directly to the growth and
223 virulence of the MTBC, by generating products that feed directly into the Electron Transport Chain
224 (67). Therefore, deleterious mutations in these genes would have obvious consequences
225 downstream, for energy generation, survival and ultimately, the virulence of members of the MTBC.

226 We found mutations in major genes encoding enzymes in these pathways that together drive central
227 carbon metabolism. MAF lineages had predicted deleterious mutations in *eno*, *pykA*, *mdh*, *pca* and
228 *icl2a* (Figure 2, Figure 6, Table 1 and Supplementary File S1). For MTB lineages, mutations were
229 detected in *pgi*, *pfkB*, *pca*, *icd2*, *icl2a* and *pckA* (Figure 3, Figure 7, Table 1 and Supplementary File
230 S2). In MTB, predicted deleterious mutations were mostly found in L3.

231 **3.1.1 The Glycolytic Pathway**

232 In MAF lineages, all genes involved in the reversible reactions from the start of glycolysis leading to
233 the breakdown of sugars to 2-phosphoglycerate, at the seventh step (Figure 6), were generally
234 conserved. However, at the critical stage of pyruvate metabolism, deleterious mutations were
235 detected in Enolase (*eno*), the gene responsible for the penultimate step of glycolysis, where 2-
236 phosphoglycerate is converted to Phosphoenolpyruvate, in the gene responsible for the irreversible
237 reaction converting Phosphoenolpyruvate to Pyruvate (*pykA*) and in that responsible for the
238 conversion of Pyruvate to Oxaloacetate, Pyruvate carboxylase (*pca*). Of these three genes, it had
239 previously been shown, in *M. bovis* and 3 MAF strains that, the mutation in *pykA* rendered the
240 critical enzyme pyruvate kinase inactive, resulting in a metabolic block at the level of pyruvate
241 biosynthesis (35). We also confirm this mutation in the MAF strains we analyzed in our study (Figure
242 2). However, to the best of our knowledge this is the first report on lineage-specific potentially
243 deleterious mutations in the essential gene *eno*, in MAF L6. *eno* is found on the surface of many
244 pathogenic bacteria and beyond its primary role in glycolysis, aids in tissue remodeling and invasion
245 of host cells (68). This gene was recently reported as a novel target for the 2-aminothiazoles of the
246 aminothiazole (AT) series, that exerted its effect by inhibiting *eno* in MTB H37Rv (L4 strain) and
247 interfering with energy metabolism, glucose metabolism, ATP homeostasis and the metabolism of
248 cell wall precursors. Therefore, the potentially deleterious mutations we detected in *eno* in our
249 analysis of all MAF L6 may affect energy metabolism and potentially contributes to the metabolic
250 block at the level of pyruvate biosynthesis, reducing the fitness of this lineage further.

251 Unlike we found for MAF lineages, MTB lineages, notably, had no deleterious mutations in *eno* and
252 *pykA* (Figure 7), implying a general ability to complete glycolysis, however, mutations were detected
253 in Glucose-6-phosphate isomerase (*pgi*) in L4 LAM and in Phosphofructokinase B (*pfkB*) in MTB L3,
254 at the second and third steps of glycolysis respectively, where Glucose-6-phosphate is converted to
255 Fructose-6-phosphate and subsequently to Fructose-1-6-biphosphate (Figure 7). *pgi* is essential for
256 the *in vitro* growth of *M. tuberculosis*, and *M. smegmatis pgi* mutants are glucose auxotrophs.

257 Moreover, an additional role for *pgi* in cell wall biosynthesis was reported (38, 69-71). Therefore,
258 *pgi* is important and the effect of the deleterious mutations we detected in *pgi* in L4 LAM should be
259 investigated to determine if L4 LAM shows a distinct growth phenotype, such as slower growth or
260 altered colony morphology, compared to other lineages/sublineages. However, the defect in *pfkB*
261 in MTB L3 is unlikely to affect glycolysis given that Phosphofructokinase activity has so far only been
262 associated with *pfkA*. In *M. tuberculosis*, a *pfkA* deletion mutant was neither able to grow on glucose
263 *in vitro* nor to have any detectable Phosphofructokinase activity and the mutant could not be
264 rescued by expressing *pfkB* (33, 72).

265 Interestingly, additional genes, Pyruvate carboxylase (*pca*) in MAF L5 and MTB L3 and
266 Phosphoenolpyruvate carboxylase (*pckA*) in MTB L4 Cameroon, were mutated at the level of
267 pyruvate metabolism (Figure 6 and Figure 7).

268 The genes*, pca* and *pckA* secrete enzymes contribute significantly to the control of metabolic flux to
269 glycolysis, gluconeogenesis and anaplerosis. These genes carry out functions related to cholesterol
270 detoxification and lipogenesis during intracellular survival. Disruption of *pca* and *pckA* lead to severe
271 *in vitro* and intracellular growth defects and *pckA* was shown to be essential for virulence in *M. bovis*
272 (73, 74). Indeed, oxaloacetate, also generated from pyruvate via the activity of Pyruvate carboxylase
273 (*pca*), feeds into the Methylcitrate cycle (75, 76). Therefore, the mutations we identified in these
274 key Central Carbon Metabolic genes (Table 1) could disrupt energy metabolism and growth in the
275 respective lineages they occur in.

276 **3.1.2 The Tricarboxylic Acid Cycle**

277 At the final stage of glycolysis, pyruvate is generated and converted to Acetyl Coenzyme A (Figure
278 1, Figure 6 and Figure 7), that is fed into the critical Tricarboxylic Acid Cycle (TCA), for the release of
279 reducing equivalents into the Electron Transport Chain. Therefore, the ability to complete this cycle,
280 whether by progression through all stages of the cycle or via the crucial glyoxylate by-pass, is
281 essential for the generation of ATP.

282 Overall, genes encoding enzymes of the TCA cycle were largely conserved in MAF and MTB lineages,
283 however, in MTB L1-L3, L4 Haarlem, L4 Ghana and MAF lineages, we found potentially deleterious
284 mutations in Isocitrate lyase (*icl2a*) and Isocitrate dehydrogenase (*icd2*).

285 The genes encoding isocitrate dehydrogenase are reportedly non-essential for growth, based on
286 transposon mutagenesis experiments. This may be because *M. tuberculosis* is able to use the

287 glyoxylate shunt, mediated by Isocitrate lyase, encoded by ICD1 (*icl*) and ICD2 (*aceAa*/*aceAb*), as a
288 shortcut to divert isocitrate, without going through additional steps in the TCA cycle (33, 69, 70).
289 Therefore, the deleterious mutations we found in *icd2* of MTB L3 may have no direct impact on
290 growth. However, further *in vitro* experiments should confirm this.

291 Emphasizing the importance of the glyoxylate shunt, *M. tuberculosis* strains lacking both ICDs are
292 unable to grow on fatty acids *in vitro*, establish and maintain a chronic infection in mice and were
293 said to be the most severely attenuated strains (33, 77). It is not certain whether the frameshift
294 mutation previously found in H37Rv, an L4 strain (78), which we, for the first time to the best of our
295 knowledge, also report in MTB L4 Haarlem and Ghana, MTB L1, L2, L3, MAF L5 and L6 (Table 1), will
296 affect growth in these lineages, however, it is reported that the activity of ICD2 is lower than that
297 of ICD1, and ICD2 does not play a significant role in the glyoxylate pathway (79). Yet, interestingly,
298 we did not detect this frameshift in MTB L4 Cameroon and LAM (Figure 4).

299 As stated earlier, in section 3.1.1, Pyruvate carboxylase (*pca*), that converts pyruvate to
300 oxaloacetate towards the final stage of glycolysis, was mutated in L5, however, at the last step of
301 the TCA cycle, malate dehydrogenase (*mdh*) that converts malate to oxaloacetate, was also mutated
302 (Figure 6 and Table 1), implying that both routes for the production of oxaloacetate in MAF L5 have
303 mutated genes. Given that oxaloacetate feeds not only into the TCA cycle but also into the
304 methylcitrate cycle, impaired function of *mdh* may affect energy metabolism in MAF L5.

305 **3.1.3 Additional Mutations in the Central Carbon Metabolic Pathway**

306 We detected other mutations in genes of the Central Carbon Metabolic pathway that could
307 potentially impact on enzyme function during metabolism even though their PROVEAN scores were
308 slightly above the cutoff score for classifying mutations as potentially deleterious (Supplementary
309 File S5). We found these mutations in *fba*, again in *mdh* and *icd2*, in *korB*, *aceE* and *glpK* in MAF
310 lineages, in *pfkA* in L4 Cameroon and in *pckA* in L4 LAM and MAF L5.
311 Having discussed the role of *pfkA*, *mdh*, *pckA* and *icd2* in sections 3.1.1 and 3.1.2, we now discuss
312 the role of *korB* in the TCA cycle and of *fba*, *glpK* and *aceE*, in the glycolytic pathway.
313 The Ketoglutarate ferredoxin oxidoreductase (*korAB*), mediating the conversion of α-ketoglutarate
314 to succinate, has its mutants showing reduced or no growth at all (33). Defects in *korB* in MAF
315 lineages may produce a similar effect. Fructose biphosphate aldolase (*fba*) that functions at the
316 fourth stage of the glycolytic pathway is a known essential gene for *in vitro* growth. Therefore, a
317 defective gene could also affect the glycolytic process in MAF lineages.

318 In MTB H37Rv, defects in the essential gene *aceE*, encoding Pyruvate decarboxylase E1 Component,
319 required for the generation of Acetyl Coenzyme A that is indispensable for initiation of the TCA
320 cycle- at the last stage of glycolysis, produced a slow growth mutant (70). In an analysis comparing
321 MTB H37Ra and H37Rv, *aceE* was expressed at high levels in virulent H37Rv throughout the course
322 of infection (80). Due to defects in *aceE*, an earlier report described this gene as a potential cause
323 of MAF growth attenuation (14). Therefore, the mutation we detected in our analysis is potentially
324 linked to MAF growth attenuation.
325 During glycolysis, glycerol metabolism is also important and Glycerol kinase (*glpK*) mediates the
326 metabolism of glycerol as it enters into the glycolytic pathway (35). According to an earlier study,
327 *glpK* is a pseudogene in GM041182, the MAF L6 reference strain. Moreover, in 3 other MAF strains
328 tested in that study, *glpK* was non-synonymously mutated (14). In our analysis, all 289 MAF strains
329 we report on, had a non-synonymously mutated *glpK*, confirming results of previous studies.
330 Recently, mutations in *glpK* have been associated with drug tolerance (81), making our findings
331 critical. Notably, another recent study reported slower response to TB first-line treatment and
332 slower sputum smear conversion for *M. africanum* L6 infected patients (82). It will be interestingly
333 to determine if the *glpK* mutations we detected in MAF lineages in our study promote drug
334 tolerance.

335 To summarize our findings on Central Carbon Metabolism, we observe that for MAF lineages,
336 deleterious mutations largely accrue at critical stages of energy metabolism: 1. At the final stage of
337 glycolysis, the level of pyruvate metabolism and the conversion of pyruvate to Acetyl CoA,
338 indispensable for initiation of the TCA cycle, and 2. At the end of the TCA cycle where intermediates
339 feed into the Methylcitrate cycle.

340 Overall, with major blocks in Central Carbon Metabolism in MAF lineages, the number of reducing
341 equivalents produced via the central carbon metabolic pathway for electron transport may be
342 lower.

343 MTB lineages largely had a conserved glycolytic pathway and TCA cycle, however, different MTB
344 lineages have previously been shown to have different growth rates and patterns (83-85). Notably,
345 the growth rate of MTB L3 is reportedly lower compared to other MTB lineages (83, 84). Therefore,
346 in MTB lineages/sublineages where we found more potentially deleterious mutations, particularly
347 in MTB L3, further investigations on the effect of these mutations on energy metabolism need to be

348 carried out, because, slower growth of these lineages may be directly linked to impaired energy
349 metabolism (83, 84).

350

351 **3.2 Mutations in genes encoding Electron Transport Chain enzymes**

352 As reducing equivalents, like NADH, from the Central Carbon Metabolic Pathway deliver electrons
353 into the Electron Transport Chain, NADH dehydrogenases serve as the gateway of the Electron
354 Transport Chain in the MTBC and electrons are transferred from NADH oxidation to quinone
355 reduction and ultimately to ATP synthase for ATP production in greater quantities (Figure 1).
356 Therefore, defects in the Electron Transport Chain are bound to affect the net yield of ATP
357 generated.

358 Relative to MTB lineages, MAF lineages had multiple predicted deleterious mutations in genes
359 encoding key enzymes of the Electron Transport Chain (Figure 4 and Figure 8). Potentially
360 deleterious mutations were detected in *ndhA*, *pruB*, *qcrC*, *ctaB*, *frdA*, *atpHG*, *sdhA* and *ald* (Figure 4,
361 Table 2 and Supplementary File S3). Interestingly, for some MTB lineages, predicted deleterious
362 mutations were also found in *ndhA*, *Rv0249c*, *frdB*, *atpD*, and *nuoDHF* (Figure 5, Table 2 and
363 Supplementary File S4).

364 To receive NADH from central metabolism into the Electron Transport Chain, *M. tuberculosis*
365 possesses two NADH dehydrogenases, NDH-1 and -2. NDH-1 has 14 subunits (*nuo*A-N) while two
366 copies of NDH-2 exist in *M. tuberculosis*, *ndh* and *ndhA*. Between the two NADH dehydrogenases,
367 an essential role of *ndh* for the growth of *M. tuberculosis* was shown and a function of *ndh* in
368 recycling NADH and maintaining an energized membrane was reported (34, 86). Even though *ndhA*
369 is dispensable for growth, it was recently shown that both *ndh* and *ndhA* differentially control
370 oxygen consumption (86, 87). We found a potentially deleterious mutation in *ndhA* of all MAF L5.
371 Given the adaptation of MAF lineages towards a microaerophilic lifestyle, *ndhA* may be redundant
372 in MAF L5 and defects in this gene may have coincided with the adaptation to a microaerophilic
373 lifestyle.

374 For all genomes in our analysis, though, no potentially deleterious mutations were detected in the
375 essential *ndh*. Of note, *M. leprae* has deleted the entire *nuo* operon (NDH-1) and maintained only
376 *ndh* in its reduced genome, emphasizing the significance of this gene for survival of the
377 mycobacteria (88). Missense mutations found in *ndh* in a clade of Lineage 6, were not predicted to

378 be deleterious for function of the enzyme (Table 2). This implies that the function of *ndh* is likely
379 conserved across the MTBC lineages. Clofazimine, the anti-leprosy drug (prodrug), repurposed for
380 the treatment of TB, is activated to exert its killing effect through its reduction by the essential NDH-
381 2, leading to the release of Reactive Oxygen Species (ROS) following reoxidation by oxygen (55, 86).

382 At the second stage of electron transport, electrons are transferred from NADH dehydrogenase
383 (*nuoA-N*, NDH-1) to the essential complex, Succinate dehydrogenase 1 (*Rv0247c-Rv0249c*, SDH-1)
384 during aerobic respiration when oxygen and nutrients are abundant or from NDH-2 (*ndh*/*ndhA*) to
385 SDH-2 (*sdhABCD*) during hypoxia/anaerobiosis and when nutrients are limited (34, 87, 89) (Figure 8
386 and Figure 9). Succinate dehydrogenases are a direct link between the TCA cycle and electron
387 transport and typically reduce succinate to fumarate (90).

388 The only potentially deleterious mutations we detected in the Succinate dehydrogenases were in
389 *sdhA* of SDH2 in a clade of MAF L6 and in SDH1, *Rv0249c*, in a clade of MTB L1 (Figure 5 and Figure
390 9). Given the essential role of SDH1 for growth and survival, where the deletion of SDH1 was shown
391 to impair the rate of respiration through the Electron Transport Chain and to reduce cell viability
392 (87), future studies should determine if electron transport and aerobic respiration are affected in
393 certain MTB L1 strains, as recent studies report slower growth and a lower odds to grow in culture
394 for MTB L1 and the MAF lineages (65, 83).

395 Electrons from Succinate dehydrogenases, move into the quinone pool, ready for transfer to the
396 third and fourth complexes Cytochrome bc1-aa3 (*qcrABC* and *ctaBCDE*), during aerobic respiration,
397 when oxygen is abundant, and to the less efficient cytochrome bd oxidase (*cydABCD*) when oxygen
398 is limited (Figure 8 and Figure 9).

399 The bc1-aa3 complexes are the major respiratory route in mycobacteria under standard aerobic
400 conditions and are essential for growth where they play a key role in oxidative phosphorylation and
401 electron transport that yields more ATP (34, 91), yet in our analysis, we detected potentially
402 deleterious mutations in *qcrC* and *ctaB* of the bc1-aa3 complexes in MAF lineages and intact gene
403 complexes in MTB lineages (Figure 4 and Figure 5). To the best of our knowledge, this is the first
404 work demonstrating that the critical bc1-aa3 complex is mutated in MAF lineages. We submit that
405 this finding is crucial to understanding the difference in energy metabolism, particularly oxidative
406 phosphorylation, between the MTB and MAF lineages, especially because MAF lineages
407 preferentially grow microaerobically (15, 92-95) and significantly underexpress the dormancy
408 regulon required for adaptation to oxygen limitation (15). Notably, the Imidazopyridine amide in

409 Phase 2 clinical trials, Telacebec (Q203), inhibits *qcrB* of Cytochrome bc1, further emphasizing the
410 importance of this complex for the survival of MTB. Therefore, the predicted deleterious mutations
411 we found in *qcrC*, the heme of the cytochrome bc1-complex, in MAF L6 (Figure 4 and Figure 8) likely
412 impairs aerobic respiration and growth in this lineage severely, given that *qcrC*, like *qcrB*, is essential
413 for survival (89).

414 *M. smegmatis* strains that had mutations in the bc1-aa3 complex were significantly growth
415 impaired, confirming the essentiality of the bc1-aa3 respiratory pathway for mycobacterial growth.
416 Taken together, our analysis provides further support for the view that MAF is adapted to a distinct
417 niche, less dependent on aerobic respiration and more adapted to a microaerobic lifestyle (15).
418 Furthermore, mutations in bc1-aa3 complex genes likely contribute to the slow growth phenotype
419 of the MAF lineages. Reasons for this potential niche adaptation and the benefit to the pathogen in
420 its interaction with its host should be investigated further.

421 As electrons are transferred through the various complexes, protons are pumped outside the
422 membrane, generating the proton motive force that produces a membrane potential, keeps the
423 membrane energized and drives the ATP Synthase pump to produce ATP from ADP and Pi, at the
424 final stage of oxidative phosphorylation.

425 Ultimately, with an impaired bc1-aa3 complex, the function of ATP synthase, the terminal electron
426 acceptor during aerobic respiration, will be impaired, resulting in reduced ATP yield overall. This is
427 likely the case in MAF lineages. Moreover, should MAF lineages engage the less efficient cytochrome
428 bd oxidase, net ATP yield will still be reduced. Therefore, we postulate that ATP yield through
429 oxidative phosphorylation in the MAF lineages is lower compared to the MTB lineages.

430 The F_1-F_0 ATP synthase itself, the target for Bedaquiline (96), is rather conserved in the different
431 lineages. However, in MAF L5 and MTB L2 we detected potentially deleterious mutations in *atp*
432 genes (Table 2). The F_1-F_0 ATP synthase operon is encoded by *atpIBEFHAGDC* and is required for
433 survival as all genes in the operon are essential (69, 97). A defective ATP synthase may coincide with
434 the microaerophilic lifestyle of MAF. However, for MTB L2, it is not clear what the advantage is for
435 acquiring mutations in genes encoding ATP synthase. Interestingly though, compared to MTB L4,
436 MTB L2 was found to have a lower growth rate (83).

437 The branched respiratory chain of *the* MTBC permits anaerobic survival. During hypoxic or anaerobic
438 conditions when oxygen is limited, Fumarate reductase (FRD, *frdABCD*) and Nitrate reductase (NR,

439 *narGHJI*) can serve as terminal electron acceptors to maintain the membrane potential. Therefore,
440 these enzymes are critical. Moreover, further studies of *sdh* suggested that FRD could partially
441 compensate for a lack of SDH activity (98). Defects in any of the subunits of *frd* could limit the overall
442 function of the FRD complex. Notably, the attenuated strain H37Ra grown under low-oxygen
443 conditions showed a lag in gene expression of *frdA* and *frdB* (80). The mutations we found in *frdA*
444 and *frdB* in MAF L5 may be related to the adaptation of this lineage to a hypoxic or microaerophilic
445 lifestyle. Therefore, the effect of mutations on gene function in MAF L5 and MTB L4 sublineage,
446 Haarlem, should be determined experimentally.

447 Other key dehydrogenases that contribute to redox balance of NADH for initiation of electron
448 transport were also mutated in the MAF lineages (Figure 4 and Figure 8), indicating potential redox
449 imbalance in MAF lineages with the likelihood to impede electron transport. From our analysis, all
450 MAF lineages had potentially deleterious mutations in Alanine dehydrogenase (*ald*) and all MAF L5
451 had potentially deleterious mutations in Proline dehydrogenase (*pruB*). Proline dehydrogenase is
452 associated with the adaptation to hypoxia, slow growth rate and is essential for growth (69, 97, 99,
453 100). Alanine dehydrogenase has been shown to play a role in redox balance during low oxygen
454 conditions and the downshift of *M. tuberculosis* to the state of nonreplicating persistence (NRP). *ald*
455 mutants had altered NADH/NAD ratios and significant delays in growth resumption after reaeration.
456 Additionally, induction of *ald* rescued the bc1-aa3 complex mutant while its disruption made the
457 growth defect of the mutant worse (100, 101). Some suggestions have been made to administer a
458 combination of *ald* inhibitors, Telacebec and Bedaquiline although the synergy between drugs in
459 such a combination is yet to be shown (102).

460 Given that *ald* and the bc1-aa3 complex were mutated in all MAF in our analysis, MAF lineages are
461 most probably natural mutants of *ald* and the bc1-aa3 complex. It is possible that these mutations
462 in the MAF lineages contribute significantly to their slower growth compared to MTB due to
463 impaired energy production. Moreover, these polymorphisms further support the adaptation of
464 MAF to a hypoxic niche.

465 **3.2.1 Additional Mutations in the Electron Transport Chain**

466 Like we did for the Central Carbon Metabolic Pathway, we report on other mutated genes, *cydB* and
467 *narG*, in the Electron Transport that are noteworthy, given their significance in respiration and the
468 fact that their provean scores were only slightly above the cut off score indicating potentially
469 deleterious mutations (Supplementary File S5).

470 *cydB* was mutated in all L6 and *narG* in all L5 (Figure 4). Interestingly, both are critical components
471 of the mycobacterial Electron Transport Chain, specifically the anaerobic respiratory chain (34).

472 *narGHJI* allows nitrate respiration during oxygen limitation and *narG* is an important part of this
473 operon. A *narG* mutant of *M. bovis* BCG was unable to reduce nitrate under anaerobic conditions.
474 Moreover, SCID mice infected with the *narG* mutant showed no signs of clinical disease even after
475 > 200 days relative to wild type MTB infected mice that showed signs after 50 days and succumbed
476 after 80 days, implying that *narG* mutants are less virulent (103).

477 *cydAB* encodes cytochrome-bd menaquinol oxidase, which functions under microaerophilic
478 conditions and supports aerobic growth during inhibition of the cytochrome bc1 complex (34, 104,
479 105), also mutated in MAF based on our findings. Notably, the respiratory inhibitor, Q203, which
480 targets the primary terminal oxidase, cytochrome bc1-aa3, is bacteriostatic in *M. tuberculosis*
481 because *M. tuberculosis* has the alternate cytochrome bd-type terminal oxidase that, to an extent,
482 serves as a rescue when the bc1-aa3 complex is defective. In contrast, *M. ulcerans* strains from
483 Africa and Australia lack cytochrome bd and rely solely on the bc1-aa3 complex during respiration.
484 Therefore, Q203 was found to be bactericidal against *M. ulcerans* (106). It is very possible that MAF
485 L6 will be comparatively more susceptible to Q203 than MTB, given that the cytochrome bd rescue
486 is also mutated in MAF L6. Indeed, MAF lineages may be generally more susceptible to respiratory
487 inhibitors.

488

489 **3.3 The case of convergence in electron transport and central metabolism pathways**

490 For multiple genes in the Electron Transport Chain and the Central Carbon Metabolic pathway, we
491 observed convergent evolution within the MTBC. In some cases, the same mutation was shared
492 between lineages in the same gene and in other cases different mutations occurred. These
493 mutations may confer a selective advantage and/or contribute to adaptation. In the Central Carbon
494 Metabolic Pathway, we found a high level of convergence (Table 1).

495 In the glycolytic pathway, we detected the same mutations in *ppgk* at the first stage of glycolysis
496 where glucose is converted to Glucose-6-phosphate, in all MTB and MAF lineages (Table 1). In MTB
497 L1, L2, L3, MAF L5 and L6, we found the same mutation in the essential gene, *eno* and in *prpD* that
498 functions in the Methylcitrate cycle. In *glpK*, encoding glycerol kinase for glycerol metabolism, the
499 same mutation was detected in MTB L1 and in all MAF lineages. Additionally, convergence was

500 detected in *pca* in the glycolytic cycle and *korB* in the Tricarboxylic Acid Cycle, in MTB L3 and MAF
501 L5, although different, and in MTB L3 and MAF L6 in *icd2*. Convergence also occurred in MTB L4
502 Cameroon, a clade of L4 LAM and MAF L5 in *pckA*.

503 Notably, the same frameshift was seen in *icl2a* in MTB L1, L2, L3, L4 Haarlem and Ghana and in MAF
504 L5 and L6 (Table 1). However, MTB Cameroon and LAM may have evolved to retain an intact copy
505 of *icl2a*, because no frameshift was detected in these lineages.

506 Only those mutations detected in *pca* in MTB L3 and MAF L5 and those detected in *icl2a* were
507 predicted to be deleterious.

508 In the Electron Transport Chain, the same mutations were found in *gpdA2*, *frdC* and *nuoG* in all MTB
509 and MAF lineages and in *nuoI* in MTB L3 and all MAF lineages. In MTB L1, MTB L3 and MAF L5, we
510 found mutations in *ndhA* and in MTB L4 LAM and a clade of MAF L6, in *sdhA*. In all lineages, we
511 detected the same mutation in *sdhD*.

512 Of these mutations, only those found in *ndhA* were predicted to be deleterious.

513 Overall, more convergence occurred between MTB L1, MTB L3 and the MAF lineages. This is
514 interesting as the MAF lineages, MTB L1 and L3 lineages all reportedly grow relatively slower than
515 the MTB L2 and L4 lineages (14, 65, 83, 84, 107). We postulate that the slower growth rate of these
516 lineages and the convergence we observe in their metabolic pathways, may be linked to their similar
517 migration and dispersal patterns in Africa and Eurasia (108), where L2 and L4 have become widely
518 dispersed, while L5, L6, and L7, had more geographically restricted expansion, adapting to more
519 specific hosts. This may have influenced niche adaptation, where L2 and L4, in line with increased
520 dispersion and range expansion, also increased their replicative/growth capacity and ability to
521 transmit, while the more host restricted lineages maintained a lower replicative/growth potential
522 in line with expansion in situ.

523

524 **3.4 Limitations**

525 Limitations of our analysis include the small sample size of MAF L5 analyzed. In the Gambia and
526 other countries in Western West Africa, the prevalence of L5 is significantly lower than L6 and is less
527 likely to be isolated compared to L6. Over an 8-year period, the prevalence of MTBC lineages isolated
528 in The Gambia remained constant at 1.0% for MAF L5 and 35.4% for MAF L6 (37). In our isolation of

529 the 289 MAF strains included in this study, only 3 from The Gambia were MAF L5. However, to
530 ensure that the mutations we detected in L5 from The Gambia were more representative of the
531 MAF L5 lineage, we included L5 from Nigeria (Eastern West Africa), where the prevalence of L5 is
532 high and L6 is significantly less likely to be isolated (109). Another limitation is that we did not
533 experimentally confirm any of our in silico phenotype predictions.

534

535 **4 Conclusion**

536 In this comparative analysis, we describe genomic differences between the MTB and MAF lineages
537 in genes encoding enzymes of the Electron Transport Chain and the Central Carbon Metabolic
538 pathway, which may explain the differences in the clinical- and *in vitro* phenotype described for the
539 MAF and MTB lineages. Generally, it appears from our analysis that compared to MTB lineages, MAF
540 lineages had the most mutated Central Carbon Metabolic and Electron Transport Pathways, with
541 mutations occurring in critical components of each pathway. The combined effect of a defective
542 Carbon Metabolic Pathway and Electron Transport Chain in MAF lineages, likely contributes to the
543 reduced fitness of the MAF lineages. We speculate that our findings may contribute to 1 - slower
544 growth of MAF lineages, 2 - relative attenuation of the MAF L6 lineage compared to MTB lineages
545 and 3 - host specificity to West Africans. Further validation experiments are needed to test these
546 hypotheses. We also suggest that future studies should investigate the Electron Transport Chain and
547 Central Carbon Metabolic Pathway of the different MTB lineages beyond MTB L4 and H37Rv to
548 determine if the different growth patterns observed for MTB lineages are directly linked to defects
549 in energy generating pathways.

550 It is intriguing that in the different MTBC lineages multiple deleterious mutations occurred in the
551 same gene. This convergence largely found in MTB lineages 1, 2, 3 and the MAF lineages, indicates
552 there may be a selective advantage for this. Interestingly, these lineages, compared to MTB L4
553 strains, have been associated with slower growth and cytokine induction patterns suggestive of
554 immune evasion (8, 14, 65, 83-85, 107, 110, 111).

555 If the potentially deleterious mutations we report on in this analysis affect energy metabolism in
556 the different MTBC lineages, fitness will differ and ultimately the infection and transmission
557 potential of these lineages. Therefore, our findings are not only relevant for TB product

558 development but also for transmission studies and interventions. The literature already provides
559 credence to and evidence for our hypothesis (9, 85, 112, 113).

560 ## 4.1 Future Proposed Investigations

561 ### 4.1.1 Functional Studies

562 Significant genomic differences between MTB and MAF lineages presented in this analysis warrant
563 further studies in order to properly characterize the regulatory network of MAF. Recently, the
564 regulatory network of H37Rv (MTB L4) was characterized yet what pertains to MAF lineages is not
565 clearly defined, even though it has been shown that master regulators like *PhoPR* and *Rv0081* as
566 well as the DosR regulon are underexpressed and/or mutated in MAF (15, 114-116). A single
567 mutation in PhoP, which abrogates ESX1 secretion, explains the complete loss of virulence of H37Ra.
568 Moreover, the PhoPR regulatory system has also been associated with the regulation of
569 components of the respiratory pathway (*ald* and *nuo*) and the glyoxylate pathway (ICL) (116).
570 Interestingly, *ald* and *icl2a* were also mutated in MAF in our analysis. It is possible that the PhoPR
571 system controls other components of the respiratory and central metabolic pathway. However, this
572 is yet to be shown. We present the H37Rv regulatory network showing differentially regulated and
573 or mutated genes in MAF L5 and L6 compared to MTB H37Rv (15, 117), and include the mutations
574 we detected in MAF lineages in the present study (Figure 6, Figure 8 and Figure 10). We suspect that
575 the potentially deleterious mutations we report on will produce growth phenotypes (Table 3).
576 Therefore, we propose functional genomic assays involving insertion/deletion mutagenesis
577 followed subsequently by *in vitro* and *in vivo* characterization of mutants to confirm the contribution
578 each gene makes to the growth and survival of MAF under defined conditions. Such studies on the
579 MAF lineages are limited, but not for MTB, particularly the generalist MTB L4. Recent studies provide
580 data describing functional consequences of synonymous SNPs (118-120), i.e. caution needs to be
581 taken in inferring the relative significance or impact of observed genomic mutations from
582 sequencing data alone. Therefore, further studies to correlate genotype with phenotype are
583 necessary and our findings serve as a prelude to such experimental studies.

584 ### 4.1.2 Investigate Bioenergetics in MAF lineages

585 It is highly plausible that adaptation of MAF lineages to growth under low oxygen (microaerobic)
586 conditions could be a strategy to escape the harmful effects of ROS generated as electrons leak
587 from a defective respiratory chain and react with built up oxygen (121). Moreover, reduced
588 aerobic respiration or oxygen consumption in the MAF lineages could potentially affect the

589 sensitivity of rapid diagnostics like the automated MGIT 960 system, that depends on oxygen
590 consumption to detect growth (122). Interestingly, the percentage of MAF in certain parts of West
591 and Central Africa was reported to have suddenly and sharply declined over the last decade (4,
592 123-126), yet, in evolutionary terms, such a scenario would be unrealistic and could possibly be
593 due to the introduction and use of new diagnostics like the MGIT 960 system. Of note, we recently
594 detected, in a retrospective analysis based on genotyping, that 84% of strains that did not grow in
595 the MGIT 960 system were MAF L6 (Ofori-Anyinam *et al*., unpublished). Since growth and survival,
596 driven by nutrient and energy metabolism, sustain pathogen transmission, and given that the
597 success of drugs and phenotypic diagnostics rely on metabolic properties of the bacteria,
598 understanding metabolism in these bacteria is essential. As far as we know, energy metabolism,
599 including oxidative phosphorylation in MAF has not been investigated and the efficiency of the
600 Electron Transport Chain and the effectiveness of electron transport in the MAF lineages has not
601 been established yet, even though it is known that MAF lineages grow significantly slower than
602 MTB (14) and preferentially grow microaerobically (15). We hypothesize that MAF lineages
603 generate less ATP and are more exposed to ROS than MTB lineages (Table 3). Therefore, we
604 propose studies to investigate energy metabolism in MAF lineages relative to the MTB lineages,
605 including whether MAF lineages are more exposed to ROS during growth.

606 **4.1.3 Host Genetics**

607 One common feature of host specificity is genomic decay. Genomes of specialists usually show signs
608 of genome decay evidenced by gene deletions or gene inactivation via point mutations. A driving
609 force for the accumulation of mutations in some key electron transport and carbon metabolism
610 genes in the specialist/host-restricted MAF lineages could indicate adaptation of the MAF lineages
611 to a specific host, West Africans, or a niche within the host, due to altered or loss of function of
612 these genes. This potential niche adaptation could be driven by a precise feature of the host
613 environment that favors the association between MAF lineages and West Africans such as has been
614 described for some diseases (127). Some findings have been made, including MAF lineage-specific
615 mutations in genes and pseudogenes involved in vitamin B12 and vitamin B3 metabolism, important
616 cofactor biosynthetic pathways for many cellular functions. Unlike MAF though, MTB is fully capable
617 of synthesizing vitamin B12. Therefore, it has been suggested that the mutations in the vitamin B12
618 pathway of the MAF lineages may affect their host range to West Africans. According to a study in
619 the United States, black persons, reportedly have higher levels of vitamin B12 relative to other
620 ethnicities (8, 10, 128-130). Future studies should investigate the molecular mechanisms underlying

621 host specificity. MAF lineages appear to provide a snapshot of the features that have contributed to
622 the virulence of the MTBC.

623

624 5 Supplementary Information

625 Supplementary File S1

626 Supplementary File S2

627 Supplementary File S3

628 Supplementary File S4

629 Supplementary File S5

630

631 6 Author Contributions

632 BO, FG, BdJ and CM designed the study. BO, FG, AR, TJ, BS and CM undertook analyses. BO, BdJ, FG
633 and CM wrote the manuscript. All authors contributed discussion and reviewed the final manuscript.

634

635 7 Competing financial interests

636 The authors declare no competing financial interests.

637

638 8 Funding

639 This work was supported by the European Research Council-INTERRUPTB starting grant nr.311725
640 (to BdJ, FG, CM, LR, BO, MA) and The UK Medical Research Council and the European & Developing
641 Countries Clinical Trials Partnership (EDCTP) Grant No. CB. 2007. 41700.007.

642

643 **9 Acknowledgments**

644 The authors thank all study participants and colleagues for their contribution to the present analysis.

645 **References**

646 1. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new
647 evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proceedings of the national
648 academy of Sciences. 2002;99(6):3684-9.
649 2. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration
650 and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nature genetics.
651 2013;45(10):1176-82.
652 3. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable
653 host-pathogen compatibility in *Mycobacterium tuberculosis*. Proceedings of the National Academy
654 of Sciences of the United States of America. 2006;103(8):2869-73.
655 4. Michel KG, Guislaine R, Serge PD, Sanou A, Antoinette K, Adama O, et al. Spoligotyping of
656 *Mycobacterium africanum*, Burkina Faso. Emerging Infectious Disease journal.
657 2012;18(1):117.
658 5. Coscolla M, Gagneux S. Consequences of genomic diversity in *Mycobacterium tuberculosis*.
659 Semin Immunol. 2014;26.
660 6. Bentley SD, Comas I, Bryant JM, Walker D, Smith NH, Harris SR. The Genome of
661 *Mycobacterium Africanum* West African 2 Reveals a Lineage-Specific Locus and Genome Erosion
662 Common to the M. tuberculosis Complex. PLoS Negl Trop Dis. 2012;6.
663 7. Mostowy S, Onipede A, Gagneux S, Niemann S, Kremer K, Desmond EP, et al. Genomic
664 analysis distinguishes *Mycobacterium africanum*. Journal of clinical microbiology. 2004;42(8):3594-
665 9.
666 8. Winglee K, Manson McGuire A, Maiga M, Abeel T, Shea T, Desjardins CA, et al. Whole
667 Genome Sequencing of *Mycobacterium africanum* Strains from Mali Provides Insights into the
668 Mechanisms of Geographic Restriction. PLoS neglected tropical diseases. 2016;10(1):e0004332.
669 9. Asare P, Asante-Poku A, Prah DA, Borrell S, Osei-Wusu S, Otchere ID, et al. Reduced
670 transmission of *Mycobacterium africanum* compared to *Mycobacterium tuberculosis* in urban
671 West Africa. International Journal of Infectious Diseases. 2018;73:30-42.
672 10. Sharma A, Bloss E, Heilig CM, Click ES. Tuberculosis Caused by *Mycobacterium africanum*,
673 United States, 2004-2013. Emerging infectious diseases. 2016;22(3):396-403.

674 11. De Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM. Progression to Active
675 Tuberculosis, but Not Transmission, Varies by *Mycobacterium tuberculosis* Lineage in The Gambia.
676 J Infect Dis. 2008;198.
677 12. de Jong BC, Hill PC, Brookes RH, Otu JK, Peterson KL, Small PM, et al. *Mycobacterium*
678 *africanum*: a new opportunistic pathogen in HIV infection? Aids. 2005;19(15):1714-5.
679 13. de Jong BC, Adetifa I, Walther B, Hill PC, Antonio M, Ota M, et al. Differences between
680 tuberculosis cases infected with *Mycobacterium africanum*, West African type 2, relative to Euro-
681 American Mycobacterium tuberculosis: an update. FEMS immunology and medical microbiology.
682 2010;58(1):102-5.
683 14. Gehre F, Otu J, DeRiemer K, de Sessions PF, Hibberd ML, Mulders W, et al. Deciphering the
684 growth behaviour of *Mycobacterium africanum*. PLoS neglected tropical diseases.
685 2013;7(5):e2220.
686 15. Ofori-Anyinam B, Dolganov G, Van T, Davis JL, Walter ND, Garcia BJ, et al. Significant under
687 expression of the DosR regulon in *M. tuberculosis* complex lineage 6 in sputum. Tuberculosis.
688 2017;104:58-64.
689 16. Leistikow RL, Morton RA, Bartek IL, Frimpong I, Wagner K, Voskuil MI. The *Mycobacterium*
690 *tuberculosis* DosR regulon assists in metabolic homeostasis and enables rapid recovery from
691 nonrespiring dormancy. Journal of bacteriology. 2010;192(6):1662-70.
692 17. Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList--10 years after. Tuberculosis
693 (Edinburgh, Scotland). 2011;91(1):1-7.
694 18. Cook GM, Berney M, Gebhard S, Heinemann M, Cox RA, Danilchanka O, et al. Physiology of
695 mycobacteria. Advances in microbial physiology. 2009;55:81-319.
696 19. Warner DF. *Mycobacterium tuberculosis* metabolism. Cold Spring Harbor perspectives in
697 medicine.5(4):a021121.
698 20. David HL, Jahan MT, Jumin A, Grandry J, Lehman EH. Numerical taxonomy analysis of
699 *Mycobacterium africanum*. Int J Syst Bacteriol. 1978;28.
700 21. Dubos RJ. The effect of lipids and serum albumin on bacterial growth. The Journal of
701 experimental medicine. 1947;85(1):9.
702 22. Edson N. The intermediary metabolism of the mycobacteria. Bacteriological reviews.
703 1951;15(3):147.
704 23. Goldman DS. ENZYME SYSTEMS IN THE MYCOBACTERIA XV. Initial Steps in the Metabolism
705 of Glycerol. Journal of bacteriology. 1963;86(1):30-7.
706 24. Hunter G. The oxidation of glycerol by mycobacteria. Biochemical Journal. 1953;55(2):320.

707 25. Merrill MH. Studies on Carbon Metabolism of Organisms of the Genus Mycobacterium: II.
708 Utilization of Organic Compounds in a Synthetic Medium. Journal of bacteriology. 1931;21(5):361.
709 26. Pattyn SR, Portaels F, Spanoghe L, Magos J. Further studies on African strains of
710 *Mycobacterium tuberculosis*: comparison with *M. bovis* and *M.microti*. Annales des Societes belges
711 de medecine tropicale, de parasitologie, et de mycologie. 1970;50.
712 27. Tepper BS. Differences in the utilization of glycerol and glucose by *Mycobacterium phlei*.
713 Journal of bacteriology. 1968;95(5):1713-7.
714 28. Youmans AS, Youmans GP. STUDIES ON THE METABOLISM OF MYCOBACTERIUM
715 TUBERCULOSIS IV.: The Effect of Fatty Acids on the Growth of M. tuberculosis var. hominis1.
716 Journal of bacteriology. 1954;67(6):731.
717 29. Youmans GP, Youmans AS. STUDIES ON THE METABOLISM OF MYCOBACTERIUM
718 TUBERCULOSIS I.: The Effect of Carbohydrates and Alcohols on the Growth of *Mycobacterium*
719 *tuberculosis* var. hominis1. Journal of bacteriology. 1953;65(1):92.
720 30. Pattyn SR, Portaels F, Spanoghe L, Magos J. Further studies on African strains of
721 *Mycobacterium tuberculosis*: comparison with M. bovis and M. microti. Annales des Societes
722 belges de medecine tropicale, de parasitologie, et de mycologie. 1970;50(2):211-27.
723 31. de Medeiros D, Castets M, Diop S. [Comparative morphological aspects of *Mycobacterium*
724 *tuberculosis*, classical type, and *Mycobacterium africanus* (preliminary note)]. Bulletin de la Societe
725 medicale d'Afrique noire de langue francaise. 1973;18(4):477-84.
726 32. BCCM Belgian Coordinated Collections of Microorganisms. 2017 [cited 2017 12 May].
727 Available from: http://bccm.belspo.be
728 33. Baughn AD, Rhee KY. Metabolomics of Central Carbon Metabolism in *Mycobacterium*
729 *tuberculosis*. Microbiology spectrum. 2014;2(3).
730 34. Cook GM, Hards K, Vilcheze C, Hartman T, Berney M. Energetics of Respiration and
731 Oxidative Phosphorylation in Mycobacteria. Microbiology spectrum. 2014;2(3).
732 35. Keating LA, Wheeler PR, Mansoor H, Inwald JK. The pyruvate requirement of some
733 members of the *Mycobacterium tuberculosis* complex is due to an inactive pyruvate kinase:
734 implications for in vivo growth. Mol Microbiol. 2005;56.
735 36. Gehre F, Otu J, Kendall L, Forson A, Kwara A, Kudzawu S, et al. The emerging threat of pre-
736 extensively drug-resistant tuberculosis in West Africa: preparing for large-scale tuberculosis
737 research and drug resistance surveillance. BMC Medicine. 2016;14(1):160.

738 37. Gehre F, Kumar S, Kendall L, Ejo M, Secka O, Ofori-Anyinam B, et al. A Mycobacterial
739 Perspective on Tuberculosis in West Africa: Significant Geographical Variation of M. africanum and
740 Other *M. tuberculosis* Complex Lineages. PLoS neglected tropical diseases. 2016;10(3):e0004408.
741 38. Singhal N, Sharma P, Kumar M, Joshi B, Bisht D. Analysis of intracellular expressed proteins
742 of *Mycobacterium tuberculosis* clinical isolates. Proteome Science. 2012;10(1):14.
743 39. Hosek J, Svastova P, Moravkova M, Pavlik I, Bartos M. Methods of mycobacterial DNA
744 isolation from different biological material: a review. Veterinarni medicina. 2006;51(5):180-92.
745 40. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv
746 preprint arXiv:13033997. 2013.
747 41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
748 Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009;25(16):2078-9.
749 42. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv
750 preprint arXiv:12073907. 2012.
751 43. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating
752 and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
753 Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80-92.
754 44. Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, Viveiros M, et al. A robust
755 SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nature communications.
756 2014;5:4812.
757 45. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a
758 Web Tool Delineating *Mycobacterium tuberculosis* Antibiotic Resistance and Lineage from Whole-
759 Genome Sequencing Data. Journal of clinical microbiology. 2015;53(6):1908-14.
760 46. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
761 phylogenies. Bioinformatics (Oxford, England). 2014;30(9):1312-3.
762 47. Leache AD, Banbury BL, Felsenstein J, de Oca AN, Stamatakis A. Short Tree, Long Tree, Right
763 Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. Systematic
764 biology. 2015;64(6):1032-47.
765 48. Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological
766 character data. Systematic biology. 2001;50(6):913-25.
767 49. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display
768 and annotation. Bioinformatics (Oxford, England). 2007;23(1):127-8.
769 50. Choi Y. A fast computation of pairwise sequence alignment scores between a protein and a
770 set of single-locus variants of another protein. Proceedings of the ACM Conference on

771 Bioinformatics, Computational Biology and Biomedicine; Orlando, Florida. 2382989: ACM; 2012. p.
772 414-7.
773 51. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid
774 substitutions and indels. Bioinformatics (Oxford, England). 2015;31(16):2745-7.
775 52. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino
776 Acid Substitutions and Indels. PloS one. 2012;7(10):e46688.
777 53. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on
778 protein function using the SIFT algorithm. Nature protocols. 2009;4(7):1073-81.
779 54. Barry VC, Belton JG, Conalty ML, Denneny JM, Edward DW, O'Sullivan JF, et al. A new series
780 of phenazines (rimino-compounds) with high antituberculosis activity. Nature.
781 1957;179(4568):1013-5.
782 55. Lechartier B, Cole ST. Mode of Action of Clofazimine and Combination Therapy with
783 Benzothiazinones against <span class="named-content genus-species"
784 id="named-content-1">*Mycobacterium tuberculosis*</span&gt. Antimicrobial
785 Agents and Chemotherapy. 2015;59(8):4457.
786 56. Lewis JM, Sloan DJ. The role of delamanid in the treatment of drug-resistant tuberculosis.
787 Therapeutics and clinical risk management. 2015;11:779-91.
788 57. Xavier AS, Lakshmanan M. Delamanid: A new armor in combating drug-resistant
789 tuberculosis. Journal of pharmacology & pharmacotherapeutics. 2014;5(3):222-4.
790 58. Yano T, Kassovska-Bratinova S, Teh JS, Winkler J, Sullivan K, Isaacs A, et al. Reduction of
791 clofazimine by mycobacterial type 2 NADH:quinone oxidoreductase: a pathway for the generation
792 of bactericidal levels of reactive oxygen species. The Journal of biological chemistry.
793 2011;286(12):10276-87.
794 59. Van Deun A, Decroo T, Piubello A, de Jong BC, Lynen L, Rieder HL. Principles for
795 constructing a tuberculosis treatment regimen: the role and definition of core and companion
796 drugs. The international journal of tuberculosis and lung disease : the official journal of the
797 International Union against Tuberculosis and Lung Disease. 2018;22(3):239-45.
798 60. Brodie AF, Kalra VK, Lee SH, Cohen NS. Properties of energy-transducing systems in
799 different types of membrane preparations from *Mycobacterium phlei*--preparation, resolution,
800 and reconstitution. Methods in enzymology. 1979;55:175-200.
801 61. Cohen T, Colijn C, Murray M. Modeling the effects of strain diversity and mechanisms of
802 strain competition on the potential performance of new tuberculosis vaccines. Proceedings of the
803 National Academy of Sciences. 2008;105(42):16302.

804 62. Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and
805 implications for tuberculosis product development. Lancet Infect Dis. 2007;7.
806 63. Homolka S, Niemann S, Russell DG, Rohde KH. Functional Genetic Diversity among
807 *Mycobacterium tuberculosis* Complex Clinical Isolates: Delineation of Conserved Core and Lineage-
808 Specific Transcriptomes during Intracellular Survival. PLoS pathogens. 2010;6(7):e1000988.
809 64. Ofori-Anyinam B, Kanuteh F, Agbla SC, Adetifa I, Okoi C, Dolganov G, et al. Impact of the
810 *Mycobaterium africanum* West Africa 2 Lineage on TB Diagnostics in West Africa: Decreased
811 Sensitivity of Rapid Identification Tests in The Gambia. PLoS neglected tropical diseases.
812 2016;10(7):e0004801.
813 65. Sanoussi CN, Affolabi D, Rigouts L, Anagonou S, de Jong B. Genotypic characterization
814 directly applied to sputum improves the detection of *Mycobacterium africanum* West African 1,
815 under-represented in positive cultures. PLoS neglected tropical diseases. 2017;11(9):e0005900.
816 66. Sanoussi CN, de Jong BC, Odoun M, Arekpa K, Ali Ligali M, Bodi O, et al. Low sensitivity of
817 the MPT64 identification test to detect lineage 5 of the *Mycobacterium tuberculosis* complex.
818 Journal of medical microbiology. 2018;67(12):1718-27.
819 67. Dalton T, Cegielski P, Akksilp S, Asencios L, Campos Caoili J, Cho SN, et al. Prevalence of and
820 risk factors for resistance to second-line drugs in people with multidrug-resistant tuberculosis in
821 eight countries: a prospective cohort study. Lancet (London, England). 2012;380(9851):1406-17.
822 68. Rahi A, Matta SK, Dhiman A, Garhyan J, Gopalani M, Chandra S, et al. Enolase of
823 *Mycobacterium tuberculosis* is a surface exposed plasminogen binding protein. Biochimica et
824 biophysica acta General subjects. 2017;1861(1 Pt A):3355-64.
825 69. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution
826 phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism.
827 PLoS pathogens. 2011;7(9):e1002251.
828 70. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high
829 density mutagenesis. Molecular microbiology. 2003;48(1):77-84.
830 71. Tuckman D, Donnelly RJ, Zhao FX, Jacobs WR, Jr., Connell ND. Interruption of the
831 phosphoglucose isomerase gene results in glucose auxotrophy in *Mycobacterium smegmatis*.
832 Journal of bacteriology. 1997;179(8):2724-30.
833 72. Phong WY, Lin W, Rao SP, Dick T, Alonso S, Pethe K. Characterization of
834 phosphofructokinase activity in *Mycobacterium tuberculosis* reveals that a functional glycolytic
835 carbon flow is necessary to limit the accumulation of toxic metabolic intermediates under hypoxia.
836 PloS one. 2013;8(2):e56037.

837 73. Collins DM, Wilson T, Campbell S, Buddle BM, Wards BJ, Hotter G, et al. Production of
838 avirulent mutants of *Mycobacterium bovis* with vaccine properties by the use of illegitimate
839 recombination and screening of stationary-phase cultures. Microbiology (Reading, England).
840 2002;148(Pt 10):3019-27.
841 74. Liu K, Yu J, Russell DG. pckA-deficient *Mycobacterium bovis* BCG shows attenuated
842 virulence in mice and in macrophages. Microbiology (Reading, England). 2003;149(Pt 7):1829-35.
843 75. Daffe M, Crick DC, Jackson M. Genetics of Capsular Polysaccharides and Cell Envelope
844 (Glyco)lipids. Microbiology spectrum. 2014;2(4):Mgm2-0021-2013.
845 76. Singh A, Crossman DK, Mai D, Guidry L, Voskuil MI, Renfrow MB, et al. *Mycobacterium*
846 *tuberculosis* WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to
847 modulate macrophage response. PLoS pathogens. 2009;5(8):e1000545.
848 77. Munoz-Elias EJ, McKinney JD. *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are
849 jointly required for in vivo growth and virulence. Nat Med. 2005;11(6):638-44.
850 78. Nandakumar M, Nathan C, Rhee KY. Isocitrate lyase mediates broad antibiotic tolerance in
851 *Mycobacterium tuberculosis*. Nature communications. 2014;5:4306.
852 79. Honer Zu Bentrup K, Miczak A, Swenson DL, Russell DG. Characterization of activity and
853 expression of isocitrate lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*. Journal of
854 bacteriology. 1999;181(23):7161-7.
855 80. Li AH, Lam WL, Stokes RW. Characterization of genes differentially expressed within
856 macrophages by virulent and attenuated *Mycobacterium tuberculosis* identifies candidate genes
857 involved in intracellular growth. Microbiology (Reading, England). 2008;154(8):2291-303.
858 81. Bellerose MM, Baek SH, Huang CC, Moss CE, Koh EI, Proulx MK. Common Variants in the
859 Glycerol Kinase Gene Reduce Tuberculosis Drug Efficacy. 2019;10(4).
860 82. Diarra B, Kone M, Togo ACG, Sarro YDS, Cisse AB, Somboro A, et al. *Mycobacterium*
861 *africanum* (Lineage 6) shows slower sputum smear conversion on tuberculosis treatment than
862 *Mycobacterium tuberculosis* (Lineage 4) in Bamako, Mali. 2018;13(12):e0208603.
863 83. Sarkar R, Lenders L, Wilkinson KA, Wilkinson RJ, Nicol MP. Modern Lineages of
864 *Mycobacterium tuberculosis* Exhibit Lineage-Specific Patterns of Growth and Cytokine Induction in
865 Human Monocyte-Derived Macrophages. PloS one. 2012;7(8):e43170.
866 84. Tanveer M, Hasan Z, Kanji A, Hussain R, Hasan R. Reduced TNF-α and IFN-γ responses to
867 Central Asian strain 1 and Beijing isolates of *Mycobacterium tuberculosis* in comparison with
868 H37Rv strain. Transactions of The Royal Society of Tropical Medicine and Hygiene.
869 2009;103(6):581-7.

870 85. Yimer SA, Norheim G, Namouchi A, Zegeye ED, Kinander W, Tønjum T, et al. <span
871 class="named-content genus-species" id="named-content-
872 1">*Mycobacterium tuberculosis* Lineage 7 Strains Are Associated with
873 Prolonged Patient Delay in Seeking Treatment for Pulmonary Tuberculosis in Amhara Region,
874 Ethiopia. Journal of clinical microbiology. 2015;53(4):1301.
875 86. Vilcheze C, Weinrick B, Leung LW, Jacobs WR, Jr. Plasticity of *Mycobacterium tuberculosis*
876 NADH dehydrogenases and their role in virulence. Proceedings of the National Academy of
877 Sciences of the United States of America. 2018;115(7):1599-604.
878 87. Hartman T, Weinrick B, Vilchèze C, Berney M, Tufariello J, Cook GM, et al. Succinate
879 dehydrogenase is the regulator of respiration in *Mycobacterium tuberculosis*. PLoS pathogens.
880 2014;10(11):e1004510.
881 88. EPFL.
882 89. Minato Y, Gohl DM, Thiede JM, Chacón JM, Harcombe WR, Maruyama F, et al. Genome-
883 wide assessment of *Mycobacterium tuberculosis* conditionally essential
884 metabolic pathways. bioRxiv. 2019:534289.
885 90. Goldberg I, Lonberg-Holm K, Bagley EA, Stieglitz B. Improved conversion of fumarate to
886 succinate by Escherichia coli strains amplified for fumarate reductase. Applied and environmental
887 microbiology. 1983;45(6):1838-47.
888 91. Matsoso LG, Kana BD, Crellin PK, Lea-Smith DJ, Pelosi A, Powell D, et al. Function of the
889 cytochrome bc1-aa3 branch of the respiratory network in mycobacteria and network adaptation
890 occurring in response to its disruption. Journal of bacteriology. 2005;187(18):6300-8.
891 92. De Jong BC, Antonio M, Gagneux S. *Mycobacterium africanum*—Review of an Important
892 Cause of Human Tuberculosis in West Africa. PLoS Negl Trop Dis. 2010;4.
893 93. Frothingham R, Strickland PL, Bretzel G, Ramaswamy S, Musser JM, Williams DL.
894 Phenotypic and Genotypic Characterization of *Mycobacterium africanum* Isolates from West
895 Africa. Journal of clinical microbiology. 1999;37(6):1921-6.
896 94. Niemann S, Richter E, Rusch-Gerdes S. Differentiation among members of the
897 *Mycobacterium tuberculosis* complex by molecular and biochemical features: evidence for two
898 pyrazinamide-susceptible subtypes of M. bovis. Journal of clinical microbiology. 2000;38(1):152-7.
899 95. Parsons LM, Brosch R, Cole ST, Somoskovi A, Loder A, Bretzel G, et al. Rapid and simple
900 approach for identification of *Mycobacterium tuberculosis* complex isolates by PCR-based genomic
901 deletion analysis. Journal of clinical microbiology. 2002;40(7):2339-45.

902 96. Trauner A, Sassetti CM, Rubin EJ. Genetic Strategies for Identifying New Drug Targets.
903 Microbiology spectrum. 2014;2(4):Mgm2-0030-2013.
904 97. Zhang YJ, Ioerger TR, Huttenhower C, Long JE, Sassetti CM, Sacchettini JC, et al. Global
905 Assessment of Genomic Regions Required for Growth in *Mycobacterium tuberculosis*. PLoS
906 pathogens. 2012;8(9):e1002946.
907 98. Guest JR. Partial replacement of succinate dehydrogenase function by phage- and plasmid-
908 specified fumarate reductase in *Escherichia coli*. Journal of general microbiology. 1981;122(2):171-
909 9.
910 99. Berney M, Cook GM. Unique flexibility in energy metabolism allows mycobacteria to
911 combat starvation and hypoxia. PloS one. 2010;5(1):e8614.
912 100. Giffin MM, Shi L, Gennaro ML, Sohaskey CD. Role of alanine dehydrogenase of
913 *Mycobacterium tuberculosis* during recovery from hypoxic nonreplicating persistence. PloS one.
914 2016;11(5):e0155522.
915 101. Jeong J-A, Park SW, Yoon D, Kim S, Kang H-Y, Oh J-I. Roles of Alanine Dehydrogenase and
916 Induction of Its Gene in <span class="named-content genus-species"
917 id="named-content-1">*Mycobacterium smegmatis* under
918 Respiration-Inhibitory Conditions. Journal of bacteriology. 2018;200(14):e00152-18.
919 102. Jeong JA, Oh JI. Alanine dehydrogenases in mycobacteria. Journal of microbiology (Seoul,
920 Korea). 2019;57(2):81-92.
921 103. Weber I, Fritz C, Ruttkowski S, Kreft A, Bange FC. Anaerobic nitrate reductase (narGHJI)
922 activity of *Mycobacterium bovis* BCG in vitro and its contribution to virulence in immunodeficient
923 mice. Molecular microbiology. 2000;35(5):1017-25.
924 104. Arora K, Ochoa-Montaño B, Tsang PS, Blundell TL, Dawes SS, Mizrahi V, et al. Respiratory
925 Flexibility in Response to Inhibition of Cytochrome c Oxidase in <span
926 class="named-content genus-species" id="named-content-
927 1">*Mycobacterium tuberculosis*</span&gt. Antimicrobial Agents and Chemotherapy.
928 2014;58(11):6962.
929 105. Kalia NP, Hasenoehrl EJ, Ab Rahman NB, Koh VH, Ang MLT, Sajorda DR, et al. Exploiting the
930 synthetic lethality between terminal respiratory oxidases to kill *Mycobacterium*
931 *tuberculosis* and clear host infection. Proceedings of the National Academy of
932 Sciences. 2017;114(28):7426.
933 106. Scherr N, Bieri R, Thomas SS, Chauffour A. Targeting the *Mycobacterium ulcerans*
934 cytochrome bc1:aa3 for the treatment of Buruli ulcer. 2018;9(1):5370.

935 107. Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M, et al. Clade-Specific
936 Virulence Patterns of <span class="named-content genus-species"
937 id="named-content-1">*Mycobacterium tuberculosis* Complex Strains
938 in Human Primary Macrophages and Aerogenically Infected Mice. mBio. 2013;4(4):e00250-13.
939 108. O'Neill MB, Shockey A, Zarley A, Aylward W, Eldholm V, Kitchen A, et al. Lineage specific
940 histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. Molecular ecology. 2019.
941 109. Molina-Moya B, Gomgnimbou MK, Spinasse L, Obasanya J, Oladimeji O, Dacombe R, et al.
942 *Mycobacterium tuberculosis* complex genotypes circulating in Nigeria based on spoligotyping
943 obtained from Ziehl-Neelsen stained slides extracted DNA. PLoS neglected tropical diseases.
944 2018;12(2):e0006242.
945 110. Ates LS, Dippenaar A, Sayes F, Pawlik A, Bouchier C, Ma L, et al. Unexpected Genomic and
946 Phenotypic Diversity of *Mycobacterium africanum* Lineage 5 Affects Drug Resistance, Protein
947 Secretion, and Immunogenicity. Genome Biology and Evolution. 2018;10(8):1858-74.
948 111. Newton SM, Smith RJ, Wilkinson KA, Nicol MP, Garton NJ, Staples KJ, et al. A deletion
949 defining a common Asian lineage of *Mycobacterium tuberculosis* associates with immune
950 subversion. Proceedings of the National Academy of Sciences of the United States of America.
951 2006;103(42):15594-8.
952 112. Albanna AS, Reed MB, Kotar KV, Fallow A, McIntosh FA, Behr MA, et al. Reduced
953 Transmissibility of East African Indian Strains of *Mycobacterium tuberculosis*. PloS one.
954 2011;6(9):e25075.
955 113. Nebenzahl-Guimaraes H, Verhagen LM, Borgdorff MW, Van Soolingen D. Transmission and
956 progression to disease of *Mycobacterium tuberculosis* phylogenetic lineages in The Netherlands.
957 Journal of clinical microbiology. 2015;53(10):3264-71.
958 114. Broset E, Martín C, Gonzalo-Asensio J. Evolutionary Landscape of the <span
959 class="named-content genus-species" id="named-content-
960 1">*Mycobacterium tuberculosis* Complex from the Viewpoint of PhoPR:
961 Implications for Virulence Regulation and Application to Vaccine Development. mBio.
962 2015;6(5):e01289-15.
963 115. Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, et al.
964 Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence
965 regulator. Proceedings of the National Academy of Sciences of the United States of America.
966 2014;111(31):11491-6.

967 116. Gonzalo-Asensio J, Mostowy S, Harders-Westerveen J, Huygen K, Hernandez-Pando R,
968 Thole J, et al. PhoP: a missing piece in the intricate puzzle of *Mycobacterium tuberculosis*
969 virulence. PloS one. 2008;3(10):e3496.
970 117. Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, et al. The *Mycobacterium*
971 *tuberculosis* regulatory network and hypoxia. Nature. 2013;499(7457):178-83.
972 118. Capon F, Allen MH, Ameen M, Burden AD, Tillman D, Barker JN, et al. A synonymous SNP of
973 the corneodesmosin gene leads to increased mRNA stability and demonstrates association with
974 psoriasis across diverse ethnic groups. Human molecular genetics. 2004;13(20):2361-8.
975 119. Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, et al. Synonymous
976 mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the
977 receptor. Human molecular genetics. 2003;12(3):205-16.
978 120. Golby P, Nunez J, Witney A, Hinds J, Quail MA, Bentley S, et al. Genome-level analyses of
979 *Mycobacterium bovis* lineages reveal the role of SNPs and antisense transcription in differential
980 gene expression. BMC genomics. 2013;14:710-.
981 121. Baez A, Shiloach J. Effect of elevated oxygen concentration on bacteria, yeasts, and cells
982 propagated for production of biological compounds. Microbial cell factories. 2014;13:181-.
983 122. Siddiqi SaR-G, S. MGIT™ Procedure Manual, For BACTEC™ MGIT 960™ TB System
984 (Also applicable for Manual MGIT), Mycobacteria Growth Indicator Tube (MGIT)
985 Culture and Drug Susceptibility Demonstration Projects. 2006.
986 123. Godreuil S, Torrea G, Terru D, Chevenet F, Diagbouga S, Supply P, et al. First molecular
987 epidemiology study of *Mycobacterium tuberculosis* in Burkina Faso. Journal of clinical
988 microbiology. 2007;45(3):921-7.
989 124. Groenheit R, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, et al. The
990 Guinea-Bissau family of *Mycobacterium tuberculosis* complex revisited. PloS one.
991 2011;6(4):e18601.
992 125. Niobe-Eyangoh SN, Kuaban C, Sorlin P, Cunin P, Thonnon J, Sola C, et al. Genetic
993 Biodiversity of *Mycobacterium tuberculosis* Complex Strains from Patients with Pulmonary
994 Tuberculosis in Cameroon. Journal of clinical microbiology. 2003;41(6):2547-53.
995 126. Ouassa T, Borroni E, Loukou GY, Faye-Kette H, Kouakou J, Menan H, et al. High prevalence
996 of shared international type 53 among *Mycobacterium tuberculosis* complex strains in retreated
997 patients from Cote d'Ivoire. PloS one. 2012;7(9):e45363.

127. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annual review of genomics and human genetics. 2008;9:403-33.

128. Amouzou EK, Chabi NW, Adjalla CE, Rodriguez-Gueant RM, Feillet F, Villaume C, et al. High prevalence of hyperhomocysteinemia related to folate deficiency and the 677C-->T mutation of the gene encoding methylenetetrahydrofolate reductase in coastal West Africa. The American journal of clinical nutrition. 2004;79(4):619-24.

129. Gagneux S. Strain variation in the *Mycobacterium tuberculosis* complex: its role in biology, epidemiology and control. Springer, Heidelberg. 2017.

130. Moore SE, Mansoor MA, Bates CJ, Prentice AM. Plasma homocysteine, folate and vitamin B12 compared between rural Gambian and UK adults. British Journal of Nutrition. 2008;96(3):508-15.

Figures:

Figure 1: Link between central carbon metabolism and the Electron Transport Chain. Reducing equivalents like NADH and FADH2 enter into the Electron Transport Chain that drives the production of high amounts of ATP for cellular processes

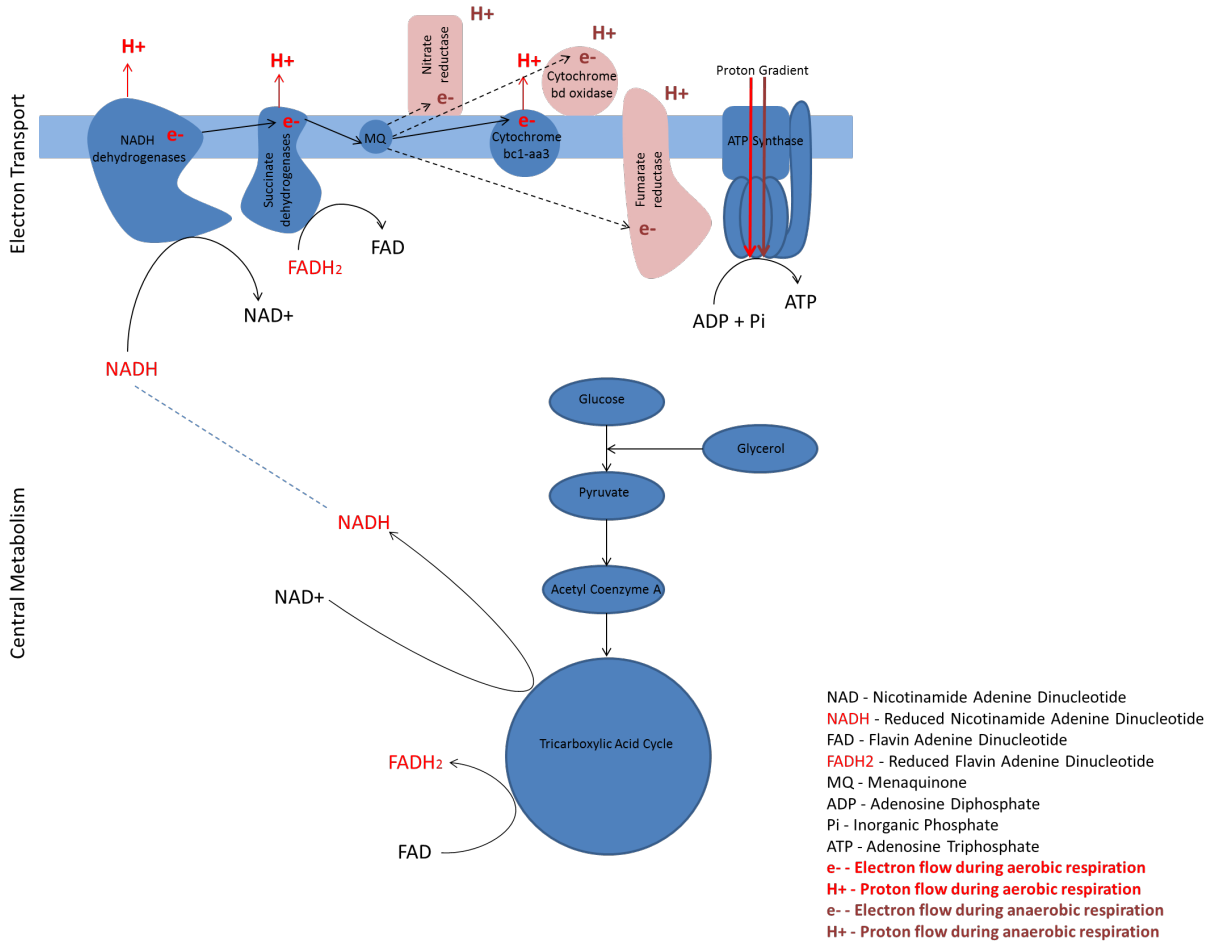

Figure 2: Mutations detected in the Central Carbon Metabolic Pathway of MAF L6 and MAF L5. Genes in **bold** have potentially deleterious mutations that can affect gene function.

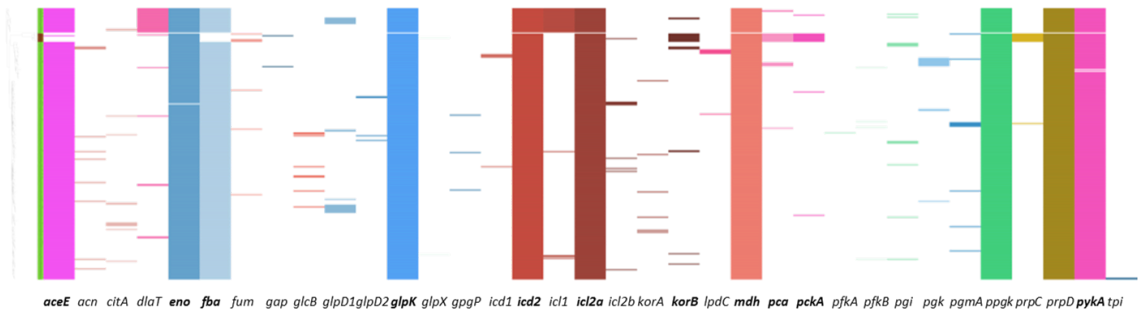

Figure 3: Mutations detected in the Central Carbon Metabolic Pathway of MTB L4 Ghana, L4 Haarlem, L1, L3, L2, L4 Cameroon and L4 LAM. Genes in **bold** have potentially deleterious mutations that can affect gene function.

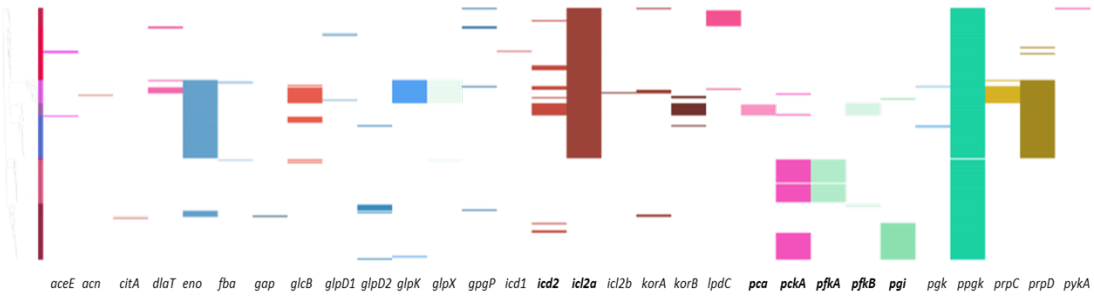

Figure 4: Mutations detected in the Electron Transport Chain of *M. africanum* MAF L6 and MAF L5. Genes in **bold** have potentially deleterious mutations that can affect gene function.

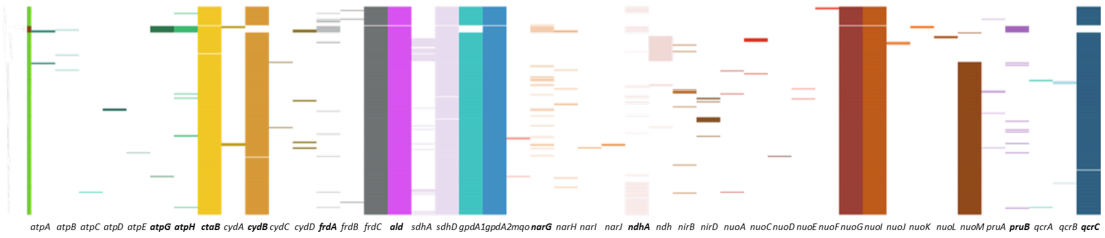

Figure 5: Mutations detected in the Electron Transport Chain of MTB L4 Ghana, L4 Haarlem, L1, L3, L2, L4 Cameroon and L4 LAM. Genes in **bold** have potentially deleterious mutations that can affect gene function.

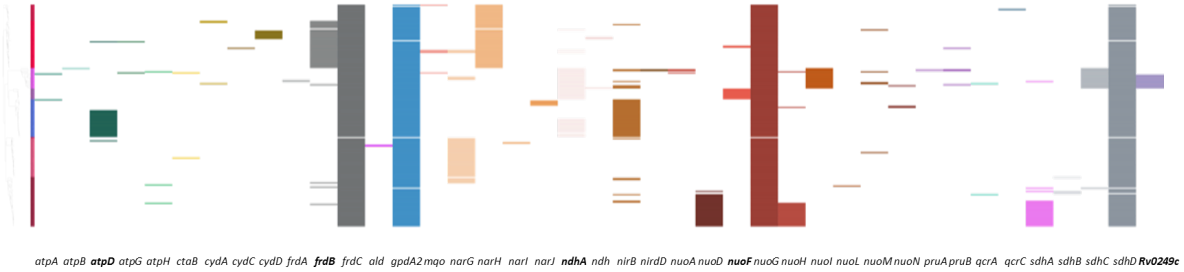

Figure 6: Mutations in the Central Carbon Metabolic Pathway of MAF. Genes with a **red** colour code have deleterious mutations in the

MAF lineages indicated next to the gene.

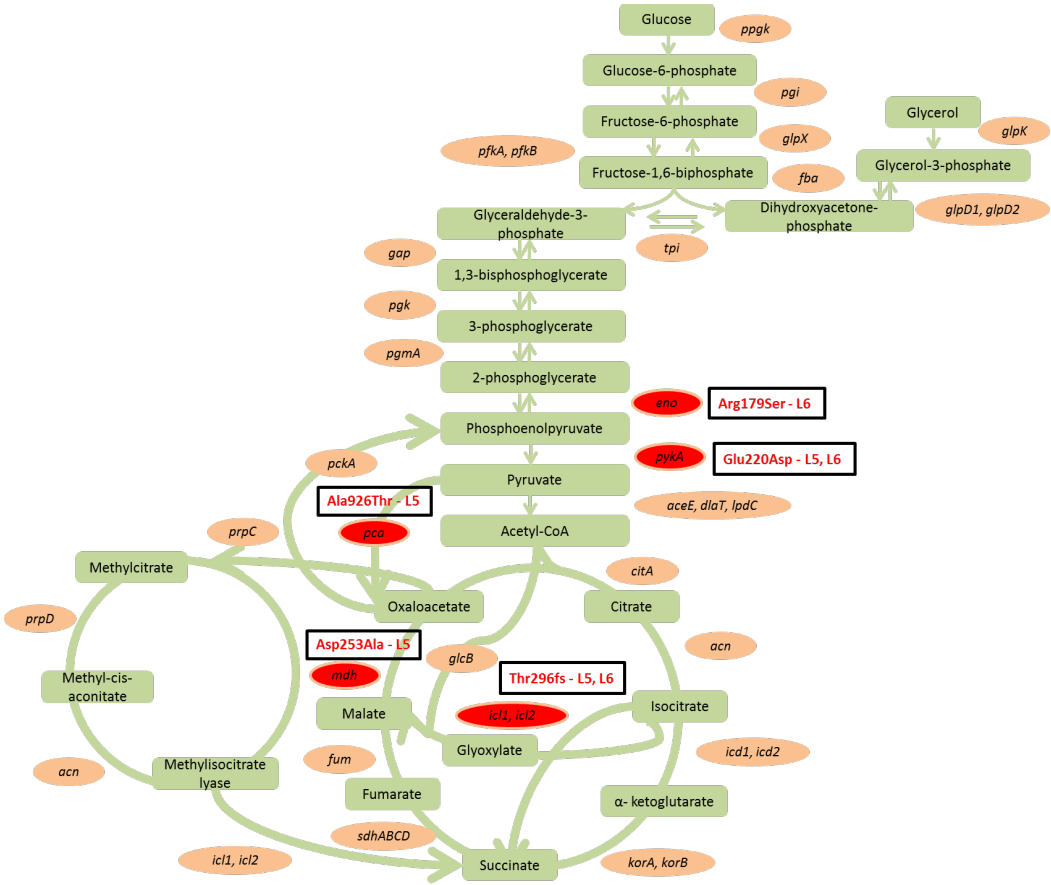

Figure 7: Mutations in the Central Carbon Metabolic Pathway of MTB. Genes with a **red** colour code have deleterious mutations in the

MTB lineages indicated next to the gene.

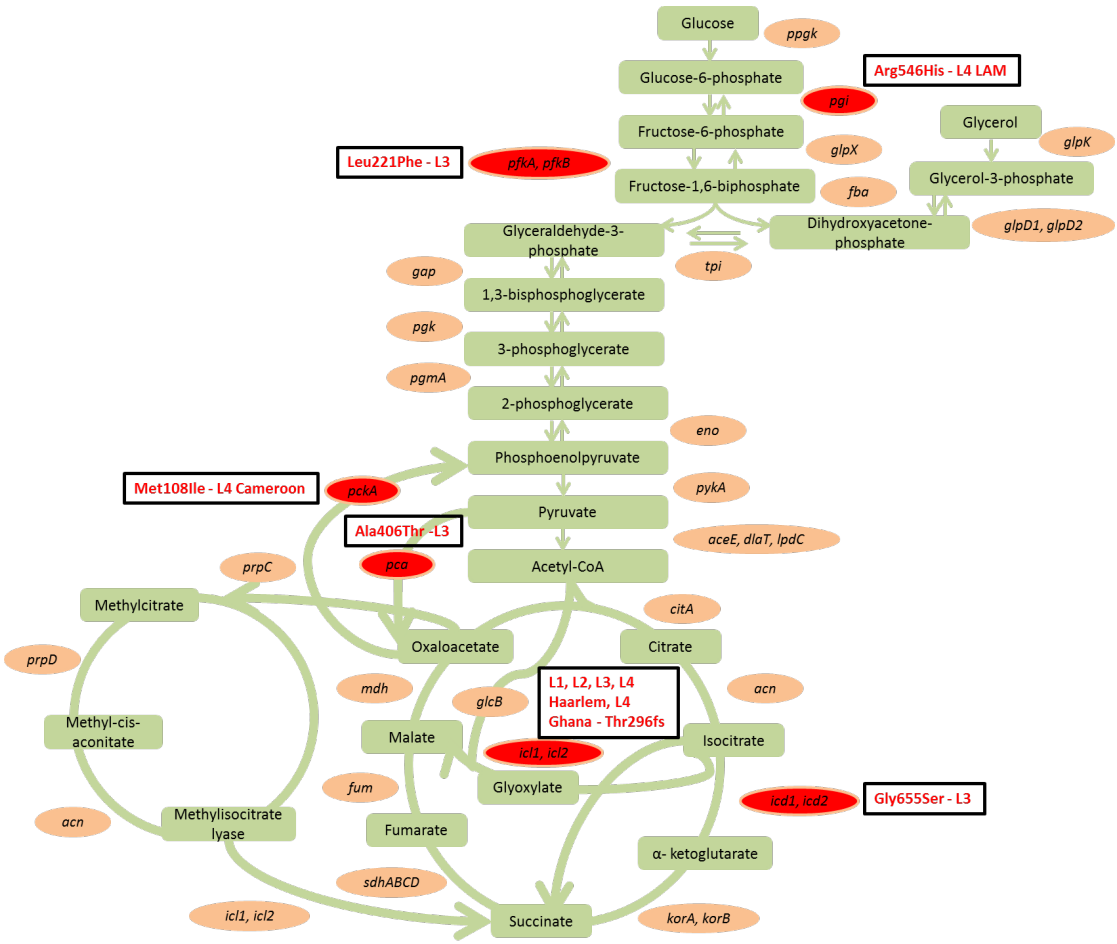

Figure 8: The core and alternate respiratory chain of mycobacteria during *in vitro* exponential growth when oxygen is abundant and when oxygen is limited. Mutations in MAF lineages are indicated in red boxes next to each affected complex. The core chain consists of type I NADH:menaquinone oxidoreductase (NuoA-N), succinate: menaquinone oxidoreductase 1 (SDH1), cytochrome aa3-bc supercomplex (Qcr-Cta) and F1F0-ATPase while the route used during oxygen limitation is composed of the Type II NADH:menaquinone oxidoreductase (Ndh); succinate:menaquinone oxidoreductase (SDH2); nitrate reductase (Nar); Fumarate Reductase (Frd) and cytochrome bd oxidase (Cyd), a high affinity terminal oxidase allowing hypoxic survival.

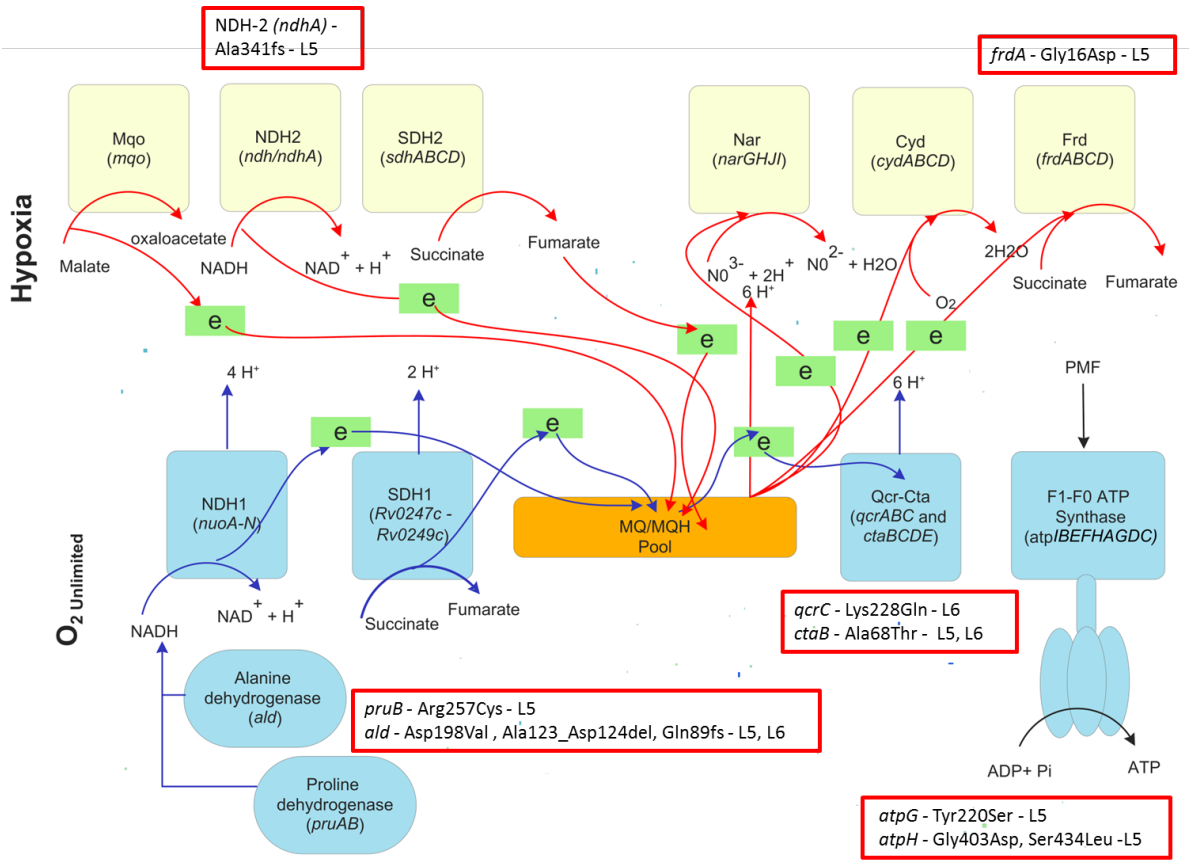

Figure 9: The core and alternate respiratory chain of mycobacteria during *in vitro* exponential growth when oxygen is abundant and when oxygen is limited. Mutations in MTB lineages are indicated in red boxes next to each affected complex. The core chain consists of type I NADH:menaquinone oxidoreductase (NuoA-N), succinate: menaquinone oxidoreductase 1 (SDH1), cytochrome aa3-bc supercomplex (Qcr-Cta) and F1F0-ATPase while the route used during oxygen limitation is composed of Type II NADH:menaquinone oxidoreductase (Ndh); succinate:menaquinone oxidoreductase (SDH2); nitrate reductase (Nar); Fumarate Reductase (Frd) and cytochrome bd oxidase (Cyd), a high affinity terminal oxidase allowing hypoxic survival.

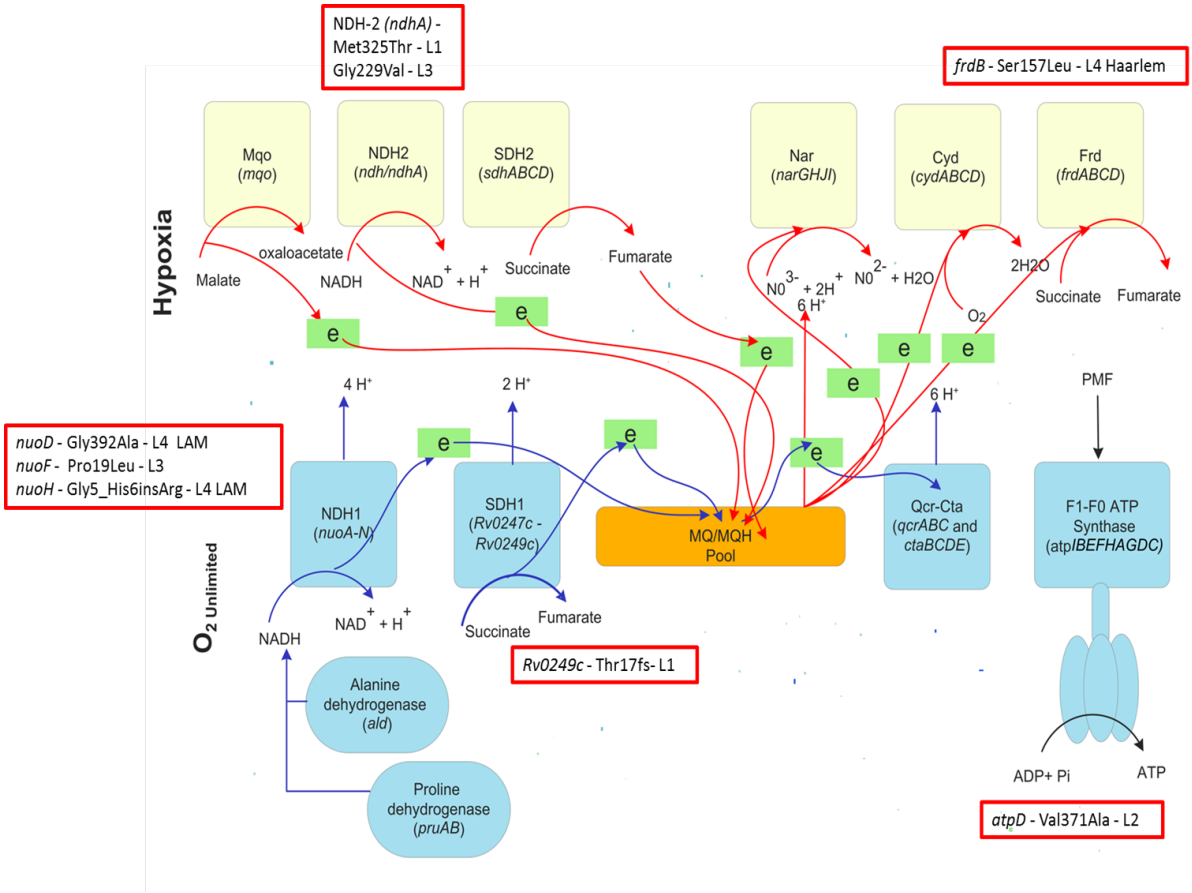

Figure 10: The MTB H37Rv proposed regulatory network showing differentially regulated and or mutated genes in MAF lineages based on findings from our present analysis, Ofori-Anyinam *et al.*, Tuberculosis, 2017, Galagan *et al.*, Nature, 2013 and Gonzalo-Asensio *et al.*, Plos One, 2008).

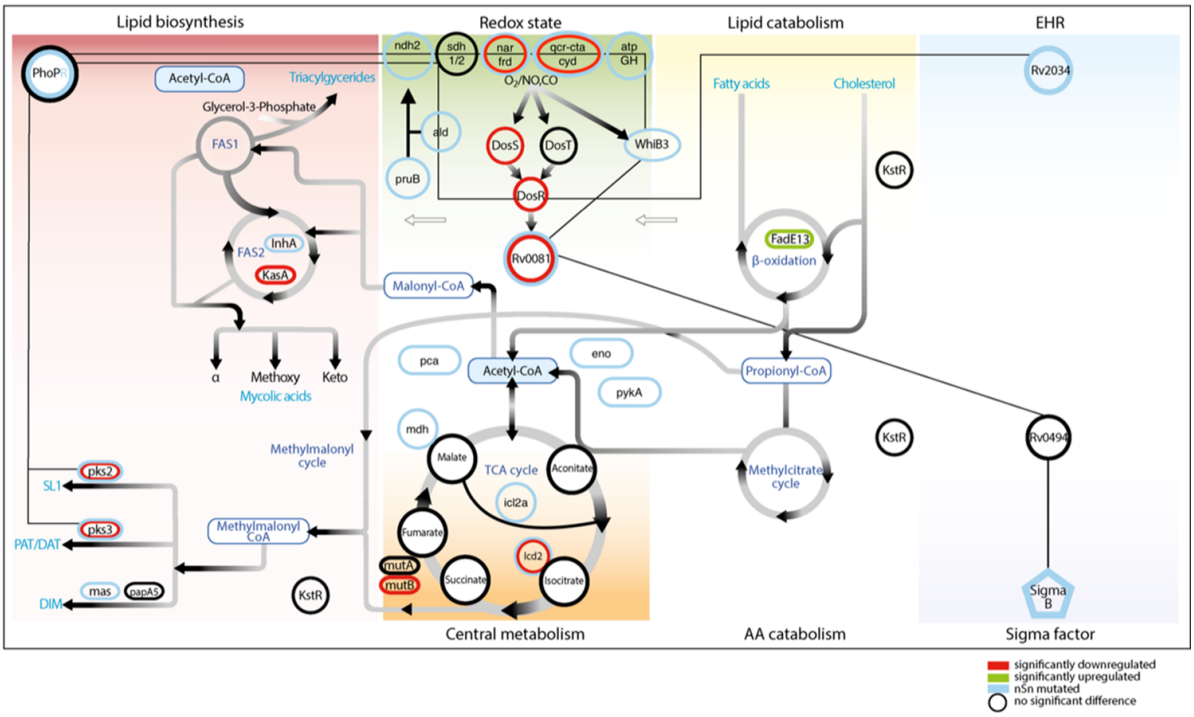

Table 1: Mutations in genes encoding enzymes in the Central Carbon Metabolic Pathway. Information in **red** indicates potentially deleterious changes and the corresponding lineages they were detected in. **x** implies convergence occurred between MTB and MAF lineages and **x** implies the mutation is potentially deleterious.

Gene	MTB amino acid change	MTB Mutation	MTB Lineage	MAF amino acid change	MAF Mutation	MAF Lineage	Convergent evolution
ppgk	Ile203Thr	I203T	All lineages	Ile203Thr	I203T	L5, L6	x
pgi	Arg546His	R546H	clade of L4 LAM				
pfkA	Gln51His	Q51H	L4 Cameroon				
pfkB	Leu221Phe	L221F	L3				
fba				Ile137Val	I137V	L6	
glpK	Val460Ala	V460A	L1	Val460Ala Glu209Lys Val192fs	V460A E209K V192fs	L5,L6	x
eno	Arg8Gly Lys429Gln	R8G K429Q	L1, L2, L3	Arg8Gly Arg179Ser	R8G R179S	L5, L6	x
pykA				Glu220Asp	E220D	L5,L6	
aceE				Ala777Thr	A777T	L6	

				Ala528Thr	A528T		
				Ser221Phe	S221F		
pca	Ala406Thr	A406T	L3	Ala926Thr	A926T	L5	x
icd2	Gly655Ser	G655S	L3	Val447Met	V447M	L6	
	Leu254Val	L254V		Lys117Asn	K117N		
mdh				Leu326Ile	L326I	L5, L6	
				Ala112Thr	A112T		
				Asp253Ala	D253A		
korB	Val293Ala	V293A	L3	Ala309Gly	A309G	L5	x
icl2a	Gly179Asp	G179D	L1,L2,L3,L4	Ala146Thr	A146T	L5, L6	x
	Thr296fs	T296fs	Haarlem, L4	His151Arg	H151R		
			Ghana	Gly179Asp	G179D		
				Thr296fs	T296fs		
prpC	Val46Met	V46M	L1				
prpD	Asp3Gly	D3G	L1,L2,L3	Asp3Gly	D3G	L5, L6	x
pckA	Met108Ile	M108I	L4	Lys422Thr	K422T	L5	x
	Ala536Gly	A536G	Cameroon,				
			clade of L4				
			LAM				

Table 2: Mutations in genes and gene subunits encoding Electron Transport Chain enzymes. Information in **red** indicates potentially deleterious changes and the corresponding lineages they were detected in. **x** indicates that convergence occurred between MTB and MAF lineages and **x** indicates that the mutation is potentially deleterious.

Gene	MTB amino acid change	MTB Mutation	MTB Lineage	MAF amino acid change	MAF Mutation	MAF Lineage	Convergent evolution
ndh				Val117ile	V117I	clade of L6	
				Lys341Glu	K341E		
ndhA	Asn360Ser	N360S	clade of L2	Ala341fs	A341fs	L5	x
	Met325Thr	M325T	L1				
	Gly229Val	G229V	L3				
gpdA1				Thr190Ala	T190A	L6	
gpdA2	Pro133Ala	P133A	All lineages	Pro133Ala	P133A	L5, L6	x
pruB				Arg257Cys	R257C	L5	
qcrC				Lys228Gln	K228Q	L6	
ctaB				Ala68Thr	A68T	L5, L6	
cydB				Met98Ile	M98I	L6	
frdA				Gly16Asp	G16D	L5	
frdB	Ser157Leu	S157L	L4 Haarlem				
frdC	Met40Val	M40V	All lineages	Met40Val	M40V	L5, L6	x

narG				Ser435Arg	S435R	L5	
narH	Ala467Val	A467V	L4 Ghana, L4 Haarlem				
nirB	Val775Leu	V775L	L2				
atpG				Tyr220Ser	Y220S	L5	
atpH				Gly403Asp Ser434Leu	G403D S434L	L5	
atpD	Val371Ala	V371A	L2				
nuoD	Gly392Ala	G392A	clade of L4 LAM				
nuoG	Ile474Met Thr604Ala	I474M T604A	All lineages	Ile474Met Thr604Ala	I474M T604A	L5, L6	x
nuoH	Gly5_His6insArg	G5_H6insR	clade of L4 LAM				
nuoF	Pro19Leu	P19L	L3				
nuoM				Asp309Glu	D309E	clade of L6	
nuoI	Thr180Ala	T180A	L3	Thr180Ala	T180A	L6, L5	x
sdhA	Glu269Asp	E269D	L4 LAM	Gly533Asp	G533D	clade of L6	x
Rv0249c	Thr17fs	T17fs	clade of L1				
sdhC	Leu54Ile	L54I	L1				

sdhD	Val54Leu	V54	All lineages	Val54Leu	V54L	L5, L6	x
	Thr112Ala	T112A					
ald				Gln89fs	Q89fs	L5, L6	
				Asp198Val	D198V		
				Ala123_Asp124del	A123_D124del		

Table 3: Future proposed experiments.

Examination	Proposed Experiments	Predicted Phenotype/ Expected outcome
Experimentally determine if mutations detected in *eno*, *pca*, *pgi*, *pckA*, *korB*, *fba*, *aceE* and *glpK* of the Glycolytic Pathway and TCA cycle contribute to slow growth.	Mutagenesis experiments, complementation studies, Growth studies.	MAF, MTB L1 and L3 will show attenuated growth relative to MTB L2 and L4.
Experimentally determine if mutations detected in *ndhA*, *pruB*, *qcrC*, *ctaB*, *frdAB*, *sdhA*, *atpDHG*, *ald*, *Rv0249c*, *narG* and *cydB* of the respiratory pathway contribute to slow growth.	Mutagenesis experiments, complementation studies, Growth studies.	MAF, MTB L1 and L3 will show attenuated growth compared to MTB L2 and L4.
Energy generation, in the form of ATP, for cellular processes	ATP quantitation assays.	MAF lineages will generate less ATP via the respiratory chain compared to MTB lineages
ROS accumulation and DNA damage in MAF cultures relative to MTB	Flow cytometry experiments and terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) assay.	Higher levels of ROS in MAF cultures and greater DNA damage of MAF relative to MTB.
Host-Pathogen interactions driving geographic restriction of MAF lineages to West Africa	Host genetics and Genome-wide association studies.	Specific genes in West Africans have undergone strong positive selection and increased the susceptibility of West

		Africans to MAF infection, permitting host restrictedness.