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 *Mycobacterium 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 order 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 mutations with the potential to affect enzymes in these pathways in MAF lineages compared to MTB lineages. We also found that similar mutations occurred in these pathways 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 indicate potential adaptation of MAF lineages to a distinct ecological niche, which we suggest includes 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

53 **1 Introduction**

54 The *Mycobacterium tuberculosis* complex (MTBC) consists of a group of human–adapted
55 ecotypes- *Mycobacterium tuberculosis* sensu stricto, *Mycobacterium africanum*,
56 *Mycobacterium canetti* and animal-adapted ecotypes (1-4). There are seven known MTBC
57 lineages (L) associated with particular geographic regions and adapted to specific human
58 populations. These are the five lineages that make up *M. tuberculosis* sensu stricto (lineages
59 1-4 and lineage 7) and the two *M. africanum* lineages (lineages 5 and 6). Africa uniquely has
60 a representation of all seven lineages.

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

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

78 Genes that encode proteins involved in nutrient metabolism and respiration, which are
79 closely linked and together govern bacterial growth and survival, and those of unknown
80 function, the conserved hypotheticals, are highest among the 4173 MTB H37Rv genes
81 reported, emphasizing the importance of the nutrient metabolic and respiratory pathways
82 (17).

83 As obligate aerobes, the mycobacteria respire and produce energy from varied nutritional or
84 energy sources, such as carbon sources (18, 19). These enable the bacteria even to maintain
85 metabolism without growth. The nutritional demands of the mycobacteria have been a topic
86 of interest for over 100 years. Pioneering work throughout the 20th century elegantly showed
87 that mycobacteria had unique nutritional requirements (20-29). Central to these findings was
88 that different members of the MTBC were supported by different nutritional sources and
89 consistent results of multiple phenotypic studies led to differentiating members of the MTBC
90 based on their nutritional requirements (20, 30). For instance, it was observed that MTB
91 showed eugonic growth on glycerol, while colonies of MAF and *M. bovis* were dysgonic,
92 indicating an inability to properly utilize this carbon source. MAF and *M. bovis* were only able
93 to show luxurious growth in the presence of sodium pyruvate (20, 30). Furthermore, MAF was
94 found to grow small umbilicated colonies, which on paraffin embedded thin sections revealed
95 extension deep into the media, rather than the surface (31).

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

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

104 Genome sequencing analyses show that the mycobacteria possess a branched respiratory
105 pathway for electron transfer from electron donors to acceptors under different growth
106 conditions. However, it appears that the transfer of electrons to oxygen, which seems to be
107 the most preferred terminal electron acceptor, occurs with little plasticity, given that only two
108 terminal oxidases have been found in mycobacteria, the bioenergetically more efficient *aa3*-
109 type cytochrome c and the less efficient cytochrome *bd*-type menaquinol oxidases (33).
110 During hypoxia or anoxia, mycobacterial growth is inhibited, even in the presence of alternate
111 terminal electron acceptors within the branched respiratory chain such as fumarate and

112 nitrate reductase. However, mycobacteria are still able to adapt and maintain metabolic
113 functions (33).

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

116 The importance of comparative genomics in unravelling the basis of distinct metabolic
117 phenotypes in the MTBC has been shown. Using molecular genomic approaches, previous
118 authors found that the inability of *M. bovis* to use glycerol and carbohydrates as sole carbon
119 sources and its requirement for pyruvate in growth media was caused by a single nucleotide
120 polymorphism in the *pykA* gene, encoding pyruvate kinase, resulting in a Glu220Asp amino
121 acid substitution and causing the disruption of sugar catabolism (34). This mutation was also
122 found in 3 MAF strains tested in the same analysis. Additionally, the authors showed that a
123 frameshift at codon 191 of the *glpK* gene of the same *M. bovis* strain led to an incomplete
124 coding sequence and the inability to use glycerol, although this *glpK* mutation was not present
125 in all *M. bovis* strains.

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

131

132 **2 Materials and Methods**

133 **2.1 Ethical Statement**

134 The study was conducted within the framework of an intervention trial of Enhanced Case
135 Finding (ECF) in the Greater Banjul Area of The Gambia (Clinicaltrials.gov NCT01660646),
136 piloted in 2012 and conducted between 2013 and 2017. This study was carried out in
137 accordance with the recommendations of the Joint Gambia Government/MRC Ethics
138 Committee and the Institute of Tropical Medicine, Antwerp Institutional Review Board. The
139 protocol, including bacterial sub-studies, was approved by the Joint Gambia
140 Government/MRC Ethics Committee and the Institute of Tropical Medicine, Antwerp

141 Institutional Review Board. Nigerian isolates were collected from Southwest Nigeria within
142 the West African Node of Excellence for TB, AIDS and Malaria (WANETAM) with the
143 recommendations of the University of Ibadan and University College Hospital, Ibadan Joint
144 Ethical Review Committees and the Nigerian Institute of Medical Research, Institutional Board
145 (35). All subjects gave written informed consent in accordance with the Declaration of Helsinki
146 and were anonymized.

147

148 **2.2 Bacterial isolates**

149 In The Gambia, MTB L4 followed by MAF L6 are the most isolated MTBC lineages. For almost
150 a decade, the prevalence of all the MTBC lineages isolated in The Gambia has remained
151 constant at 4.3% (L1), 2.5% (L2), 0.8% (L3), 57.2% (L4), 1.0% (L5), and 35.4% (L6) (36).

152 Within the framework of the ECF study conducted in the Greater Banjul Area, we sequenced
153 280 MAF L6 (32%), 3 MAF L5 (0.3%), 19 MTB L1 (2.2%), 36 MTB L2 (4%)(Beijing), 10 MTB L3
154 (1%) and 534 (60.5%) MTB L4 consisting of 85 MTB L4 Cameroon (9.6%), 15 MTB L4 Ghana
155 (1.7%), 224 MTB L4 Haarlem (25.3%), and 211 MTB L4 LAM (23.9%). Given that only 3 MAF
156 L5 were isolated from The Gambia within the period of analysis, we included 6 MAF L5 from
157 Nigeria (Eastern West Africa), to improve the representativeness of this lineage.

158 Of this dataset, we analyzed the whole genome sequences of all 280 MAF L6 strains, the 3
159 MAF L5 isolated from The Gambia and the 6 MAF L5 from Nigeria, resulting in a total 289
160 MAF strains. For the MTB lineages, we analyzed a total 205 strains consisting of all 19 MTB
161 L1, all 10 MTB L3 and 15 MTB L4 Ghana, 35 MTB L2 and a random number of L4 Haarlem
162 (44), Cameroon (36) and LAM (46), while ensuring that all MTB L4 sublineages isolated were
163 represented and isolates from each year of isolation, 2012 to 2014, were included.

164 **2.3 DNA Extraction**

165 Genomic DNA was extracted from loopfuls of pure MTBC colonies grown on Lowenstein-
166 Jensen media (37) using the Maxwell 16 DNA Purification Kit (Promega). DNA from Nigerian
167 strains was extracted using the Cetyl trimethylammonium bromide (CTAB) method (38).

168 ## 2.4 Whole-genome sequencing

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

175

176 ## 2.5 Bioinformatics Analysis

177 ### 2.5.1 Mapping and Variant calling

178 We used Snippy version 3.1 for the analysis of genomes. Briefly, paired-end raw reads of each
179 sample were mapped to the *M. tuberculosis* H37Rv reference genome (GenBank accession
180 number: NC_000962.3) using BWA-MEM 0.7.12 (39). Mapped reads were converted to the
181 SAM/BAM format and sorted using Samtools 1.3 1 (40). Variant calling was done using
182 Freebayes 0.9.20 (41). Variants were called only if ≥10 reads covered variant positions and
183 ≥90% of those reads differed from the reference. Genes were annotated with SnpEff 4.1 (42).
184 Samples were assigned to MTBC lineages based on the classification of Coll and colleagues
185 (43) using the VCF output of snippy and the PhyResSE SNP list (44).

186 ### 2.5.2 Phylogenetic analysis

187 Using the Snippy output folders of all isolates and custom python scripts, a SNP alignment and
188 a count of all excluded invariant sites were created. A maximum-likelihood (ML) phylogeny
189 was inferred using RAxML version 8.2.9 (45) executing a thousand rapid bootstrap inferences
190 under the general time-reversible (GTR) model, with ascertainment bias correction using the
191 Stamatakis reconstituted DNA approach (46, 47). The resulting tree was exported to
192 interactive Tree of Life (iTOL) for visualization (48).

193 ### 2.5.3 Protein Predictions

194 The effect of amino acid substitutions were predicted using PROVEAN software tools at
195 default settings (PROVEAN protein) (49-51). The PROVEAN NCBI non redundant 2012

196 database is a newer and larger sequence database than SIFT databases (52) but comparable
197 to SIFT in prediction accuracy. A PROVEAN score of ≤ -2.5 implies that the amino acid
198 substitution could impact negatively on protein function.

199

200 **3 Results and Discussion**

201 Given that generalists and specialists differ in the vastness of their ecological niches (53), and
202 thus carbon sources, we compared the specialist MAF lineages (289 strains) to the generalist
203 MTB lineages (205 strains) for genomic differences in their Electron Transport and Carbon
204 metabolic pathways. These pathways are intrinsically linked and are central to deriving energy
205 (ATP) from carbon sources (Figure 1).

206 Understanding differences in metabolism and respiration between the MTBC lineages is also
207 pertinent for tackling TB as mycobacterial central metabolism and respiration have re-
208 emerged as potential targets for TB chemotherapy. The new TB drug, Bedaquiline (TMC207)
209 and the drug candidate, Telacebec (Q203), both target the respiratory chain (54-56) and even
210 the repurposed drug Clofazimine, and another new drug, Delamanid, reportedly interfere
211 with redox cycling and cellular respiration by generating Reactive Intermediates (57-61). Thus
212 respiratory inhibitors may offer the next generation of core drugs against mycobacterial
213 diseases (62) and there is a pressing need to understand the differences in these processes
214 between the different lineages of the MTBC.

215 It is now increasingly apparent that genetic and metabolic differences between the MTBC
216 human-adapted lineages have the potential to affect transmission, diagnostics and treatment
217 (5, 63-68). Moreover, niche adaptation is influenced by the metabolic requirements of an
218 organism, a consequence of evolution. However, the extent of such genetic changes and their
219 potential metabolic effects has not been well explored. In this study, we found a large number
220 of mutations with the potential to negatively affect gene function (hereafter referred to as
221 harmful mutations), particularly in MAF lineages (Figure 2 and Figure 4). However, we also
222 observed that similarities in these pathways occurred between some MTB and MAF lineages
223 potentially due to convergent evolution.

224

225 ## 3.1 Mutations in genes encoding Central Carbon Metabolic pathway enzymes.

226 We examined mutations in all genes in the glycolytic pathway, the Tricarboxylic Acid cycle and
227 the Methylcitrate Cycle. Genes and enzymes in these pathways contribute directly to the
228 growth and virulence of the MTBC, by generating products that feed directly into the Electron
229 Transport Chain (69). Therefore, mutations, with the potential to affect proper function of
230 these genes and their gene products, would have obvious consequences downstream, for
231 energy generation, survival and ultimately, the virulence of members of the MTBC.

232 ### 3.1.1 The Glycolytic Pathway

233 In MAF lineages, all genes leading to the breakdown of sugars to 2-phosphoglycerate at the
234 seventh step (Figure 2 and Figure 3), were generally conserved. However, at the critical stage
235 of pyruvate metabolism, mutations with the potential to affect the normal function of Enolase
236 (*eno*), Pyruvate kinase (*pykA*) and Pyruvate carboxylase (*pca*) were detected. Enolase is
237 responsible for the penultimate step of glycolysis, where 2-phosphoglycerate is converted to
238 Phosphoenolpyruvate. Pyruvate kinase modulates the irreversible reaction converting
239 Phosphoenolpyruvate to Pyruvate while Pyruvate carboxylase drives the conversion of
240 Pyruvate to Oxaloacetate. Of these three genes, it had previously been shown, in *M. bovis*
241 and 3 MAF strains, that the mutation in *pykA* rendered the critical enzyme pyruvate kinase
242 inactive, resulting in a metabolic block at the level of pyruvate biosynthesis (34). We also
243 confirm this mutation in the MAF strains we analyzed in our study (Figure 2). However, to the
244 best of our knowledge this is the first report on lineage-specific mutations with potentially
245 harmful effects in the essential gene *eno*, in MAF L6. *eno* is found on the surface of many
246 pathogenic bacteria and beyond its primary role in glycolysis, aids in tissue remodeling and
247 invasion of host cells (70). This gene was recently reported as a novel target for the 2-
248 aminothiazoles of the aminothiazole (AT) series, that exerted its effect by inhibiting *eno* in
249 MTB H37Rv (L4 strain). Therefore, the potentially harmful mutations we detected in *eno* may
250 affect energy metabolism and potentially contributes to the metabolic block at the level of
251 pyruvate biosynthesis, further reducing the fitness of L6.

252 Unlike the MAF lineages, MTB lineages, notably, did not have any potentially harmful
253 mutations in *eno* and *pykA* (Figure 2 and Figure 3), implying a general ability to complete
254 glycolysis. However, mutations were detected in Glucose-6-phosphate isomerase (*pgi*) in L4

255 LAM and in Phosphofructokinase B (*pfkB*) in MTB L3, at the second and third steps of glycolysis
256 respectively, where Glucose-6-phosphate is converted to Fructose-6-phosphate and
257 subsequently to Fructose-1-6-biphosphate (Figure 3). *pgi* is essential for the in vitro growth
258 of *M. tuberculosis*, and *M. smegmatis pgi* mutants are glucose auxotrophs. Moreover, an
259 additional role for *pgi* in cell wall biosynthesis was reported (37, 71-73). Therefore, *pgi* is
260 important and the effect of the predicted harmful mutations we detected in *pgi* in L4 LAM
261 should be investigated. However, the defect in *pfkB* in MTB L3 is unlikely to affect glycolysis
262 given that Phosphofructokinase activity has so far only been associated with *pfkA,* where a
263 *pfkA* deletion mutant was neither able to grow on glucose in vitro nor to have any detectable
264 Phosphofructokinase activity; the mutant could not be rescued by expressing *pfkB* (32, 74).

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 3). *pca* and *pckA* secrete enzymes that contribute significantly
268 to the control of metabolic flux to glycolysis, gluconeogenesis and anaplerosis. These genes
269 carry out functions related to cholesterol detoxification and lipogenesis during intracellular
270 survival and *pckA* was shown to be essential for virulence in *M. bovis* (75, 76).

271 **3.1.2 The Tricarboxylic Acid Cycle**

272 At the final stage of glycolysis, pyruvate is generated and converted to Acetyl Coenzyme A
273 (Figure 1 and Figure 3), that is fed into the critical Tricarboxylic Acid Cycle (TCA), for the release
274 of reducing equivalents into the Electron Transport Chain. Therefore, the ability to complete
275 this cycle, whether by progression through all stages of the cycle or via the crucial glyoxylate
276 by-pass, is essential for the generation of ATP.

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

280 Emphasizing the importance of the glyoxylate shunt, *M. tuberculosis* strains lacking both ICLs
281 are unable to grow on fatty acids in vitro, establish and maintain a chronic infection in mice
282 and were said to be the most severely attenuated strains (32, 77). In our analysis, we found
283 the frameshift mutation previously found in H37Rv, an L4 strain (78), which we, for the first
284 time to the best of our knowledge, also report in MTB L4 Haarlem and Ghana, MTB L1, L2, L3,

285 MAF L5 and L6, yet, interestingly, not in MTB L4 Cameroon and LAM (Figure 2, Figure 3 and
286 Table 1).

287 As stated earlier, in section 3.1.1, Pyruvate carboxylase (*pca*), that converts pyruvate to
288 oxaloacetate towards the final stage of glycolysis, was mutated in L5, however, at the last
289 step of the TCA cycle, malate dehydrogenase (*mdh*) that converts malate to oxaloacetate, was
290 also mutated (Figure 2, Figure 3 and Table 1), implying that both routes for the production of
291 oxaloacetate in MAF L5 have mutated genes. Given that oxaloacetate feeds not only into the
292 TCA cycle but also into the methylcitrate cycle, impaired function of *mdh* may affect energy
293 metabolism in MAF L5.

294 Overall, with major blocks in Central Carbon Metabolism in MAF lineages, the number of
295 reducing equivalents produced via the central carbon metabolic pathway for electron
296 transport may be lower. MTB lineages largely had a conserved glycolytic pathway and TCA
297 cycle, however, different MTB lineages have previously been shown to have different growth
298 rates and patterns (79-81). Notably, the growth rate of MTB L3 is reportedly lower compared
299 to other MTB lineages (79, 80). Therefore, in MTB lineages/sublineages where we found more
300 potentially harmful mutations, particularly in MTB L3, further investigations on the effect of
301 these mutations on energy metabolism need to be carried out as slower growth of these
302 lineages may be directly linked to impaired energy metabolism.

303

304 **3.2 Mutations in genes encoding Electron Transport Chain enzymes**

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

311 Relative to MTB lineages, MAF lineages had multiple mutations predicted to affect the normal
312 function of genes encoding key enzymes of the Electron Transport Chain (Figure 4 and Figure
313 5).

314 Potentially harmful mutations were detected in *ndhA*, *pruB*, *qcrC*, *ctaB*, *frdA*, *atpHG*, *sdhA* and
315 *ald* (Figure 4, Figure 5, Table 2 and Supplementary File S3). Interestingly, for some MTB
316 lineages, mutations predicted to affect gene function were also found in *ndhA*, *Rv0249c*, *frdB*,
317 *atpD*, and *nuoDHF* (Figure 4, Figure 5, Table 2 and Supplementary File S4).

318 **3.2.1 NADH dehydrogenases**

319 To receive NADH from central metabolism into the Electron Transport Chain, *M. tuberculosis*
320 possesses two NADH dehydrogenases, NDH-1 and -2. NDH-1 has 14 subunits (*nuo*A-N) while
321 two copies of NDH-2 exist in *M. tuberculosis*: *ndh* and *ndhA*. Between the two NADH
322 dehydrogenases, an essential role of *ndh* for the growth of *M. tuberculosis* was reported
323 previously and a function of *ndh* in recycling NADH and maintaining an energized membrane
324 was documented (33, 82). Even though *ndhA* was previously reported to be dispensable for
325 growth, it was recently shown that both *ndh* and *ndhA* differentially control oxygen
326 consumption (82). In fact, *ndh* was also shown to be dispensable for growth of *M. tuberculosis*
327 but deletion of both *ndh* and *ndhA* prevented growth altogether in standard media and
328 resulted in attenuated growth in mice (82, 83). We found a potentially harmful mutation in
329 *ndhA* in all MAF L5. Given the adaptation of MAF lineages towards a microaerophilic lifestyle,
330 *ndhA* may be redundant in MAF L5 and defects in this gene may have coincided with the
331 adaptation to a microaerophilic lifestyle (84).

332 No potentially harmful mutations were detected in the essential *ndh* in any analyzed genome.
333 This implies that the function of *ndh* is likely conserved across the MTBC lineages.

334 **3.2.2 Succinate dehydrogenases**

335 Succinate dehydrogenase 1 (*Rv0247c-Rv0249c*, SDH-1) functions during aerobic respiration
336 when oxygen and nutrients are abundant, while SDH-2 (*sdhABCD*) functions during
337 hypoxia/anaerobiosis, when nutrients are limited (33, 85, 86) (Figure 5). Succinate
338 dehydrogenases are a direct link between the TCA cycle and electron transport and typically
339 reduce succinate to fumarate (87).

340 The only potentially harmful mutations we detected in the Succinate dehydrogenases were
341 in *sdhA* of SDH2 in a clade of MAF L6 and in SDH1, *Rv0249c*, in a clade of MTB L1 (Figure 4).
342 Given the essential role of SDH1 for growth and survival, where the deletion of SDH1 was

343 shown to impair the rate of respiration through the Electron Transport Chain and to reduce
344 cell viability (85), future studies should determine if electron transport and aerobic respiration
345 are affected in certain MTB L1 strains, as recent studies report slower growth and a lower
346 odds to grow in culture for MTB L1 and the MAF lineages (67, 79).

347 **3.2.3 Cytochrome bc1-aa3 complexes and Cytochrome bd oxidase**

348 Electrons from Succinate dehydrogenases, move into the quinone pool, ready for transfer to
349 the third and fourth complexes Cytochrome bc1-aa3 (*qcrABC* and *ctaBCDE*), during aerobic
350 respiration, when oxygen is abundant, and to the less efficient cytochrome bd oxidase
351 (*cydABCD*) when oxygen is limited (Figure 5).

352 The bc1-aa3 complexes are the major respiratory route in mycobacteria under standard
353 aerobic conditions and are essential for growth where they play a key role in oxidative
354 phosphorylation and electron transport that yields more ATP (33, 88), yet in our analysis, we
355 detected potentially harmful mutations in *qcrC* and *ctaB* of the bc1-aa3 complexes in MAF
356 lineages and intact gene complexes in MTB lineages (Figure 4 and Figure 5). To the best of our
357 knowledge, this is the first work demonstrating that the critical bc1-aa3 complex is mutated
358 in MAF lineages. This is perhaps crucial to understanding the difference in energy metabolism,
359 particularly oxidative phosphorylation, between the MTB and MAF lineages, especially
360 because MAF lineages preferentially grow microaerobically (15, 84, 89-91) and significantly
361 under-express the dormancy regulon required for adaptation to oxygen limitation (15).
362 Notably, the Imidazopyridine amide in Phase 2 clinical trials, Telacebec (Q203), inhibits *qcrB*
363 of Cytochrome bc1, further emphasizing the importance of this complex for the survival of
364 MTB. Therefore, the predicted harmful mutations we found in *qcrC*, the heme of the
365 cytochrome bc1-complex, in MAF L6 (Figure 4 and Figure 5) likely impairs aerobic respiration
366 and growth in this lineage severely, given that *qcrC*, like *qcrB*, is essential for survival (86).
367 Similarly, *M. smegmatis* strains that had mutations in the bc1-aa3 complex were significantly
368 growth impaired, confirming the essentiality of the bc1-aa3 respiratory pathway for
369 mycobacterial growth. Recently, it was also confirmed that *M. tuberculosis* requires the bc1-
370 aa3 complex to attain optimal growth rates and high titres in mice (83).

371 Taken together, our analysis provides further support for the view that MAF is adapted to a
372 distinct niche, less dependent on aerobic respiration and more adapted to a microaerobic

373 lifestyle (15). Reasons for this potential niche adaptation and the benefit to the pathogen in
374 its interaction with its host should be investigated further.

375 Ultimately, with an impaired bc1-aa3 complex, ATP yield will be reduced overall. This is likely
376 the case in MAF lineages. Therefore, we postulate that ATP yield through oxidative
377 phosphorylation in the MAF lineages is lower compared to the MTB lineages.

378 **3.2.4 ATP Synthase**

379 The F_1-F_0 ATP synthase itself, the target for Bedaquiline (54, 92), is rather conserved in the
380 different lineages. However, in MAF L5 and MTB L2 we detected potentially harmful
381 mutations in *atp* genes (Table 2). The F_1-F_0 ATP synthase operon is encoded by *atpIBEFHAGDC*
382 and is required for survival as all genes in the operon are essential (71, 93). A defective ATP
383 synthase may coincide with the microaerophilic lifestyle of MAF. However, for MTB L2, it is
384 not clear what the advantage is for acquiring mutations in genes encoding ATP synthase.
385 Interestingly though, compared to MTB L4, MTB L2 was found to have a lower growth rate
386 (79).

387 **3.2.5 Fumarate reductase**

388 The branched respiratory chain of *the* MTBC permits anaerobic survival. During hypoxic or
389 anaerobic conditions when oxygen is limited, Fumarate reductase (FRD, *frdABCD*) and Nitrate
390 reductase (NAR, *narGHJI*) can serve as terminal electron acceptors to maintain the membrane
391 potential. Therefore, these enzymes are critical. Moreover, further studies of *sdh* suggested
392 that FRD could partially compensate for a lack of SDH activity (94). Defects in any of the
393 subunits of *frd* could limit the overall function of the FRD complex. Notably, the attenuated
394 strain H37Ra grown under low-oxygen conditions showed a lag in gene expression of *frdA* and
395 *frdB* (95). The mutations we found in *frdA* and *frdB* in MAF L5 may be related to the adaptation
396 of this lineage to a hypoxic or microaerophilic lifestyle. Therefore, the effect of mutations on
397 gene function in MAF L5 and MTB L4 sublineage, Haarlem, should be determined
398 experimentally.

399 **3.2.6 Alanine dehydrogenase and Proline dehydrogenase**

400 Other key dehydrogenases that contribute to redox balance of NADH for initiation of electron
401 transport were also mutated in the MAF lineages (Figure 4 and Figure 5), indicating potential

402 redox imbalance in MAF lineages with the likelihood to impede electron transport. From our
403 analysis, all MAF lineages had potentially harmful mutations in Alanine dehydrogenase (*ald*)
404 and all MAF L5 had potentially harmful mutations in Proline dehydrogenase (*pruB*). Proline
405 dehydrogenase is associated with the adaptation to hypoxia, slow growth rate and is essential
406 for growth (71, 93, 96, 97). Alanine dehydrogenase has been shown to play a role in redox
407 balance during low oxygen conditions and the downshift of *M. tuberculosis* to the state of
408 nonreplicating persistence. *ald* mutants had altered NADH/NAD ratios and significant delays
409 in growth resumption after reaeration. Additionally, induction of *ald* rescued the bc1-aa3
410 complex mutant while its disruption made the growth defect of the mutant worse (97, 98).

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

416

417 **3.3 Additional Mutations and similarities between MTBC lineages in the Central**
418 **Carbon Metabolic Pathway and Electron Transport Chain**

419 We detected several other mutations in genes of the Central Carbon Metabolic pathway and
420 Electron Transport Chain that could potentially impact on enzyme function (Supplementary
421 File S5). However, their PROVEAN scores were above the cut-off for harmful mutations and
422 thus any impact may be slight or only due to neutral evolution. *cydB* and *narG*, highly
423 important genes in the Electron Transport Chain also had several mutations, but none
424 predicted to be potentially harmful (Supplementary File S5).

425 For multiple genes in the Electron Transport Chain and the Central Carbon Metabolic
426 pathway, we either found the same mutation occurring in the same gene or different
427 mutations occurring in the same gene in the different MTBC lineages. These mutations may
428 confer a selective advantage and/or contribute to adaptation (Supplementary File S5).

429 Of these, only those mutations detected in *pca* in MTB L3 and MAF L5, those detected in *icl2a*
430 and those found in *ndhA* are potentially harmful.

431 Overall, more similarities occurred between MTB L1, MTB L3 and the MAF lineages. This is
432 interesting as the MAF lineages, MTB L1 and L3 lineages all reportedly grow relatively slower
433 than the MTB L2 and L4 lineages (14, 67, 79, 80, 99). We postulate that the slower growth
434 rate of these lineages and the similarities we observe in their metabolic pathways, may be
435 linked to their similar migration and dispersal patterns in Africa and Eurasia (100), where L2
436 and L4 have become widely dispersed, while L5, L6, and L7, had more geographically
437 restricted expansion, adapting to more specific hosts. This may have influenced niche
438 adaptation, where L2 and L4, in line with increased dispersion and range expansion, also
439 increased their replicative/growth capacity and ability to transmit, while the more host
440 restricted lineages maintained a lower replicative/growth potential in line with expansion in
441 situ.

442

443 ### 3.4 Limitations

444 Limitations of our analysis include the small sample size of MAF L5 analyzed. In the Gambia
445 and other countries in Western West Africa, the prevalence of L5 is significantly lower than
446 L6 (36). In our isolation of the 289 MAF strains included in this study, only 3 from The Gambia
447 were MAF L5. However, to ensure that the mutations we detected in L5 from The Gambia
448 were more representative of the MAF L5 lineage, we included L5 from Nigeria (Eastern West
449 Africa), where the prevalence of L5 is high and L6 is significantly less likely to be isolated (101).
450 Another limitation is that we did not experimentally confirm any of our in silico phenotype
451 predictions.

452

453 ## 4 Conclusion

454 In this comparative analysis, we describe genomic differences between the MTB and MAF
455 lineages in genes encoding enzymes of the Electron Transport Chain and the Central Carbon
456 Metabolic pathway, which may explain the differences in the clinical- and in vitro phenotype
457 described for the MAF and MTB lineages. In vitro, MAF lineages grow significantly slower than
458 MTB lineages and MAF L6 is, clinically, less virulent than MTB L4, as evidenced by significantly
459 lower progression of MAF L6 infected individuals to active TB disease (11). Again, in vitro, MAF

460 lineages show microaerobic growth and clinically, are associated with extrapulmonary
461 disease, implying a preference for regions with low oxygen (10, 15). Furthermore, MAF L6
462 more commonly causes disease in immunocompromised persons, implying a more
463 opportunistic pathogen (12, 13). Generally, it appears from our analysis that compared to
464 MTB lineages, MAF lineages had the most mutated Central Carbon Metabolic and Electron
465 Transport Pathways, with mutations occurring in critical components of each pathway. The
466 combined effect of a defective Carbon Metabolic Pathway and Electron Transport Chain in
467 MAF lineages, likely contributes to the reduced fitness of the MAF lineages. We speculate that
468 our findings may contribute to 1 - the slower growth of MAF lineages, 2 - relative attenuation
469 of the MAF L6 lineage compared to MTB lineages and 3 - host specificity to West Africans.

470 It is intriguing that in the different MTBC lineages multiple harmful mutations occurred in the
471 same gene. These similarities largely found in MTB lineages 1, 2, 3 and the MAF lineages,
472 indicates there may be a selective advantage for this. Interestingly, these lineages, compared
473 to MTB L4 strains, have been associated with slower growth and cytokine induction patterns
474 suggestive of immune evasion (8, 14, 67, 79-81, 99, 102, 103).

475 If the potentially harmful mutations we report on in this analysis affect energy metabolism in
476 the different MTBC lineages, fitness will differ and ultimately the infection and transmission
477 potential of these lineages. Therefore, our findings are not only relevant for TB product
478 development but also for transmission studies and interventions. The literature already
479 provides credence to and evidence for our hypothesis (9, 81, 104, 105).

480 **4.1 Future Proposed Investigations**

481 **4.1.1 Functional Studies**

482 Significant genomic differences between MTB and MAF lineages presented in this analysis
483 warrant further studies in order to properly characterize the regulatory network of MAF.
484 Recently, the regulatory network of H37Rv (MTB L4) was characterized yet what pertains to
485 MAF lineages is not clearly defined, even though it has been shown that master regulators
486 like *PhoPR* and *Rv0081* as well as the DosR regulon are underexpressed and/or mutated in
487 MAF (15, 106-108). It is possible that the PhoPR system controls other components of the
488 respiratory and central metabolic pathway although this is yet to be shown. We suspect that
489 the potentially harmful mutations we report on will produce different growth phenotypes

490 (Table 3). Therefore, we propose functional genomic assays followed subsequently by in vitro
491 and in vivo characterization of mutants to confirm the contribution to growth and survival
492 that each gene makes. Such studies on the MAF lineages are limited, but not for MTB,
493 particularly the generalist MTB L4. Recent studies provide data describing functional
494 consequences of synonymous SNPs (109-111), i.e. caution needs to be taken in inferring the
495 relative significance or impact of observed genomic mutations from sequencing data alone.
496 Therefore, further studies to correlate genotype with phenotype are necessary and our
497 findings serve as a prelude to such experimental studies.

498 **4.1.2 Investigate Bioenergetics in MAF lineages**

499 It is highly plausible that adaptation of MAF lineages to growth under low oxygen
500 (microaerobic) conditions could be a strategy to escape the harmful effects of ROS
501 generated as electrons leak from a defective respiratory chain and react with built up
502 oxygen (112). Moreover, reduced aerobic respiration or oxygen consumption in the MAF
503 lineages could potentially affect the sensitivity of rapid diagnostics like the automated MGIT
504 960 system, that depends on oxygen consumption to detect growth (113). Interestingly, the
505 percentage of MAF in certain parts of West and Central Africa was reported to have
506 suddenly and sharply declined over the last decade (4, 114-117), which could possibly be
507 due to the introduction and use of new diagnostics like the MGIT 960 system. Of note, we
508 recently detected, in a retrospective analysis based on genotyping, that 84% of strains that
509 did not grow in the MGIT 960 system were MAF L6 (Ofori-Anyinam *et al*., unpublished).
510 Since growth and survival, driven by nutrient and energy metabolism, sustain pathogen
511 transmission, and given that the success of drugs and phenotypic diagnostics rely on
512 metabolic properties of the bacteria, understanding metabolism in these bacteria is
513 essential. We hypothesize that MAF lineages generate less ATP and are more exposed to
514 ROS than MTB lineages (Table 3). Therefore, we propose studies to investigate energy
515 metabolism in MAF lineages relative to the MTB lineages, including whether MAF lineages
516 are more exposed to ROS during growth.

517 **4.1.3 Host Genetics**

518 One common feature of host specificity is genomic decay. Genomes of specialists usually
519 show signs of genome decay evidenced by gene deletions or gene inactivation via point
520 mutations. A driving force for the accumulation of mutations in some key electron transport

521 and carbon metabolism genes in the specialist/host-restricted MAF lineages could indicate
522 adaptation of the MAF lineages to a specific host, West Africans, or a niche within the host,
523 due to altered or loss of function of these genes. This potential niche adaptation could be
524 driven by a precise feature of the host environment that favors the association between MAF
525 lineages and West Africans such as has been described for some diseases (118). Some findings
526 have been made, including MAF lineage-specific mutations in genes and pseudogenes
527 involved in vitamin B12 and vitamin B3 metabolism, important cofactor biosynthetic
528 pathways for many cellular functions. Unlike MAF though, MTB is fully capable of synthesizing
529 vitamin B12. Therefore, it has been suggested that the mutations in the vitamin B12 pathway
530 of the MAF lineages may affect their host range to West Africans. According to a study in the
531 United States, black persons reportedly have higher levels of vitamin B12 relative to other
532 ethnicities (8, 10, 119-121). Future studies should investigate the molecular mechanisms
533 underlying host specificity.

534

535 **5 Supplementary Information**

536 Supplementary File S1

537 Supplementary File S2

538 Supplementary File S3

539 Supplementary File S4

540 Supplementary File S5

541

542 **6 Author Contributions**

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

7 Competing financial interests

The authors declare no competing financial interests.

8 Funding

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

9 Acknowledgments

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

References

1. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proceedings of the national academy of Sciences. 2002;99(6):3684-9.
2. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nature genetics. 2013;45(10):1176-82.
3. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(8):2869-73.
4. Michel KG, Guislaine R, Serge PD, Sanou A, Antoinette K, Adama O, et al. Spoligotyping of *Mycobacterium africanum*, Burkina Faso. Emerging Infectious Disease journal. 2012;18(1):117.
5. Coscolla M, Gagneux S. Consequences of genomic diversity in *Mycobacterium tuberculosis*. Semin Immunol. 2014;26.

575 6. Bentley SD, Comas I, Bryant JM, Walker D, Smith NH, Harris SR. The Genome of
576 *Mycobacterium Africanum* West African 2 Reveals a Lineage-Specific Locus and Genome Erosion
577 Common to the M. tuberculosis Complex. PLoS Negl Trop Dis. 2012;6.
578 7. Mostowy S, Onipede A, Gagneux S, Niemann S, Kremer K, Desmond EP, et al. Genomic
579 analysis distinguishes *Mycobacterium africanum*. Journal of clinical microbiology. 2004;42(8):3594-9.
580 8. Winglee K, Manson McGuire A, Maiga M, Abeel T, Shea T, Desjardins CA, et al. Whole
581 Genome Sequencing of *Mycobacterium africanum* Strains from Mali Provides Insights into the
582 Mechanisms of Geographic Restriction. PLoS neglected tropical diseases. 2016;10(1):e0004332.
583 9. Asare P, Asante-Poku A, Prah DA, Borrell S, Osei-Wusu S, Otchere ID, et al. Reduced
584 transmission of *Mycobacterium africanum* compared to *Mycobacterium tuberculosis* in urban West
585 Africa. International Journal of Infectious Diseases. 2018;73:30-42.
586 10. Sharma A, Bloss E, Heilig CM, Click ES. Tuberculosis Caused by *Mycobacterium africanum*,
587 United States, 2004-2013. Emerging infectious diseases. 2016;22(3):396-403.
588 11. De Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM. Progression to Active
589 Tuberculosis, but Not Transmission, Varies by *Mycobacterium tuberculosis* Lineage in The Gambia. J
590 Infect Dis. 2008;198.
591 12. de Jong BC, Hill PC, Brookes RH, Otu JK, Peterson KL, Small PM, et al. *Mycobacterium*
592 *africanum*: a new opportunistic pathogen in HIV infection? Aids. 2005;19(15):1714-5.
593 13. de Jong BC, Adetifa I, Walther B, Hill PC, Antonio M, Ota M, et al. Differences between
594 tuberculosis cases infected with *Mycobacterium africanum*, West African type 2, relative to Euro-
595 American *Mycobacterium tuberculosis*: an update. FEMS immunology and medical microbiology.
596 2010;58(1):102-5.
597 14. Gehre F, Otu J, DeRiemer K, de Sessions PF, Hibberd ML, Mulders W, et al. Deciphering the
598 growth behaviour of *Mycobacterium africanum*. PLoS neglected tropical diseases. 2013;7(5):e2220.
599 15. Ofori-Anyinam B, Dolganov G, Van T, Davis JL, Walter ND, Garcia BJ, et al. Significant under
600 expression of the DosR regulon in *M. tuberculosis* complex lineage 6 in sputum. Tuberculosis.
601 2017;104:58-64.
602 16. Leistikow RL, Morton RA, Bartek IL, Frimpong I, Wagner K, Voskuil MI. The *Mycobacterium*
603 *tuberculosis* DosR regulon assists in metabolic homeostasis and enables rapid recovery from
604 nonrespiring dormancy. Journal of bacteriology. 2010;192(6):1662-70.
605 17. Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList--10 years after. Tuberculosis
606 (Edinburgh, Scotland). 2011;91(1):1-7.
607 18. Cook GM, Berney M, Gebhard S, Heinemann M, Cox RA, Danilchanka O, et al. Physiology of
608 mycobacteria. Advances in microbial physiology. 2009;55:81-319.
609 19. Warner DF. *Mycobacterium tuberculosis* metabolism. Cold Spring Harbor perspectives in
610 medicine.5(4):a021121.
611 20. David HL, Jahan MT, Jumin A, Grandry J, Lehman EH. Numerical taxonomy analysis of
612 *Mycobacterium africanum*. Int J Syst Bacteriol. 1978;28.
613 21. Dubos RJ. The effect of lipids and serum albumin on bacterial growth. The Journal of
614 experimental medicine. 1947;85(1):9.
615 22. Edson N. The intermediary metabolism of the mycobacteria. Bacteriological reviews.
616 1951;15(3):147.
617 23. Goldman DS. ENZYME SYSTEMS IN THE MYCOBACTERIA XV. Initial Steps in the Metabolism of
618 Glycerol. Journal of bacteriology. 1963;86(1):30-7.
619 24. Hunter G. The oxidation of glycerol by mycobacteria. Biochemical Journal. 1953;55(2):320.
620 25. Merrill MH. Studies on Carbon Metabolism of Organisms of the Genus Mycobacterium: II.
621 Utilization of Organic Compounds in a Synthetic Medium. Journal of bacteriology. 1931;21(5):361.
622 26. Pattyn SR, Portaels F, Spanoghe L, Magos J. Further studies on African strains of
623 *Mycobacterium tuberculosis*: comparison with *M. bovis* and *M.microti*. Annales des Societes belges
624 de medecine tropicale, de parasitologie, et de mycologie. 1970;50.

625 27. Tepper BS. Differences in the utilization of glycerol and glucose by *Mycobacterium phlei*.
626 Journal of bacteriology. 1968;95(5):1713-7.
627 28. Youmans AS, Youmans GP. STUDIES ON THE METABOLISM OF MYCOBACTERIUM
628 TUBERCULOSIS IV.: The Effect of Fatty Acids on the Growth of *M. tuberculosis* var. hominis1. Journal
629 of bacteriology. 1954;67(6):731.
630 29. Youmans GP, Youmans AS. STUDIES ON THE METABOLISM OF MYCOBACTERIUM
631 TUBERCULOSIS I.: The Effect of Carbohydrates and Alcohols on the Growth of *Mycobacterium*
632 *tuberculosis* var. hominis1. Journal of bacteriology. 1953;65(1):92.
633 30. Pattyn SR, Portaels F, Spanoghe L, Magos J. Further studies on African strains of
634 *Mycobacterium tuberculosis*: comparison with *M. bovis* and *M. microti*. Annales des Societes belges
635 de medecine tropicale, de parasitologie, et de mycologie. 1970;50(2):211-27.
636 31. de Medeiros D, Castets M, Diop S. [Comparative morphological aspects of *Mycobacterium*
637 *tuberculosis*, classical type, and *Mycobacterium africanus* (preliminary note)]. Bulletin de la Societe
638 medicale d'Afrique noire de langue francaise. 1973;18(4):477-84.
639 32. Baughn AD, Rhee KY. Metabolomics of Central Carbon Metabolism in *Mycobacterium*
640 *tuberculosis*. Microbiology spectrum. 2014;2(3).
641 33. Cook GM, Hards K, Vilcheze C, Hartman T, Berney M. Energetics of Respiration and Oxidative
642 Phosphorylation in Mycobacteria. Microbiology spectrum. 2014;2(3).
643 34. Keating LA, Wheeler PR, Mansoor H, Inwald JK. The pyruvate requirement of some members
644 of the *Mycobacterium tuberculosis* complex is due to an inactive pyruvate kinase: implications for in
645 vivo growth. Mol Microbiol. 2005;56.
646 35. Gehre F, Otu J, Kendall L, Forson A, Kwara A, Kudzawu S, et al. The emerging threat of pre-
647 extensively drug-resistant tuberculosis in West Africa: preparing for large-scale tuberculosis research
648 and drug resistance surveillance. BMC Medicine. 2016;14(1):160.
649 36. Gehre F, Kumar S, Kendall L, Ejo M, Secka O, Ofori-Anyinam B, et al. A Mycobacterial
650 Perspective on Tuberculosis in West Africa: Significant Geographical Variation of *M. africanum* and
651 Other M. tuberculosis Complex Lineages. PLoS neglected tropical diseases. 2016;10(3):e0004408.
652 37. Singhal N, Sharma P, Kumar M, Joshi B, Bisht D. Analysis of intracellular expressed proteins
653 of *Mycobacterium tuberculosis* clinical isolates. Proteome Science. 2012;10(1):14.
654 38. Hosek J, Svastova P, Moravkova M, Pavlik I, Bartos M. Methods of mycobacterial DNA
655 isolation from different biological material: a review. Veterinarni medicina. 2006;51(5):180-92.
656 39. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv
657 preprint arXiv:13033997. 2013.
658 40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
659 Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009;25(16):2078-9.
660 41. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv
661 preprint arXiv:12073907. 2012.
662 42. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating
663 and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
664 Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80-92.
665 43. Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, Viveiros M, et al. A robust SNP
666 barcode for typing *Mycobacterium tuberculosis* complex strains. Nature communications.
667 2014;5:4812.
668 44. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a Web
669 Tool Delineating *Mycobacterium tuberculosis* Antibiotic Resistance and Lineage from Whole-Genome
670 Sequencing Data. Journal of clinical microbiology. 2015;53(6):1908-14.
671 45. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
672 phylogenies. Bioinformatics (Oxford, England). 2014;30(9):1312-3.
673 46. Leache AD, Banbury BL, Felsenstein J, de Oca AN, Stamatakis A. Short Tree, Long Tree, Right
674 Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. Systematic
675 biology. 2015;64(6):1032-47.

676 47. Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological
677 character data. Systematic biology. 2001;50(6):913-25.
678 48. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display
679 and annotation. Bioinformatics (Oxford, England). 2007;23(1):127-8.
680 49. Choi Y. A fast computation of pairwise sequence alignment scores between a protein and a
681 set of single-locus variants of another protein. Proceedings of the ACM Conference on
682 Bioinformatics, Computational Biology and Biomedicine; Orlando, Florida. 2382989: ACM; 2012. p.
683 414-7.
684 50. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid
685 substitutions and indels. Bioinformatics (Oxford, England). 2015;31(16):2745-7.
686 51. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid
687 Substitutions and Indels. PloS one. 2012;7(10):e46688.
688 52. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on
689 protein function using the SIFT algorithm. Nature protocols. 2009;4(7):1073-81.
690 53. Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al. *Mycobacterium tuberculosis*
691 lineage 4 comprises globally distributed and geographically restricted sublineages. Nature genetics.
692 2016;48(12):1535-43.
693 54. Andries K, Verhasselt P, Guillemont J, Gohlmann HW, Neefs JM, Winkler H, et al. A
694 diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. Science (New York,
695 NY). 2005;307(5707):223-7.
696 55. Koul A, Dendouga N, Vergauwen K, Molenberghs B, Vranckx L, Willebrords R, et al.
697 Diarylquinolines target subunit c of mycobacterial ATP synthase. Nature chemical biology.
698 2007;3(6):323-4.
699 56. Pethe K, Bifani P, Jang J, Kang S, Park S, Ahn S, et al. Discovery of Q203, a potent clinical
700 candidate for the treatment of tuberculosis. Nat Med. 2013;19(9):1157-60.
701 57. Barry VC, Belton JG, Conalty ML, Denneny JM, Edward DW, O'Sullivan JF, et al. A new series
702 of phenazines (rimino-compounds) with high antituberculosis activity. Nature. 1957;179(4568):1013-
703 5.
704 58. Lechartier B, Cole ST. Mode of Action of Clofazimine and Combination Therapy with
705 Benzothiazinones against <span class="named-content genus-species"
706 id="named-content-1">*Mycobacterium tuberculosis*</span&gt. Antimicrobial
707 Agents and Chemotherapy. 2015;59(8):4457.
708 59. Lewis JM, Sloan DJ. The role of delamanid in the treatment of drug-resistant tuberculosis.
709 Therapeutics and clinical risk management. 2015;11:779-91.
710 60. Xavier AS, Lakshmanan M. Delamanid: A new armor in combating drug-resistant
711 tuberculosis. Journal of pharmacology & pharmacotherapeutics. 2014;5(3):222-4.
712 61. Yano T, Kassovska-Bratinova S, Teh JS, Winkler J, Sullivan K, Isaacs A, et al. Reduction of
713 clofazimine by mycobacterial type 2 NADH:quinone oxidoreductase: a pathway for the generation of
714 bactericidal levels of reactive oxygen species. The Journal of biological chemistry.
715 2011;286(12):10276-87.
716 62. Van Deun A, Decroo T, Piubello A, de Jong BC, Lynen L, Rieder HL. Principles for constructing
717 a tuberculosis treatment regimen: the role and definition of core and companion drugs. The
718 international journal of tuberculosis and lung disease : the official journal of the International Union
719 against Tuberculosis and Lung Disease. 2018;22(3):239-45.
720 63. Cohen T, Colijn C, Murray M. Modeling the effects of strain diversity and mechanisms of
721 strain competition on the potential performance of new tuberculosis vaccines. Proceedings of the
722 National Academy of Sciences. 2008;105(42):16302.
723 64. Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and
724 implications for tuberculosis product development. Lancet Infect Dis. 2007;7.

725 65. Homolka S, Niemann S, Russell DG, Rohde KH. Functional Genetic Diversity among
726 *Mycobacterium tuberculosis* Complex Clinical Isolates: Delineation of Conserved Core and Lineage-
727 Specific Transcriptomes during Intracellular Survival. PLoS pathogens. 2010;6(7):e1000988.
728 66. Ofori-Anyinam B, Kanuteh F, Agbla SC, Adetifa I, Okoi C, Dolganov G, et al. Impact of the
729 *Mycobaterium africanum* West Africa 2 Lineage on TB Diagnostics in West Africa: Decreased
730 Sensitivity of Rapid Identification Tests in The Gambia. PLoS neglected tropical diseases.
731 2016;10(7):e0004801.
732 67. Sanoussi CN, Affolabi D, Rigouts L, Anagonou S, de Jong B. Genotypic characterization
733 directly applied to sputum improves the detection of *Mycobacterium africanum* West African 1,
734 under-represented in positive cultures. PLoS neglected tropical diseases. 2017;11(9):e0005900.
735 68. Sanoussi CN, de Jong BC, Odoun M, Arekpa K, Ali Ligali M, Bodi O, et al. Low sensitivity of the
736 MPT64 identification test to detect lineage 5 of the *Mycobacterium tuberculosis* complex. Journal of
737 medical microbiology. 2018;67(12):1718-27.
738 69. Iqbal IK, Bajeli S, Akela AK, Kumar A. Bioenergetics of Mycobacterium: An Emerging
739 Landscape for Drug Discovery. Pathogens. 2018;7(1):24.
740 70. Rahi A, Matta SK, Dhiman A, Garhyan J, Gopalani M, Chandra S, et al. Enolase of
741 *Mycobacterium tuberculosis* is a surface exposed plasminogen binding protein. Biochimica et
742 biophysica acta General subjects. 2017;1861(1 Pt A):3355-64.
743 71. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution
744 phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism.
745 PLoS pathogens. 2011;7(9):e1002251.
746 72. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high
747 density mutagenesis. Molecular microbiology. 2003;48(1):77-84.
748 73. Tuckman D, Donnelly RJ, Zhao FX, Jacobs WR, Jr., Connell ND. Interruption of the
749 phosphoglucose isomerase gene results in glucose auxotrophy in *Mycobacterium smegmatis*. Journal
750 of bacteriology. 1997;179(8):2724-30.
751 74. Phong WY, Lin W, Rao SP, Dick T, Alonso S, Pethe K. Characterization of phosphofructokinase
752 activity in *Mycobacterium tuberculosis* reveals that a functional glycolytic carbon flow is necessary to
753 limit the accumulation of toxic metabolic intermediates under hypoxia. PloS one. 2013;8(2):e56037.
754 75. Collins DM, Wilson T, Campbell S, Buddle BM, Wards BJ, Hotter G, et al. Production of
755 avirulent mutants of *Mycobacterium bovis* with vaccine properties by the use of illegitimate
756 recombination and screening of stationary-phase cultures. Microbiology (Reading, England).
757 2002;148(Pt 10):3019-27.
758 76. Liu K, Yu J, Russell DG. pckA-deficient *Mycobacterium bovis* BCG shows attenuated virulence
759 in mice and in macrophages. Microbiology (Reading, England). 2003;149(Pt 7):1829-35.
760 77. Munoz-Elias EJ, McKinney JD. *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly
761 required for in vivo growth and virulence. Nat Med. 2005;11(6):638-44.
762 78. Nandakumar M, Nathan C, Rhee KY. Isocitrate lyase mediates broad antibiotic tolerance in
763 *Mycobacterium tuberculosis*. Nature communications. 2014;5:4306.
764 79. Sarkar R, Lenders L, Wilkinson KA, Wilkinson RJ, Nicol MP. Modern Lineages of
765 *Mycobacterium tuberculosis* Exhibit Lineage-Specific Patterns of Growth and Cytokine Induction in
766 Human Monocyte-Derived Macrophages. PloS one. 2012;7(8):e43170.
767 80. Tanveer M, Hasan Z, Kanji A, Hussain R, Hasan R. Reduced TNF-α and IFN-γ responses to
768 Central Asian strain 1 and Beijing isolates of *Mycobacterium tuberculosis* in comparison with H37Rv
769 strain. Transactions of The Royal Society of Tropical Medicine and Hygiene. 2009;103(6):581-7.
770 81. Yimer SA, Norheim G, Namouchi A, Zegeye ED, Kinander W, Tønjum T, et al. <span
771 class="named-content genus-species" id="named-content-
772 1">Mycobacterium tuberculosis Lineage 7 Strains Are Associated with
773 Prolonged Patient Delay in Seeking Treatment for Pulmonary Tuberculosis in Amhara Region,
774 Ethiopia. Journal of clinical microbiology. 2015;53(4):1301.

82. Vilcheze C, Weinrick B, Leung LW, Jacobs WR, Jr. Plasticity of *Mycobacterium tuberculosis* NADH dehydrogenases and their role in virulence. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(7):1599-604.
83. Beites T, O'Brien K, Tiwari D, Engelhart CA, Walters S, Andrews J, et al. Plasticity of the *Mycobacterium tuberculosis* respiratory chain and its impact on tuberculosis drug development. Nature communications. 2019;10(1):4970-.
84. De Jong BC, Antonio M, Gagneux S. *Mycobacterium africanum*—Review of an Important Cause of Human Tuberculosis in West Africa. PLoS Negl Trop Dis. 2010;4.
85. Hartman T, Weinrick B, Vilchèze C, Berney M, Tufariello J, Cook GM, et al. Succinate dehydrogenase is the regulator of respiration in *Mycobacterium tuberculosis*. PLoS pathogens. 2014;10(11):e1004510.
86. Minato Y, Gohl DM, Thiede JM, Chacón JM, Harcombe WR, Maruyama F, et al. Genome-wide assessment of *Mycobacterium tuberculosis* conditionally essential metabolic pathways. bioRxiv. 2019:534289.
87. Goldberg I, Lonberg-Holm K, Bagley EA, Stieglitz B. Improved conversion of fumarate to succinate by *Escherichia coli* strains amplified for fumarate reductase. Applied and environmental microbiology. 1983;45(6):1838-47.
88. Matsoso LG, Kana BD, Crellin PK, Lea-Smith DJ, Pelosi A, Powell D, et al. Function of the cytochrome bc1-aa3 branch of the respiratory network in mycobacteria and network adaptation occurring in response to its disruption. Journal of bacteriology. 2005;187(18):6300-8.
89. Frothingham R, Strickland PL, Bretzel G, Ramaswamy S, Musser JM, Williams DL. Phenotypic and Genotypic Characterization of *Mycobacterium africanum* Isolates from West Africa. Journal of clinical microbiology. 1999;37(6):1921-6.
90. Niemann S, Richter E, Rusch-Gerdes S. Differentiation among members of the *Mycobacterium tuberculosis* complex by molecular and biochemical features: evidence for two pyrazinamide-susceptible subtypes of M. bovis. Journal of clinical microbiology. 2000;38(1):152-7.
91. Parsons LM, Brosch R, Cole ST, Somoskovi A, Loder A, Bretzel G, et al. Rapid and simple approach for identification of *Mycobacterium tuberculosis* complex isolates by PCR-based genomic deletion analysis. Journal of clinical microbiology. 2002;40(7):2339-45.
92. Trauner A, Sassetti CM, Rubin EJ. Genetic Strategies for Identifying New Drug Targets. Microbiology spectrum. 2014;2(4):Mgm2-0030-2013.
93. Zhang YJ, Ioerger TR, Huttenhower C, Long JE, Sassetti CM, Sacchettini JC, et al. Global Assessment of Genomic Regions Required for Growth in *Mycobacterium tuberculosis*. PLoS pathogens. 2012;8(9):e1002946.
94. Guest JR. Partial replacement of succinate dehydrogenase function by phage- and plasmid-specified fumarate reductase in *Escherichia coli*. Journal of general microbiology. 1981;122(2):171-9.
95. Li AH, Lam WL, Stokes RW. Characterization of genes differentially expressed within macrophages by virulent and attenuated *Mycobacterium tuberculosis* identifies candidate genes involved in intracellular growth. Microbiology (Reading, England). 2008;154(8):2291-303.
96. Berney M, Cook GM. Unique flexibility in energy metabolism allows mycobacteria to combat starvation and hypoxia. PloS one. 2010;5(1):e8614.
97. Giffin MM, Shi L, Gennaro ML, Sohaskey CD. Role of alanine dehydrogenase of *Mycobacterium tuberculosis* during recovery from hypoxic nonreplicating persistence. PloS one. 2016;11(5):e0155522.
98. Jeong J-A, Park SW, Yoon D, Kim S, Kang H-Y, Oh J-I. Roles of Alanine Dehydrogenase and Induction of Its Gene in *Mycobacterium smegmatis* under Respiration-Inhibitory Conditions. Journal of bacteriology. 2018;200(14):e00152-18.
99. Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M, et al. Clade-Specific Virulence Patterns of <span class="named-content genus-species" id="named-

825 content-1">*Mycobacterium tuberculosis* Complex Strains in Human Primary
826 Macrophages and Aerogenically Infected Mice. mBio. 2013;4(4):e00250-13.
827 100. O'Neill MB, Shockey A, Zarley A, Aylward W, Eldholm V, Kitchen A, et al. Lineage specific
828 histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. Molecular ecology. 2019.
829 101. Molina-Moya B, Gomgnimbou MK, Spinasse L, Obasanya J, Oladimeji O, Dacombe R, et al.
830 *Mycobacterium tuberculosis* complex genotypes circulating in Nigeria based on spoligotyping
831 obtained from Ziehl-Neelsen stained slides extracted DNA. PLoS neglected tropical diseases.
832 2018;12(2):e0006242.
833 102. Ates LS, Dippenaar A, Sayes F, Pawlik A, Bouchier C, Ma L, et al. Unexpected Genomic and
834 Phenotypic Diversity of *Mycobacterium africanum* Lineage 5 Affects Drug Resistance, Protein
835 Secretion, and Immunogenicity. Genome Biology and Evolution. 2018;10(8):1858-74.
836 103. Newton SM, Smith RJ, Wilkinson KA, Nicol MP, Garton NJ, Staples KJ, et al. A deletion
837 defining a common Asian lineage of *Mycobacterium tuberculosis* associates with immune
838 subversion. Proceedings of the National Academy of Sciences of the United States of America.
839 2006;103(42):15594-8.
840 104. Albanna AS, Reed MB, Kotar KV, Fallow A, McIntosh FA, Behr MA, et al. Reduced
841 Transmissibility of East African Indian Strains of *Mycobacterium tuberculosis*. PloS one.
842 2011;6(9):e25075.
843 105. Nebenzahl-Guimaraes H, Verhagen LM, Borgdorff MW, Van Soolingen D. Transmission and
844 progression to disease of *Mycobacterium tuberculosis* phylogenetic lineages in The Netherlands.
845 Journal of clinical microbiology. 2015;53(10):3264-71.
846 106. Broset E, Martín C, Gonzalo-Asensio J. Evolutionary Landscape of the <span
847 class="named-content genus-species" id="named-content-
848 1">*Mycobacterium tuberculosis* Complex from the Viewpoint of PhoPR:
849 Implications for Virulence Regulation and Application to Vaccine Development. mBio.
850 2015;6(5):e01289-15.
851 107. Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, et al.
852 Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence
853 regulator. Proceedings of the National Academy of Sciences of the United States of America.
854 2014;111(31):11491-6.
855 108. Gonzalo-Asensio J, Mostowy S, Harders-Westerveen J, Huygen K, Hernandez-Pando R, Thole
856 J, et al. PhoP: a missing piece in the intricate puzzle of *Mycobacterium tuberculosis* virulence. PloS
857 one. 2008;3(10):e3496.
858 109. Capon F, Allen MH, Ameen M, Burden AD, Tillman D, Barker JN, et al. A synonymous SNP of
859 the corneodesmosin gene leads to increased mRNA stability and demonstrates association with
860 psoriasis across diverse ethnic groups. Human molecular genetics. 2004;13(20):2361-8.
861 110. Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, et al. Synonymous
862 mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the
863 receptor. Human molecular genetics. 2003;12(3):205-16.
864 111. Golby P, Nunez J, Witney A, Hinds J, Quail MA, Bentley S, et al. Genome-level analyses of
865 *Mycobacterium bovis* lineages reveal the role of SNPs and antisense transcription in differential gene
866 expression. BMC genomics. 2013;14:710-.
867 112. Baez A, Shiloach J. Effect of elevated oxygen concentration on bacteria, yeasts, and cells
868 propagated for production of biological compounds. Microbial cell factories. 2014;13:181-.
869 113. Siddiqi SaR-G, S. MGIT™ Procedure Manual, For BACTEC™ MGIT 960™ TB System

870 (Also applicable for Manual MGIT), Mycobacteria Growth Indicator Tube (MGIT)

871 Culture and Drug Susceptibility Demonstration Projects. 2006.
872 114. Godreuil S, Torrea G, Terru D, Chevenet F, Diagbouga S, Supply P, et al. First molecular
873 epidemiology study of *Mycobacterium tuberculosis* in Burkina Faso. Journal of clinical microbiology.
874 2007;45(3):921-7.

115. Groenheit R, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, et al. The Guinea-Bissau family of *Mycobacterium tuberculosis* complex revisited. PloS one. 2011;6(4):e18601.
116. Niobe-Eyangoh SN, Kuaban C, Sorlin P, Cunin P, Thonnon J, Sola C, et al. Genetic Biodiversity of *Mycobacterium tuberculosis* Complex Strains from Patients with Pulmonary Tuberculosis in Cameroon. Journal of clinical microbiology. 2003;41(6):2547-53.
117. Ouassa T, Borroni E, Loukou GY, Faye-Kette H, Kouakou J, Menan H, et al. High prevalence of shared international type 53 among *Mycobacterium tuberculosis* complex strains in retreated patients from Cote d'Ivoire. PloS one. 2012;7(9):e45363.
118. 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.
119. 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.
120. Gagneux S. Strain variation in the *Mycobacterium tuberculosis* complex: its role in biology, epidemiology and control. Springer, Heidelberg. 2017.
121. 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.

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

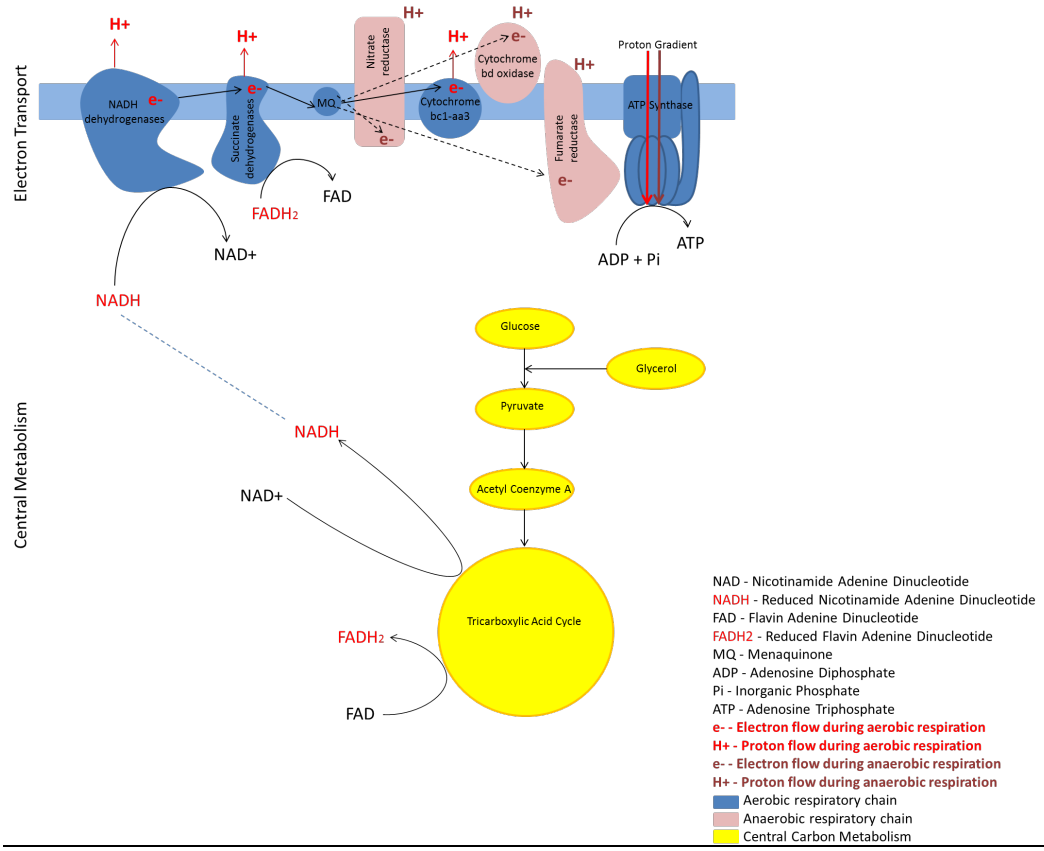

Figure 2: Mutations detected in the Central Carbon Metabolic Pathway of a) MAF L6 and MAF L5 and b) MTB L4 Ghana, L4 Haarlem, L1, L3, L2, L4 Cameroon and L4 LAM. Red and blue coloured bars represent nonsynonymous mutations detected in 50-100% of strains per lineage in the associated genes shown below each set of bars indicating mutations. Genes emboldened red have mutations detected predicted to be harmful in MAF lineages, genes emboldened blue have mutations detected predicted to be harmful in MTB lineages and genes emboldened purple have mutations detected predicted to be harmful in the same gene in MTB and MAF lineages. Grey bars indicate that no nonsynonymous and/or predicted harmful mutations were detected in the corresponding gene below each set of grey bars. First column colour indicates the lineage: a) MAF L6 (green) and MAF L5 (brown), b) MTB L4 Ghana (brown), L4 Haarlem (red), L1 (lilac), L3 (purple), L2 (blue), L4 Cameroon (pink) and L4 LAM (dark red).

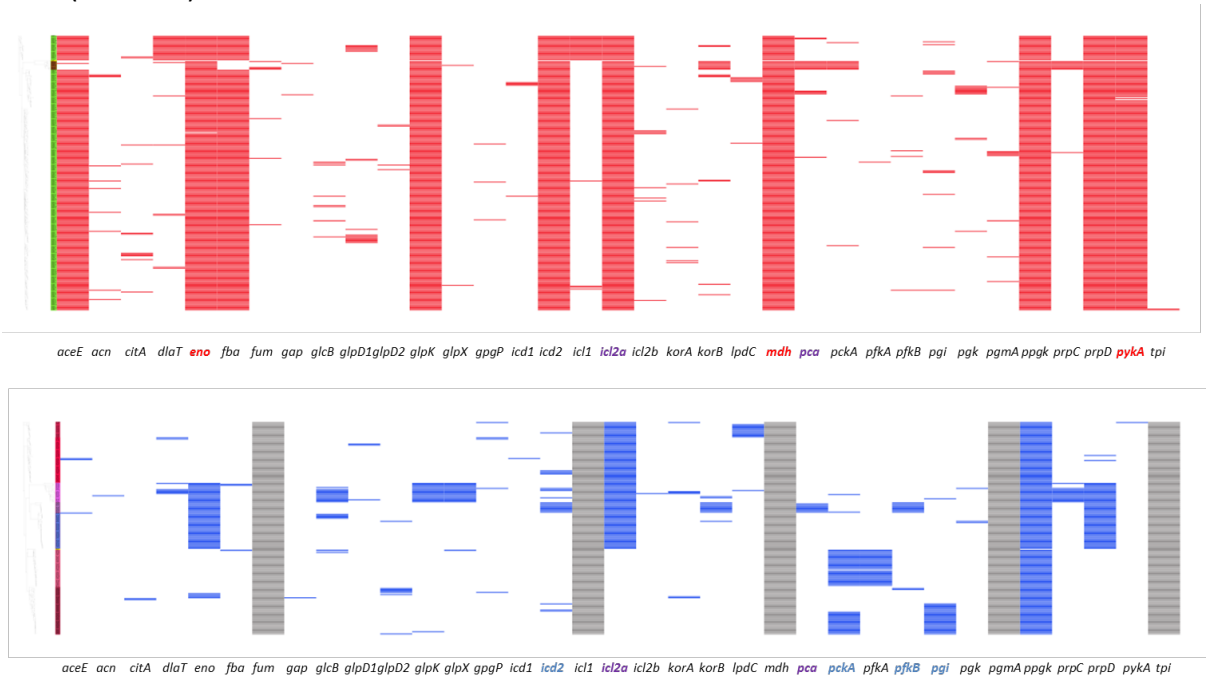

Figure 3: Mutations in the Central Carbon Metabolic Pathway of MAF and MTB. Genes with a **red** colour code have harmful mutations in MAF lineages, indicated next to the gene. Genes with a **blue** colour code have harmful mutations in MTB lineages, while genes in **purple** have harmful mutations in the same gene in MAF and MTB lineages.

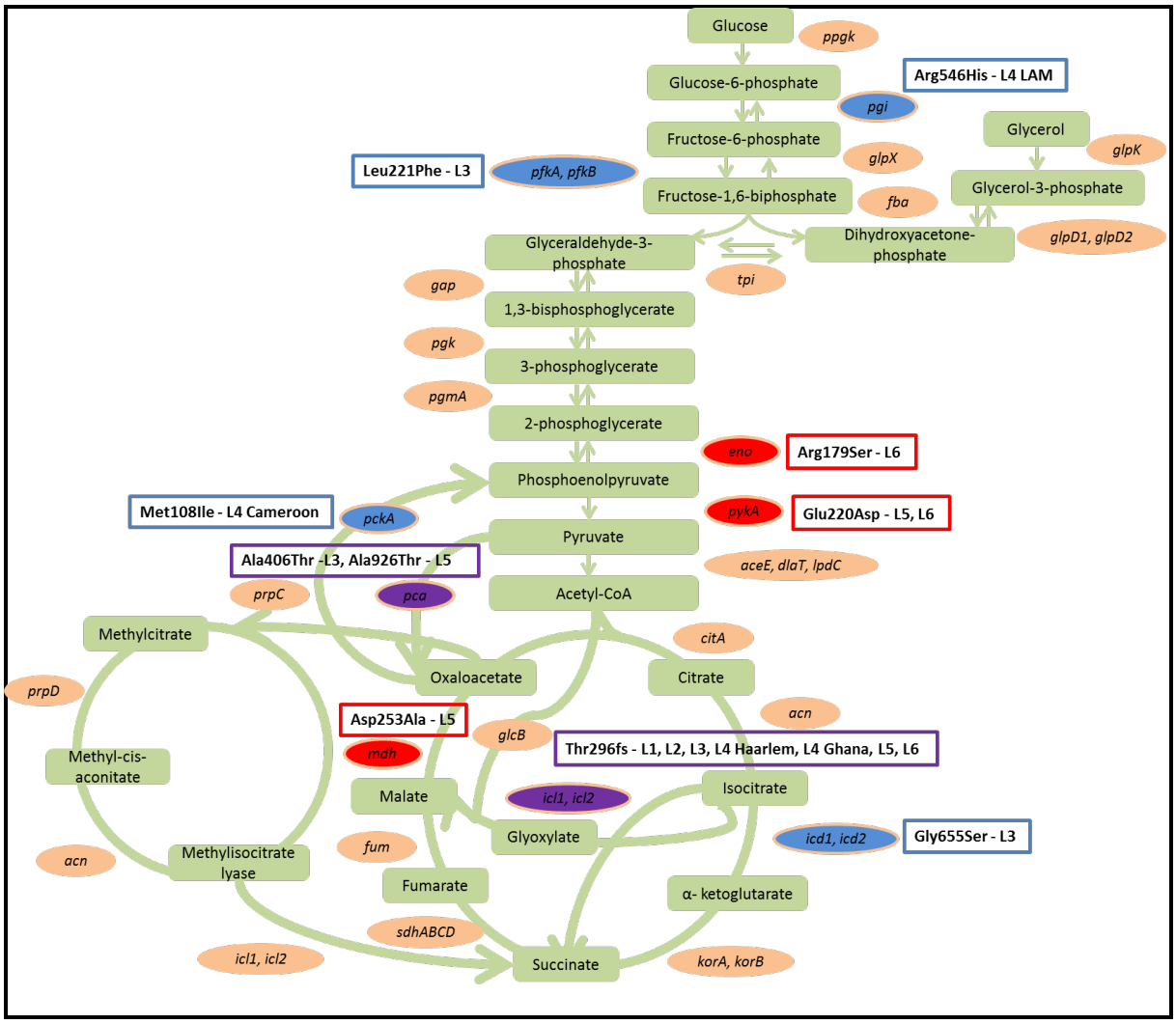

Figure 4: Mutations detected in the Electron Transport Chain of a) MAF L6 and MAF L5 and b) MTB L4 Ghana, L4 Haarlem, L1, L3, L2, L4 Cameroon and L4 LAM. Red and blue coloured bars represent nonsynonymous mutations detected in 50-100% of strains per lineage in associated genes shown below each set of bars indicating mutations. Genes emboldened red have mutations detected predicted to be harmful in MAF lineages, genes emboldened blue have mutations detected predicted to be harmful in MTB lineages and genes emboldened purple have mutations detected predicted to be harmful in the same gene in MTB and MAF lineages. Grey bars indicate that no nonsynonymous and/or predicted harmful mutations were detected in the corresponding gene below each set of grey bars. First column colour indicates the lineage: a) MAF L6 (green) and MAF L5 (brown), b) MTB L4 Ghana (brown), L4 Haarlem (red), L1 (lilac), L3 (purple), L2 (blue), L4 Cameroon (pink) and L4

LAM (dark red).

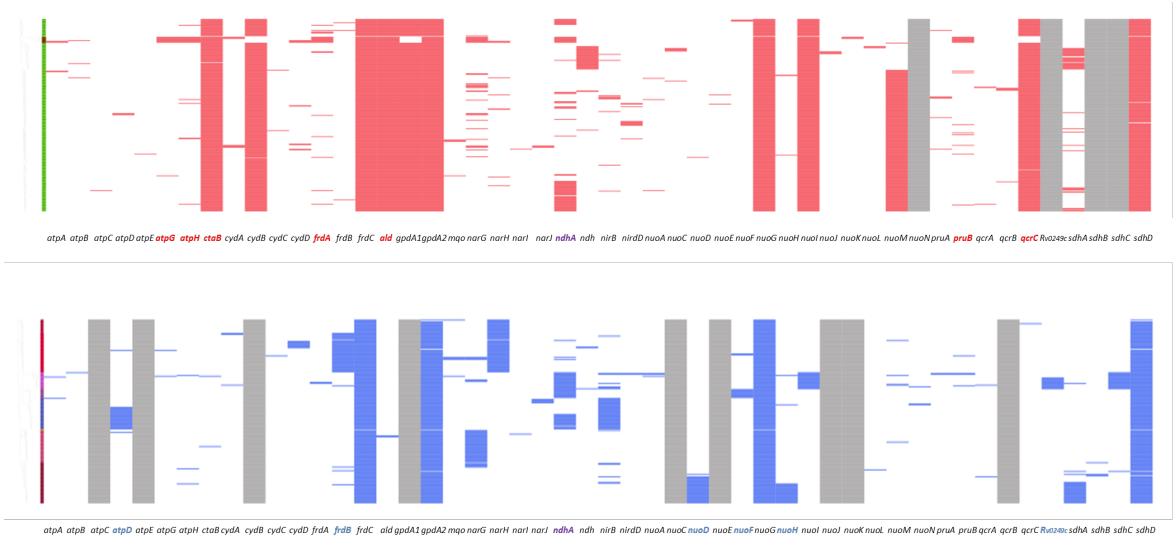

Figure 5: The core and alternate respiratory chain of mycobacteria during in vitro exponential growth when oxygen is abundant and when oxygen is limited. Mutations in MAF lineages are indicated in **red** boxes next to each affected complex. Mutations in MTB are indicated in **blue** boxes next to the affected gene while mutations found in the same gene in both MAF and MTB are shown in **purple** boxes. 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.

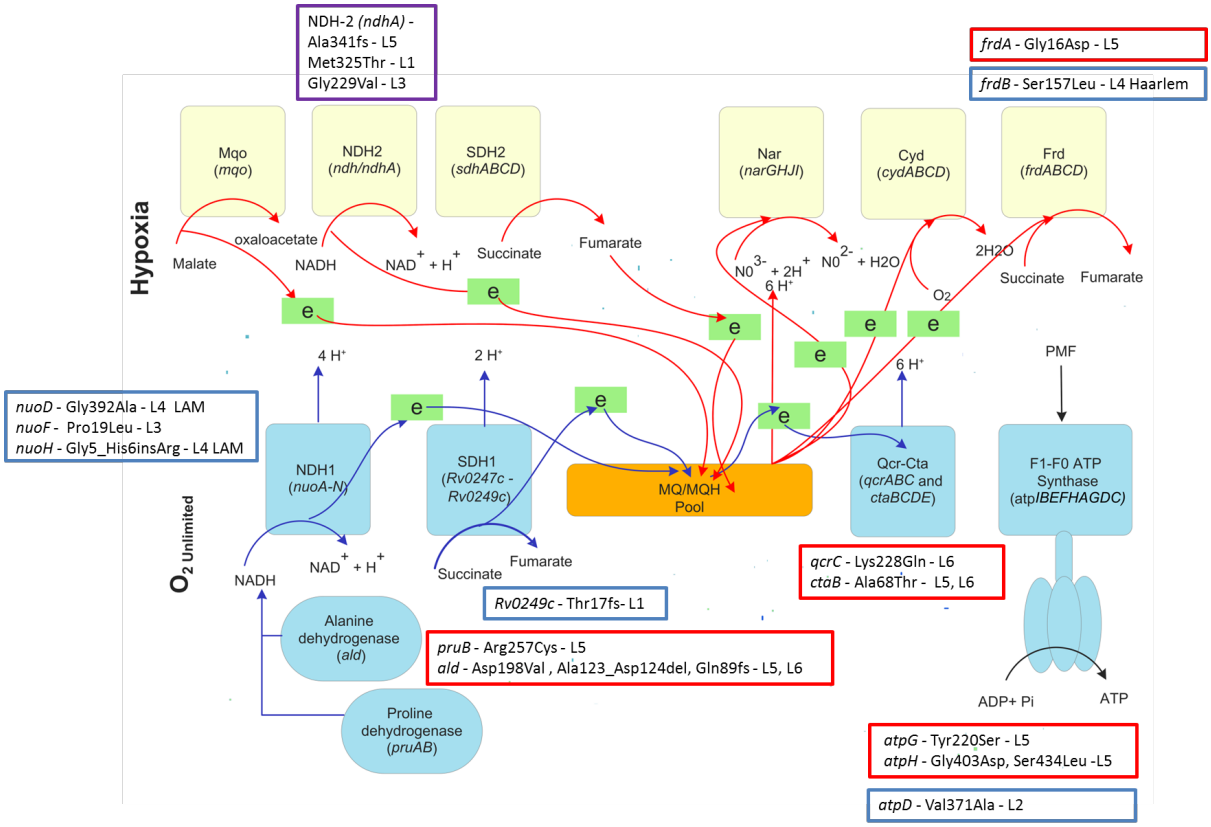
NDH-2 (ndhA) - Ala341fs - L5 Met325Thr - L1 Gly229Val - L3
frdA - Gly16Asp - L5
frdB - Ser157Leu - L4 Haarlem
Hypoxia
Mqo (mqo)
NDH2 (ndh/ndhA)
SDH2 (sdhABCD)
Nar (narGHJI)
Cyd (cydABCD)
Frd (frdABCD)
Malate
oxaloacetate
NADH
NAD+ + H+
Succinate
Fumarate
N03- + 2H+ 6 H+
N02- + H2O
2H2O
O2
Succinate
Fumarate
e
4 H+
2 H+
6 H+
PMF
nuoD - Gly392Ala - L4 LAM
nuoF - Pro19Leu - L3
nuoH - Gly5_His6insArg - L4 LAM
NDH1 (nuoA-N)
SDH1 (Rv0247c - Rv0249c)
MQ/MQH Pool
Qcr-Cta (qcrABC and ctaBCDE)
F1-F0 ATP Synthase (atpIBEFHAGDC)
O2 Unlimited
NADH
NAD+ + H+
Succinate
Fumarate
Rv0249c - Thr17fs- L1
qcrC - Lys228Gln - L6
ctaB - Ala68Thr - L5, L6
Alanine dehydrogenase (ald)
pruB - Arg257Cys - L5
ald - Asp198Val , Ala123_Asp124del, Gln89fs - L5, L6
Proline dehydrogenase (pruAB)
ADP+ Pi
ATP
atpG - Tyr220Ser - L5
atpH - Gly403Asp, Ser434Leu -L5
atpD - Val371Ala - L2

Table 1: Potentially deleterious mutations in genes encoding enzymes in the Central Carbon Metabolic Pathway. List of all mutations found to be potentially harmful by PROVEAN (marked by a *) are listed. The full list of mutations is given in supplementary table S5.

Gene	MTB amino acid change	MTB Lineage	MAF amino acid change	MAF Lineage
pgi*	Arg546His	clade of L4 LAM		
pfkB*	Leu221Phe	L3		
eno*	Arg8Gly Lys429Gln	L1, L2, L3	Arg8Gly Arg179Ser	L5, L6
pykA*			Glu220Asp	L5, L6
pca*	Ala406Thr	L3	Ala926Thr	L5
icd2*	Gly655Ser Leu254Val	L3	Val447Met Lys117Asn	L6
mdh*			Leu326Ile Ala112Thr Asp253Ala	L5, L6
icl2a*	Gly179Asp Thr296fs	L1,L2,L3,L4 Haarlem, L4 Ghana	Ala146Thr His151Arg Gly179Asp Thr296fs	L5, L6
pckA*	Met108Ile Ala536Gly	L4 Cameroon, clade of L4 LAM	Lys422Thr	L5

Table 2: Potentially deleterious mutations in genes and gene subunits encoding Electron Transport Chain enzymes. List of all mutations found to be potentially harmful by PROVEAN (marked by a *). The full list of mutations is given in supplementary table S5.

Gene	MTB amino acid change	MTB Lineage	MAF amino acid change	MAF Lineage
ndhA*	Asn360Ser Met325Thr Gly229Val	clade of L2 L1 L3	Ala341fs	L5
pruB*			Arg257Cys	L5
qcrC*			Lys228Gln	L6
ctaB*			Ala68Thr	L5, L6
cydB			Met98Ile	L6
frdA*			Gly16Asp	L5
frdB*	Ser157Leu	L4 Haarlem		
atpG*			Tyr220Ser	L5
atpH*			Gly403Asp Ser434Leu	L5
atpD*	Val371Ala	L2		
nuoD*	Gly392Ala	clade of L4 LAM		
nuoH*	Gly5_His6insArg	clade of L4 LAM		
nuoF*	Pro19Leu	L3		
Rv0249c*	Thr17fs	clade of L1		
ald*			Gln89fs Asp198Val Ala123_Asp124del	L5, L6

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. Mutagenesis studies would involve introducing the mutations detected into L2, L4 or H37Rv and testing for attenuated growth through growth studies or by complementing MAF and MTB L1 and L3 with a wild-type allele and testing for complementation of growth.	It is expected that the reported mutations will lead to attenuated growth.
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. Mutagenesis studies would involve introducing the mutations detected into L2, L4 or H37Rv and testing for attenuated growth in growth studies or by complementing MAF and MTB L1 and L3 with a wild-type allele and testing for	It is expected that the reported mutations will lead to attenuated growth.

	complementation of growth.	
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.