1 **Title**

2 Genomic insights into adaptations of TMA-utilizing methanogens to diverse habitats

3 including the human gut

4 ***Running title***

5 Insights into *Methanomassiliicoccales* in the human gut

6

7 **Authors**

8 Jacobo de la Cuesta-Zuluaga[1], Timothy D. Spector[2], Nicholas D. Youngblut[1], Ruth

9 E. Ley[1#]

10

11 [1]Department of Microbiome Science, Max Planck Institute for Developmental

12 Biology, 72076 Tübingen, Germany.

13

14 [2]Department of Twin Research and Genetic Epidemiology, King's College London,

15 London SE1 7EH, UK

16

17 [#] correspondence: rley@tuebingen.mpg.de.

1

Abstract

19   Archaea of the order *Methanomassiliicoccales* use methylated-amines such

20   as trimethylamine as a substrate for methane production. They form two large

21   phylogenetic clades and reside in diverse environments, from soil to the human gut.

22   Two genera, one from each clade, inhabit the human gut: *Methanomassiliicoccus*,

23   which has one cultured representative, and "*candidatus* Methanomethylophilus",

24   which has none. Questions remain regarding their distribution across different

25   biomes and human populations, their association with other taxa in the human gut,

26   and whether host genetics correlate with their abundance. To gain insight into the

27   *Methanomassiliicoccales*, and the human-associated members in particular, we

28   performed a genomic comparison of 72 *Methanomassiliicoccales* genomes and

29   assessed their presence in metagenomes derived from the human gut (n=4472

30   representing 25 populations), nonhuman animal gut (n=145) and nonhost

31   environments (n=160). Our analyses showed that all taxa are generalists: they were

32   detected in animal gut and environmental samples. We confirmed two large clades,

33   one enriched in the gut, the other enriched in the environment, with notable

34   exceptions. Genomic adaptations to the gut include genome reduction, a set of

35   adhesion factors distinct from that of environmental taxa, and genes involved in the

36   shikimate pathway and bile resistance. Genomic adaptations differed by clade, not

37   habitat preference, indicating convergent evolution between the clades. In the

38   human gut, the relative abundance of *Methanomassiliicoccales* correlated with

39   trimethylamine-producing bacteria and was unrelated to host genotype. Our results

40   shed light on the microbial ecology of this group may help guide

41   *Methanomassiliicoccales*-based strategies for trimethylamine mitigation in

42   cardiovascular disease.

43 **Importance**

44   *Methanomassiliicoccales* are a lesser known component of the human gut

45   microbiota. This archaeal order is composed of methane producers that use

46   methylated amines, such as trimethylamine, in methane production. This group has

47   only one cultured representative; how they adapted to inhabit the mammalian gut

48   and how they interact with other microbes is largely unknown. Using bioinformatics

49   methods applied to DNA from a wide range of samples, we profiled the relative

50   abundances of these archaea in environmental and host-associated microbial

51   communities. We observed two groups of *Methanomassiliicoccales,* one largely

52   host-associated and one largely found in environmental samples, with some

53   exceptions. When host-associated, these archaea have a distinct set of genes

54   related to adhesion and possess genes related to bile resistance. We did not detect

55   *Methanomassiliicoccales* in all human populations tested but when present, they are

56   correlated with *Bacteria* known to produce trimethylamine. Since trimethylamine is

57   linked to cardiovascular disease risk, these intriguing Archaea may also be involved.

Introduction

58

59   *Archaea* generally make up a tenth or less of the biomass of the human gut

60   microbiota; however, they are widely prevalent and occupy a unique metabolic niche,

61   utilizing by-products of bacterial metabolism as substrate for methanogenesis (1).

62   The most widespread methanogens in the human gut are members of the order

63   *Methanobacteriales.* These include *Methanobrevibacter smithii*, which uses CO_2,

64   formate and H_2 as substrates for methane production (2), and *Methanosphaera*

65   *stadtmanae,* which consumes methanol and H_2 (3). Through the process of methane

66   formation, *Archaea* decrease partial pressures of H_2, thereby potentially increasing

67   the energetic efficiency of primary fermenters and the production of short-chain fatty

68   acids (4). Members of *Methanobacteriales* are the dominant species of the human

69   gut archaeome (1, 5).

70   A second archaeal lineage, the order *Methanomassiliicoccales*, is also found

71   within the human gut microbiota, yet its members are less well characterized than

72   those of *Methanobacteriales*. Members of the order *Methanomassiliicoccales*,

73   including human-derived *Methanomassiliicoccus luminyensis, "candidatus*

74   Methanomassiliicoccus intestinalis" and "*candidatus* Methanomethylophilus alvus",

75   perform H_2-dependent methylotrophic methanogenesis as sole energy source (6–8).

76   Their genomes encode several methyltransferases and associated proteins used to

77   reduce methylamines and methanol to methane. Studies based on 16S rRNA and

78   *mcrA* gene diversity analysis indicate that the order *Methanomassiliicoccales* is

79   made up of two large clades, which mostly group species that have either a free

80   living (FL) or host associated (HA) lifestyle (9, 10). Based on analyses of the

81   genomes from three human-derived species from both clades, Borrel *et al.* (11)

82   suggested each clade colonized the mammalian gut independently. Members of the

4

83   HA clade, including the human-associated "*ca.* M. alvus*"*, might be expected to show

84   adaptations similar to other methanogens from the gut microbiota (12, 13). How

85   members of the FL clade, including the human-associated *M. luminyensis* and *"ca.*

86   *M.* intestinalis", have converged on the gut niche remains to be explored.

87   A better understanding of the ecology of *Methanomassiliicoccales* may be of

88   interest to human health, as they can utilize mono-, di-, and trimethylamine (TMA) as

89   substrate for methanogenesis in the gut (14). TMA, a by-product of the bacterial

90   metabolism of carnitine, choline, and other and choline-containing compounds, is

91   absorbed by the host and transformed in the liver into trimethylamine N-oxide

92   (TMAO) (15). In turn, circulating TMAO inhibits cholesterol transport and promotes

93   its accumulation in macrophages, inducing the formation of artherosclerotic plaques

94   (16). Decreasing TMA levels in the gut, and reducing circulating TMAO levels, has

95   been proposed as a therapeutic strategy for cardiovascular disease (17). One way to

96   use the gut microbiome to this end would be to boost levels of

97   *Methanomassiliicoccales* (18). To accomplish this goal requires a deeper

98   understanding of its ecology.

99   Here, we conducted a comparative analysis of 71 *Methanomassiliicoccales*

100   genomes, together with an additional metagenome-assembled genome (MAG)

101   corresponding to a strain of "*ca.* M. alvus*"*, which we retrieved by metagenome

102   assembly of gut samples from subjects of the TwinsUK cohort (19). We used 305

103   publically available metagenomes to assess the prevalence of taxa across various

104   habitat types. While the two large clades grouping host-associated (HA) and free-

105   living (FL) taxa, are generally enriched in host-associated and environmental

106   metagenomes, a few exceptions stand out. Our results showed that the repertoire of

107   adhesion proteins encoded by the genomes of taxa from each clade differed. Genes

108   involved in bile resistance and the shikimate pathway are likely involved in the

109   adaptation to the gut environment of members of the HA clade, but not for the FL

110   clade. Thus, gut-adapted members converged on life in the gut using different

111   genomic adaptations. *Methanomassiliicoccales* genera present in the human gut

112   positively correlate with TMA-producing bacteria.

113

Materials and Methods

Genome annotation and phylogenomic tree reconstruction

116   We downloaded 78 available genomes belonging to the order

117   *Methanomassiliicoccales* from the NCBI assembly database

118   (https://www.ncbi.nlm.nih.gov/assembly) as available in June 2018, and used

119   CheckM to assess their quality. For subsequent analyses, we included 71

120   substantially complete genomes (completeness ≥70 %) with low contamination

121   (contamination <5 %) (20), plus an additional high-quality metagenome-assembled

122   genome (MAG) corresponding to "*candidatus* Methanomethylophilus alvus" (see

123   *supplementary methods* and table S1). Gene calling, proteome prediction and

124   annotation was performed on each genome using Prokka 1.12 (21). Details of each

125   genome, including the original source of isolation, can be found in (table S1).

126   Using PhyloPhlAn 0.26 (22), we constructed a maximum-likelihood

127   phylogenomic tree using a concatenated alignment of multiple universally distributed

128   single copy marker genes of 72 publicly available genomes from the order

129   *Methanomassiliicoccales*. Of these, one was retrieved from pure culture, 6 were

130   obtained from enrichment cultures and 64 were MAGs. We included an additional

131   MAG retrieved from human gut metagenomes corresponding to "*ca.* M. alvus"

132   (*supplementary results*). Briefly, universal markers were obtained from the translated

6

133   amino acid sequences of the included genomes, aligned using mafft 7.3 (23) and

134   concatenated into a single sequence. We then used the concatenated alignment to

135   reconstruct an maximum-likelihood phylogenetic tree using RAxML 8.1 (24); branch

136   support was estimated by 1000 bootstrap iterations and the tree was rooted by

137   including members of the order *Thermoplasmatales* as outgroup, namely

138   *Thermoplasma acidophilum* DSM 1728 (GenBank assembly accession:

139   GCA_000195915.1), *Picrophilus oshimae* DSM 9789 (GCA_900176435.1),

140   *Ferroplasma acidarmanus* fer1 (GCA_000152265.2), *Acidiplasma aeolicum*

141   (GCA_001402945.1) and *Cuniculiplasma divulgatum* (GCA_900090055.1). We used

142   iTOL (25) to visualize the tree.

143

144   ***Abundance of Methanomassiliicoccales in environmental and animal***

145   ***gastrointestinal metagenomes***

146   We retrieved 305 metagenome samples of gastrointestinal and environmental

147   origin (26) sequenced using the Illumina HiSeq platform (table S2). Sequences were

148   then downloaded from the Sequence Read Archive (SRA) and quality-controlled

149   (see *supplementary methods*). To avoid the issue of multiple mapping, we

150   dereplicated the 72 genomes at a species-level threshold (95 % ANI) using dRep,

151   resulting in 29 representative genomes. Next, we quantified the abundance of

152   dereplicated *Methanomassiliicoccales* genomes in these samples using KrakenUniq

153   v.0.5.8 (27). Statistical analyses were performed using R v.3.5.1 (28). We estimated

154   the enrichment of each representative *Methanomassiliicoccales* on host or

155   environmental metagenomes using DESeq2 (29) on sequence counts and

156   classifying metagenome samples as either host-derived or environmental. We

157   applied hierarchical clustering using Ward's method on the log-fold-change of

7

158   environmental vs gastrointestinal enrichment of each taxon and calculated the

159   cophenetic correlation with the phylogenomic tree using the ape package of R (30).

160

161   ***Comparative genomics***

162   The predicted proteome of each included genome was used to to assign

163   orthology clusters using panX 1.6.0 (31). We used InterProScan (32) and eggNOG

164   mapper 1.0.3 (33) with DIAMOND 0.8.36 (34) against the optimized archaeal

165   database to improve the annotation of gene clusters. Phylogenetic signal of genome

166   characteristics and gene cluster presence was tested using the phylosignal package

167   of R with the local indicator of phylogenetic association (LIPA) (35).

168   The R package micropan (36) was used to create a principal component

169   analysis (PCA) of gene cluster presence. We compared the gene cluster content

170   between clades to determine gene clusters enriched on clades FL or HA using

171   phylogenetic ANOVA using the R package phytools (37). To reduce the number of

172   comparisons we first removed low frequency gene clusters by filtering those with

173   near zero variance. The above analysis was repeated by comparing gene cluster

174   content between taxa significantly enriched on gut or environmental samples, prior

175   removal of taxa not significantly enriched in either biome class. We adjusted P

176   values with the Benjamini-Hochberg method.

177   We assessed the presence of eukaryote-like proteins (ELPs) (38) by

178   combining the counts of gene clusters classified by InterProScan as any of the

179   following: Sel1 containing proteins (Sel1), Listeria-Bacteroides repeat containing

180   proteins (List-Bact), tetratricopeptide repeats (TPRs), Ankyrin repeats (ANKs),

181   Leucine-rich repeats (LRRs), Fibronectin type III (fn3) domains, Laminin G domain,

182   Bacterial Ig-like domains, YadA-like domain (Yersinia adhesin A), TadE-like domain

183   or Invasion protein B (ialB). Likewise, we characterized the presence of parallel beta-

184   helix repeat-containing proteins, also known as adhesin-like proteins (ALPs).

185

186   ***Characterization of Methanomassiliicoccales distribution across human***

187   ***populations***

188   We obtained sample metadata from publicly available studies using the

189   curatedMetagenomicData v.1.17.0 package of Bioconductor (39). Samples were

190   selected according to the following criteria: i) shotgun gut metagenomes sequenced

191   using the Illumina HiSeq platform with a median read length > 95 bp; ii) with

192   available SRA accession; iii) labeled as adults or seniors, or with a reported age ≥ 18

193   years; iv) without report of antibiotic consumption (i.e. no or NA); v) without report of

194   pregnancy (i.e. no or NA); vi) non-lactating women (i.e. no or NA); vii) without report

195   of gangrene, pneumonia, cellulitis, adenoma, colorectal cancer, arthritis, Behcet's

196   disease, cirrhosis or inflammatory bowel disease. Only forward reads were

197   downloaded and processed. A total of 4472 samples from 34 independent studies

198   were downloaded from the SRA between December 2019 and February 2020 (table

199   S3) and quality controlled as described in *supplementary methods*.

200   Reads were classified using Kraken v.2.0 (40) and a Bayesian re-estimation

201   of the species-level abundance of each sample was then performed using Bracken

202   v.2.2 (41). We utilized custom databases created using the Struo pipeline (42) based

203   on GTDB release 86 (available at http://ftp.tue.mpg.de/ebio/projects/struo/). Taxa

204   with <100 reads in a given sample were considered as absent. We obtained

205   complete taxonomic annotations from NCBI taxIDs with TaxonKit 0.2.4

206   (https://bioinf.shenwei.me/taxonkit/). To determine the cooccurrence patterns of the

207   detected *Methanomassiliicoccales* in the human gut we used the cooccur package of

9

208   R (43); to determine their coabundance patterns, we calculated the proportionality of

209   taxa abundance (*rho*) with the propr package (44). The lme4 and lmerTest R

210   packages (45) were used to fit linear mixed effects models to test differences of

211   *Methanomassiliicoccales* genera log-transformed abundance by westernization

212   status, age and gender with F-tests and P-values determined via the Satterthwaite's

213   method (ANOVA Type II sum of squares). Similarly, we employed binomial linear

214   mixed models to test differences of *Methanomassiliicoccales* genera prevalence.

215   We assessed the heritability of *Methanomassiliicoccales* taxa by comparing

216   relative abundances within 153 monozygotic (MZ) and 200 dizygotic (DZ) twin pairs

217   using the taxonomic profiles of 706 gut metagenome samples from the United

218   Kingdom Adult Twin Registry (TwinsUK) (19, 46, 47) with a sequencing depth >5

219   million reads/sample. We aggregated abundances at the genus level and removed

220   genera with a prevalence <5 %. Absolute read counts were transformed using the

221   Yeo-Johnson transformation and adjusted by BMI, sex and sequencing depth (19,

222   46). For each genus, we calculated the intraclass correlation coefficient (ICC) in MZ

223   and DZ twins with the irr package of R, and adjusted P-values for multiple

224   comparisons using the Benjamini-Hochberg method. As control we compared the

225   mean ICC across all taxa between MZ and DZ twins using the Mann-Whitney test,

226   and by assessing the ICC of specific taxa known previously reported as heritable in

227   the same population (*Methanobrevibacter*, *Faecalibacterium*, *Christensenella* and

228   *Bifidobacterium*) (46, 48). We carried a sensitivity analysis by repeating these

229   analyses on a subset of 394 samples (80 MZ and 117 DZ twin pairs) with a

230   sequencing depth of >12 million reads/sample.

231

232   ***Data and code availability***

233   The metagenomic sequence data generated during this study have been

234   deposited in the European Nucleotide Archive with accession ID PRJEB40256. The

235   jupyter notebooks with analysis code are available at

236   https://github.com/leylabmpi/Methanomassilii. The "*candidatus*

237   Methanomethylophilus alvus" MAG here generated can be found at

238   http://ftp.tue.mpg.de/ebio/projects/Mmassilii

239

Results

241   ***Genome-based phylogeny confirms two large Methanomassiliioccales clades***

242   Based on whole-genome phylogenetic analysis, the order

243   *Methanomassiliicoccales* forms two clades with robust support (figure 1). This

244   phylogeny is in agreement with previously reported phylogenies based on 16S rRNA

245   and mcrA genes (9, 49, 50). A third distal clade was formed by two closely related

246   MAGs generated in a recent massive metagenome assembly effort (51), which we

247   labeled external (EX; figure 1). We use the terminology of Borrel *et al.,* (52): the

248   clade including *Methanomassiliicoccus* is labeled free-living (FL), and the clade

249   containing "*candidatus* Methanomethylophilus" host-associated (HA).

250   As observed previously (9), the reported source of the genomes was not

251   always consistent with the clade in which it was grouped. For instance, while publicly

252   available genomes originally retrieved from human, baboon, elephant and cow

253   gastrointestinal tracts were related to "*candidatus* Methanomethylophilus" (HA), this

254   clade also contained MAGs derived from digestor and reactors (figure 1) reportedly

255   not treating animal waste (table S1). Moreover, MAGs retrieved from pit mud of

256   solid-state fermentation reactors used for the production of Chinese liquor were

11

257   present in both the HA and FL clades (table S1). Similarly, "*ca.* M. intestinalis"

258   Issoire-Mx1, *M.* luminyensis B10, and Methanomassiliicoccales archaeon RumEn

259   M1, all retrieved from mammal hosts, grouped in the FL clade.

260

261   ***Abundance of Methanomassiliicoccales clades differs in gastrointestinal and***

262   ***environmental samples***

263   We assessed the abundance of species-level representative

264   *Methanomassiliicoccales* taxa in publicly available metagenomes that included 145

265   samples from gastrointestinal tracts of non-human animals, such as cats, pigs, elks,

266   cows, sheep, mice, white-throated woodrats, trouts, chickens and geese, and 160

267   environmental samples from sediment, ice, and diverse water and soil sources (table

268   S2).

269   Taxa from all three clades were detected in a wide range of metagenomes

270   from environmental and gut origin. We observed differences in environmental

271   preference by clade. Abundance of taxa from Clade EX was highest on

272   environmental metagenomes (0.001 % ± 0.0012) (figure 1). They were also detected

273   in gut samples (0.0002 % ± 0.0005), albeit with a very low abundance in fecal

274   (0.0003 % ± 0.0005), large intestine (0.0001 % ± 0.0002), stomach (0.0009 % ±

275   0.0006) metagenomes (figure 2). Given their low abundances, further analysis is

276   focused on the FL and HA clades.

277   The aggregated abundance of Clades FL and HA varied across biomes

278   (figure 2). In agreement with their names, HA clade members were enriched in host-

279   associated samples, and FL in non-host samples. The combined abundance of

280   members of Clade FL was higher in samples from environmental biomes (0.01 % ±

281   0.008), although non-zero abundances were observed in digestive system

12

282 metagenomes (0.008 % ± 0.015), with some samples containing levels comparable

283 to that of Clade HA (figure 2).

284   The mean abundance of Clade HA in aggregate was higher in metagenomes

285 from gut samples (0.014 % ± 0.03) compared to environmental biomes (0.004 % ±

286 0.008). However, among the environmental biomes, non-zero abundances of Clade

287 HA were detected in freshwater (0.002 % ± 0.003), marine (0.006 % ± 0.011), saline

288 and alkaline (0.002 % ± 0.002) and soil (0.004 % ± 0.003) samples.

289   We further validated the differences in clade abundances across biomes by

290 generating a dendrogram of *Methanomassiliicoccales* taxa using the fold-change

291 enrichment of individual taxa on gut versus environmental biomes, that is, the effect

292 size of their enrichment in either direction. We then compared the structure of this

293 dendrogram with that of the phylogenomic tree and found that they were positively

294 correlated (cophenetic correlation = 0.67, P val. < 0.01).

295   Overall, we observed a low abundance of individual *Methanomassiliicoccales*

296 taxa across all samples, ranging from 0 to 0.15 % (figure 2 and figure S1). The

297 enrichment analysis of individual taxa from Clade FL on diverse biomes showed that

298 while most were significantly enriched in environmental metagenomes, some taxa

299 showed the opposite enrichment. *M. luminyensis* and Methanomassiliicoccus sp.

300 UBA386 were not significantly enriched in gut or environmental biomes. "*ca.* M.

301 intestinalis" Issoire-Mx1, Methanomassiliicoccales archaeon RumEn M1 and

302 Methanomassiliicoccus sp. UBA6 were significantly enriched in gut biomes (figure 1),

303 although they were also present in multiple environmental biomes (figure S1).

304   When assessed on a per-taxon basis, the vast majority of Clade HA taxa were

305 significantly enriched in gut samples, with the exception of "*ca.* M. termitum", which

13

306   was highly abundant in soil samples from grasslands and water samples from

307   intertidal zones (figure 1).

308

309   ***Genome characteristics and core genes functions differ between***

310   ***Methanomassiliicoccales clades***

311   Given the tendency of clades FL and HA to be enriched in environmental or

312   animal metagenomes, respectively, we searched for genes and genome features

313   linked to putative adaptations of *Methanomassiliicoccales* to an animal gut. For this,

314   we compared 72 genomes from *Methanomassiliicoccales* taxa retrieved from

315   humans, non-human animals and environmental sources.

316   We observed that genomes were more similar to others closely located on the

317   phylogeny for genome GC content, genome length and total gene count (LIPA Adj. P

318   < 0.01 in all cases) (figure 3). To determine whether these features differed between

319   clades, while accounting for the autocorrelation due to evolutionary history, we

320   performed a phylogenetic ANOVA. Clade FL taxa had significantly larger genomes

321   (mean ± sd: 1985.1 Kb ± 245.1) than either the clades HA (1318.3 Kb ± 187.3) or EX

322   (1872.2 Kb ± 173.8) (phylogenetic ANOVA Adj. P = 0.028). In accord, Clade FL also

323   had the highest gene count (FL: 2153.1 genes ± 233.7; HA: 1377.7 genes ± 187.7;

324   EX 1567.0 genes ± 90.5. Adj. P = 0.025). While non-significant, clades HA and EX

325   taxa tended to have a lower GC content than Clade FL taxa (FL: 59.1 % ± 4.8; HA:

326   55.8 % ± 2.8; EX 54.4 % ± 0.5. Adj. P = 0.6).

327   To compare gene presence and absence across clades, we performed a

328   pangenome analysis. After identification of orthologous gene clusters based on

329   sequence similarity using PanX, we obtained 13,695 clusters, of which 7,312 were

330   present at least once in Clade FL, 6,592 in Clade HA, and 1,833 in Clade EX. A

14

331   large proportion of gene clusters were of unknown function according to the COG

332   functional classification (38.4 % ± 4.3); gene clusters of unknown function tended to

333   be small, with only one or two genes (figure S2 A, B). Principal component (PC)

334   analysis of gene cluster presence/absence clearly differentiated clades along PC1.

335   We defined outlier taxa as FL taxa enriched in gut biomes

336   (Methanomassiliicoccales archaeon RumEn M1, Methanomassiliicoccus sp. UBA6,

337   "ca. Methanomassiliicoccus intestinalis" Issoire-Mx1, *Methanomassiliicoccus*

338   *luminyensis* B10 and Methanomassiliicoccus sp. UBA386) and the HA taxon

339   enriched in non-host biomes ("*ca.* M. termitum"). Outliers mostly clustered with their

340   close relatives, not with the taxa enriched in the same biome (figure 4), with the

341   exception of "ca. Methanomassiliicoccus intestinalis" Issoire-Mx1, which did not

342   cluster with either clade.

343

344   ***Gene clusters enriched in Clade HA evidence adaptation to the gut***

345   ***environment***

346   Because of the small number of genomes that cluster within the Clade EX,

347   and because they are largely absent from animal-associated samples, subsequent

348   analyses focus on comparisons between the clades FL and HA.

349   To identify gene clusters potentially involved in the adaptation of members of

350   Clade HA to a host environment, we compared the gene cluster content between

351   clades. The gene cluster frequency spectrum shows many clusters with a small

352   number of genes: 7,990 (58.3 %) gene clusters were singletons and 2,002 (14.6 %)

353   were doubletons (figure S2 A, B). After removing rare gene clusters by filtering those

354   with near zero variance, we included 2937 clusters, which we then used to perform

355   in phylogenetic ANOVAs. Results reveal 14 gene clusters significantly enriched in

15

356 HA compared to FL (Adj. P < 0.1 in all cases). Three gene clusters are involved in

357 detoxification and xenobiotic metabolism, namely, bile acid:sodium symporter

358 (InterPro accession IPR002657), bleomycin resistance protein (IPR029068) and

359 HAD-superfamily hydrolase (IPR006357). Two clusters are related to shikimate or

360 chorismate metabolism: chorismate mutase II (IPR002701) and prephenate

361 dehydratase (IPR001086). Other annotated clusters include the small unit of

362 exonuclease VII (IPR003761), holliday junction resolvase Hjc (IPR002732), nitrogen

363 regulatory protein PII (IPR015867), xylose isomerase-like protein (IPR013022) and

364 metal-binding domain containing protein (IPR019271). Four had poor or no

365 annotation (table 1).

366

367 ***Genomic adaptations to the gut of members of the FL clade***

368   To determine whether outlier taxa belonging to Clade FL had similar

369 adaptations to the gut as members of Clade HA, we explored gene clusters present

370 in these outliers and in Clade HA but that were rare in other members of Clade FL.

371 We selected gene clusters present in the core genome of Clade HA (i.e. present in

372 >80 % of taxa from this clade, see supplementary results) and present in less than

373 half of FL taxa. A total of 15 gene clusters were obtained, most of them encoded by

374 only one of the outlier taxa. Two gene clusters, ferrous iron transport proteins A and

375 B (IPR030389 and IPR007167), were present in three of the outliers

376 (Methanomassiliicoccus luminyensis B10, "*candidatus* Methanomassiliicoccus

377 intestinalis" Issoire-Mx1 and Methanomassiliicoccales archaeon RumEn M1). Other

378 clusters detected in more than one outlier included an uncharacterised membrane

379 protein (IPR005182, in Methanomassiliicoccus sp. UBA6 and

380 Methanomassiliicoccales archaeon RumEn M1), a putative nickel-responsive

16

381   regulator (IPR014864, in Methanomassiliicoccus luminyensis B10 and

382   Methanomassiliicoccus sp UBA386), and an ABC transporter (IPR037294, in

383   Methanomassiliicoccus luminyensis B10 and Methanomassiliicoccus sp UBA386).

384   The remaining gene clusters, detected once, corresponded to transcriptional

385   regulators or proteins of unknown function.

386

387   ***The genomes of taxa from clades HA and FL encode distinct repertoires of***

388   ***adhesion proteins***

389   We compared between FL and HA clades two large groups of membrane

390   proteins involved in adhesion: eukaryote-like proteins (ELPs), a series of protein

391   families involved in microbial adherence to its host (38), and adhesin-like proteins

392   (ALPs), a class of proteins hypothesized to be involved in the microbe-microbe

393   interactions of *Methanobacteriales* in the gut (13). We aggregated the counts of gene

394   clusters annotated as the ALP and ELP classes, and performed phylogenetic

395   ANOVA. This analysis showed that members of each clade tended to encode a

396   different repertoire of adhesion proteins (figure 3). Taxa from Clade HA had a higher

397   mean count of tetratricopeptide repeats (Mean±SD count; HA: 16.30±06.56, FL:

398   9.55±1.70), Sel1 repeats (HA: 9.32±5.69, FL: 0.35±1.35), Listeria-Bacteroides

399   repeats (HA: 3.68±3.76, FL: 1.65±5.78) and leucine-rich repeats (HA: 1.5±2.15, FL:

400   1.1±2.02) than FL taxa, although we did not observe significant differences in their

401   frequency (Adj. $P > 0.1$ in all cases). Conversely, ALPs (FL: 2.25±1.48, HA:

402   0.14±0.61) and Ig-like domains, (FL: 1.55±1.32, HA: 0.20±0.53) tended to be more

403   abundant in the genomes of members of Clade FL.

404   Interestingly, outlier taxa from Clade FL had gene counts of several of the

405   adhesion factors higher than the mean of their own clade and more characteristic of

17

406   clade HA. In some cases, the gene counts were higher than the mean for Clade HA.

407   These included Listeria-Bacteroides repeats (gene cluster count - *M. luminyensis*: 2,

408   "*ca.* M. intestinalis" Issoire-Mx1: 26, Methanomassiliicoccales archaeon RumEn M1:

409   2), Sel1 repeats (*M. luminyensis*: 1, "*ca.* M. intestinalis" Issoire-Mx1: 6), and leucine-

410   rich repeats (*M. luminyensis*: 5, "*ca.* M. intestinalis" Issoire-Mx1: 7).

411

412   ***Methanomassiliicoccales taxa cooccur with each other, with other Archaea,***

413   ***and with TMA producing bacteria in the human gut***

414   　　　　We characterized the distribution of *Methanomassiliicoccales* across a

415   collection of human gut metagenomes derived from 34 studies. Together, the

416   combined 4472 samples represented people from 22 countries, resulting in 35

417   unique datasets (*i.e.,* study-country combination). Across the whole set, we detected

418   just two genera: *Methanomassiliicoccus* (Clade FL) and "*ca.* Methanomethylophilus"

419   (Clade HA), both rare members of the human gut microbiota (figure 5). "*ca.*

420   Methanomethylophilus" was detectable in 19 out of 35 datasets; on these 19

421   datasets it had a prevalence ranging from 0.5 % to 41.7 %, and mean abundance

422   ranged from $4.8*10^{-6}$ % to $2.2*10^{-2}$ %. Similarly, *Methanomassiliicoccus* was

423   detectable in 22 of the 35 datasets; on the 22 datasets it had a prevalence range of 1

424   % to 25.7 % and a mean abundance range of $1.5*10^{-5}$ % to $1.0*10^{-2}$ % (table S4).

425   　　　　We tested associations of these two genera with age, sex and westernization

426   status of the subjects using linear mixed models that included the dataset and

427   country as random effects. Subjects from non-westernized countries had a

428   significantly higher prevalence of "*ca.* Methanomethylophilus" (mean prevalence ±

429   SD: Non-westernized = 8.9 % ± 28.5, Westernized 1.1 % ± 10.3; P Adj. = 0.002).

430   Westernized individuals were more likely to harbor higher *Methanomassiliicoccus*,

18

431   although differences were not significant (Non-westernized = 3.9 % ± 19.4,

432   Westernized 5.0 % ± 21.7; P Adj. > 0.1). The age and sex of the individuals did not

433   explain variance in the prevalence or abundance of either genus (Adj. P > 0.1 in all

434   cases).

435   To identify other microbial taxa positively associated with members of

436   *Methanomassiliicoccales* in the human gut, we calculated a network of positively

437   associated microorganisms (*i.e.* coabundant taxa) across samples (rho >0.1 in all

438   cases) (53). In addition, we determined which taxa were present with members of

439   *Methanomassiliicoccales* more than expected by chance (*i.e.* cooccurring taxa)

440   relative to a permuted null model (43). Results showed that both "*ca.*

441   Methanomethylophilus" and *Methanomassiliicoccus* were part of the same

442   coabundance network, together with a third archaeal genus, *Methanoculleus (*order

443   *Methanomicrobiales*). We did not find evidence of positive or negative abundance

444   associations of either *Methanomassiliicoccales* genus with *Methanobrevibacter*.

445   Coocurrence analysis showed a random association pattern between these taxa (P

446   val. > 0.05 for both "*ca.* Methanomethylophilus" and *Methanomassiliicoccus*),

447   indicating that their ecological niches do not overlap with *Methanobrevibacter*.

448   Analysis of the combined network of "*ca.* Methanomethylophilus" and

449   *Methanomassiliicoccus* revealed a large overlap between taxa associated with either

450   genus (figure 6): out of 119 taxa in the network, 86 (72.3 %) were associated with

451   both. Moreover, 51 taxa (42.9 %) also had a significant positive cooccurrence pattern

452   with both genera (P val. < 0.05 in all cases). Most bacterial members of this network

453   had an overall low relative abundance. Interestingly, they included several taxa

454   whose genomes contain genes encoding enzymes involved in TMA production,

19

455   including *Bacteroides, Campylobacter, Yokenella, Mobiluncus, Proteus, Providencia*

456   and *Edwardsiella (54)*.

457

458   **Abundance of Methanomassiliicoccales is not concordant in monozygotic or**

459   **dizygotic human twins**

460   To evaluate whether host genetics influences the abundance of

461   *Methanomassiliicoccales* in the human gut, we compared the intraclass correlation

462   coefficient (ICC) of their abundances at the genus level using a set of 153

463   monozygotic (MZ) and 200 dizygotic (DZ) twin pairs from the TwinsUK cohort. As

464   control, we first compared the mean ICC across all taxa between MZ and DZ twins,

465   and found that ICC_{MZ} (0.1) was significantly higher than ICC_{DZ} (0.03) (P val. < 0.01).

466   In addition, we assessed the ICC values of bacterial (*Christensenella,*

467   *Faecalibacterium and Bifidobacterium*) and archaeal (*Methanobrevibacter*), and

468   consistently found a higher correlation on MZ compared to DZ twins (table S5). We

469   were only able to assess ICC values of *Methanomassiliicoccus*, as it was the only

470   *Methanomassiliicoccales* detected in the twins with a prevalence (8.64 %) above the

471   5 % cutoff (see methods). We did not detect a significant concordance between the

472   abundances of *Methanomassiliicoccus* in MZ (ICC_{MZ} = 0.004, Adj. P = 0.59) or in DZ

473   twins (ICC_{DZ} = 0.017, Adj. P = 0.71). Given the low abundance of

474   *Methanomassiliicoccales* taxa, we performed a sensitivity analysis using samples

475   with a high sequencing depth (>12 million reads/sample), however, we did not

476   observe differences in the abundance and prevalence of the

477   *Methanomassiliicoccales* genera nor the ICC estimates (not shown).

478

20

Discussion

479

480 While the source of the members of the *Methanomassiliicoccales* has been

481 noted in previous surveys of single markers such as 16S rRNA and mcrA genes (9,

482 10), here we searched metagenomes from host associated and environmental

483 samples for their relative abundances. Overall, the HA taxa were enriched in host

484 associated samples and the FL taxa in environmental samples; intriguingly, all taxa

485 regardless of clade were detected in both biomes. This suggests that members of

486 the order *Methanomassiliicoccales* are generalists with an overall habitat preference

487 according to clade, although there were some exceptions to the general pattern. We

488 show that members of *Methanomassiliicoccales* use many of the same adaptations

489 to the gut as other methanogens. These adaptations include genome reduction, and

490 genes involved in the shikimate pathway and bile resistance. In addition, gut-

491 enriched taxa possess a distinct repertoire of genes encoding adhesion factors. We

492 observed that potential adaptations to the gut differed by clade, not preferred habitat,

493 indicating convergence on a shared niche through different genomic solutions. In the

494 human gut, *Methanomassiliicoccales* taxa correlated with TMA-producing bacteria,

495 rather than host genetics or other host factors.

496 For members of the HA clade, adaptations to life in the gut included an

497 enrichment of genes involved in bile acid transport, efflux pumps, and hydrolases,

498 which play a role in tolerance to these compounds in the gastrointestinal tract (55).

499 This adaptation is also shared with other members of the gut microbiota, including

500 *Methanobacteriales: M. smithii* and *M. stadtmanae* are resistant to bile salts (2, 3).

501 Other gene clusters with known function enriched in Clade HA are involved in

502 metabolism of shikimate and chorismate. The shikimate pathway is involved in the

503 synthesis of aromatic amino acids in plants and microbes, but is absent in mammals.

21

504   Shikimate metabolism is carried out by archaeal (56) and bacterial (57, 58) members

505   of the animal gut microbiota, and was reported as one of the most conserved

506   metabolic modules in a large-scale gene catalogue from the human gut (59).

507   Derivatives from the aromatic amino acids are known to be bioactive in the mammal

508   host (60).

509   In addition, members of the HA clade had a particular set of adhesion factors,

510   known to be involved in the maintenance of syntrophic relationships of the

511   methanogens with bacterial (12, 61) or eukaryotic (62) microorganisms. Two groups

512   of adhesion factors, proteins containing Sel1 domains and Listeria-Bacteroides

513   repeats, have been previously studied on *Methanomassiliicoccales* taxa retrieved

514   from the gut (11, 52). Our assessment of these factors in the broader context of the

515   order *Methanomassiliicoccales* showed that these two groups of proteins are

516   characteristic of Clade HA rather than FL, with the exception of the outlier taxa.

517   Indeed, while the repertoire of ELPs and ALPs differs between HA and FL taxa, it

518   was similar between species inhabiting the gut regardless of their clade.

519   In contrast, members of Clade FL appear to be generalists that colonized the

520   animal gut independently from the HA clade. It has been previously noted that *M.*

521   *luminyensis*, an outlier from Clade FL, could have a facultative association to the

522   animal gut. It possesses genes involved in nitrogen fixation, oxidative stress (11) and

523   mercury methylation (52), which are common in soil microorganisms but rare in

524   members of the gut microbiota (63). In accord, we observed that members of Clade

525   FL are widespread and abundant on soil, water and gut metagenomes, with a

526   preference for environmental biomes. Similarities in ELP content between gut-

527   dwelling taxa from both clades indicate that interaction with the host or other

528   members of the gut microbiota might be a key factor in the adaptation of these

22

529   methanogens.

530   Analysis of the gene content of outlier taxa from Clade FL showed that they

531   tended to be more similar to members of their own clade than to taxa from Clade HA,

532   with the exception of "*ca.* M. intestinalis" Issoire-Mx1, which was distinct from either

533   Clade FL and HA. In addition there was little overlap in gene clusters commonly

534   observed in Clade HA and outlier taxa from Clade FL, with the exception of the

535   adhesion factors discussed below. These observations support the hypothesis that

536   colonization of animal guts by members of *Methanomassiliicoccales* occurred in two

537   independent events (11, 52), and suggests that there is not one solution to life in the

538   gut for these *Archaea*, as members from two clades seem to have solved the

539   problem with a different set of adaptations.

540   Characterization of the abundance of *Methanomassiliicoccales* across human

541   populations showed members of this group are rare in the microbiota of healthy

542   adults. We did not detect them in all the studied populations, and when detected,

543   they had low prevalence and abundance. Nevertheless, this analysis allowed us to

544   assess whether *Archaea* in the human gut are mutually exclusive. We observed

545   positive correlations of "*ca.* Methanomethylophilus" and *Methanomassiliicoccus* with

546   each other and with *Methanoculleus*, another rare archaeal member of the gut

547   microbiota (64). We did not find evidence of association between members of

548   *Methanomassiliicoccales* and *Methanobrevibacter*, positive or otherwise, confirming

549   the previous report that these methanogens are not mutually exclusive (46);

550   abundance of H_2 in the gut, together with differences in other substrate utilization,

551   might result in non-overlapping niches (65).

552   While genus *Methanobrevibacter* has been consistently found to have a

553   moderate heritability the TwinsUK (19, 46, 66) and other cohorts (48, 67), it was not

554   the case for members of *Methanomassiliicoccales*. Similar to humans, methane

555   production (68) and abundance of *Methanobrevibacter* (69) are also heritable in

556   bovine cattle, but not *Methanomassiliicoccales* taxa (69). Thus, host genetics might

557   be linked to particular taxa and methanogenesis pathways, not to all *Archaea* or

558   methane production as a whole.

559   Genera "*ca.* Methanomethylophilus" and *Methanomassiliicoccus* cooccur with

560   TMA-producing bacteria (54), further supporting their potential use as a way of

561   targeting intestinal TMA (70). The exact nature of the ecological relationships each of

562   these taxa establishes with other members of the microbiome remains to be

563   elucidated. In a facilitation scenario between the methanogens and H_2- and TMA-

564   producers, freely available TMA and H_2 required for methylotrophic methanogenesis

565   could be utilized by *Methanomassiliicoccales* taxa (71), without cost to the producer.

566   Alternatively, the methanogens could establish syntrophic interactions with other

567   microorganisms, whereby the consumption of these metabolites is also beneficial to

568   the producer (71).

569   The present study extends our understanding of the order

570   *Methanomassiliicoccales* by revealing genomic adaptations to life in the gut by

571   members of both clades that make up this group. Furthermore, the positive

572   correlation between the relative abundances of these TMA-utilizing archaea with

573   TMA-producing bacteria in the gut is a first step towards understanding how they

574   may be harnessed for therapeutic management of gut TMA levels in the context of

575   cardiovascular disease.

576

Acknowledgements

24

587

References

589   1.   Borrel G, Brugère J-F, Gribaldo S, Schmitz RA, Moissl-Eichinger C. 2020. The host-
590   associated archaeome. Nat Rev Microbiol.

591   2.   Miller TL, Wolin MJ. 1982. Enumeration of Methanobrevibacter smithii in human feces.
592   Arch Microbiol 131:14–18.

593   3.   Miller TL, Wolin MJ. 1985. Methanosphaera stadtmaniae gen. nov., sp. nov.: a species
594   that forms methane by reducing methanol with hydrogen. Arch Microbiol 141:116–122.

595   4.   Horz H-P, Conrads G. 2010. The discussion goes on: What is the role of Euryarchaeota
596   in humans? Archaea 2010:967271.

597   5.   Moissl-Eichinger C, Pausan M, Taffner J, Berg G, Bang C, Schmitz RA. 2018. Archaea
598   Are Interactive Components of Complex Microbiomes. Trends Microbiol 26:70–85.

599   6.   Borrel G, Harris HMB, Tottey W, Mihajlovski A, Parisot N, Peyretaillade E, Peyret P,
600   Gribaldo S, O'Toole PW, Brugère J-F. 2012. Genome Sequence of "Candidatus
601   Methanomethylophilus alvus" Mx1201, a Methanogenic Archaeon from the Human Gut
602   Belonging to a Seventh Order of Methanogens. J Bacteriol 194:6944–6945.

603   7.   Borrel G, Harris HMB, Parisot N, Gaci N, Tottey W, Mihajlovski A, Deane J, Gribaldo S,

604   Bardot O, Peyretaillade E, Peyret P, O'Toole PW, Brugère J-F. 2013. Genome

605   Sequence of "Candidatus Methanomassiliicoccus intestinalis" Issoire-Mx1, a Third

606   Thermoplasmatales-Related Methanogenic Archaeon from Human Feces. Genome

607   Announc 1:e00453–13.

608   8.   Dridi B, Fardeau M-L, Ollivier B, Raoult D, Drancourt M. 2012. Methanomassiliicoccus

609   luminyensis gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces.

610   Int J Syst Evol Microbiol 62:1902–1907.

611   9.   Söllinger A, Schwab C, Weinmaier T, Loy A, Tveit AT, Schleper C, Urich T. 2016.

612   Phylogenetic and genomic analysis of Methanomassiliicoccales in wetlands and animal

613   intestinal tracts reveals clade-specific habitat preferences. FEMS Microbiol Ecol

614   92:fiv149.

615   10. Speth DR, Orphan VJ. 2018. Metabolic marker gene mining provides insight in global

616   diversity and, coupled with targeted genome reconstruction, sheds further light on

617   metabolic potential of the. PeerJ 6:e5614.

618   11. Borrel G, Parisot N, Harris HMB, Peyretaillade E, Gaci N, Tottey W, Bardot O, Raymann

619   K, Gribaldo S, Peyret P, O'Toole PW, Brugère J-F. 2014. Comparative genomics

620   highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related

621   seventh order of methanogenic archaea that encodes pyrrolysine. BMC Genomics

622   15:679.

623   12. Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, Latreille

624   P, Kim K, Wilson RK, Gordon JI. 2007. Genomic and metabolic adaptations of

625   Methanobrevibacter smithii to the human gut. Proc Natl Acad Sci U S A 104:10643–

626   10648.

627   13. Hansen EE, Lozupone CA, Rey FE, Wu M, Guruge JL, Narra A, Goodfellow J, Zaneveld

628   JR, McDonald DT, Goodrich JA, Heath AC, Knight R, Gordon JI. 2011. Pan-genome of

629   the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in

630   twins. Proc Natl Acad Sci U S A 108 Suppl 1:4599–4606.

631   14. Söllinger A, Urich T. 2019. Methylotrophic methanogens everywhere - physiology and

632   ecology of novel players in global methane cycling. Biochem Soc Trans 47:1895–1907.

633   15. Brown JM, Hazen SL. 2018. Microbial modulation of cardiovascular disease. Nat Rev

634   Microbiol 16:171–181.

635   16. Geng J, Yang C, Wang B, Zhang X, Hu T, Gu Y, Li J. 2018. Trimethylamine N-oxide

636   promotes atherosclerosis via CD36-dependent MAPK/JNK pathway. Biomed

637   Pharmacother 97:941–947.

638   17. Wang Z, Roberts AB, Buffa JA, Levison BS, Zhu W, Org E, Gu X, Huang Y, Zamanian-

639   Daryoush M, Culley MK, DiDonato AJ, Fu X, Hazen JE, Krajcik D, DiDonato JA, Lusis

640   AJ, Hazen SL. 2015. Non-lethal Inhibition of Gut Microbial Trimethylamine Production

641   for the Treatment of Atherosclerosis. Cell 163:1585–1595.

642   18. Brugère J-F, Borrel G, Gaci N, Tottey W, O'Toole PW, Malpuech-Brugère C. 2014.

643   Archaebiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and

644   cardiovascular disease. Gut Microbes 5:5–10.

645   19. Xie H, Guo R, Zhong H, Feng Q, Lan Z, Qin B, Ward KJ, Jackson MA, Xia Y, Chen X,

646   Chen B, Xia H, Xu C, Li F, Xu X, Al-Aama JY, Yang H, Wang J, Kristiansen K, Wang J,

647   Steves CJ, Bell JT, Li J, Spector TD, Jia H. 2016. Shotgun Metagenomics of 250 Adult

648   Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome. Cell

649   systems 3:572–584.e3.

650   20. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM:

651   assessing the quality of microbial genomes recovered from isolates, single cells, and

652   metagenomes. Genome Res 25:1043–1055.

27

653 21. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
654   30:2068–2069.

655 22. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method
656   for improved phylogenetic and taxonomic placement of microbes. Nat Commun 4:2304.

657 23. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
658   improvements in performance and usability. Mol Biol Evol 30:772–780.

659 24. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
660   analysis of large phylogenies. Bioinformatics 30:1312–1313.

661 25. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display
662   and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–5.

663 26. Mitchell AL, Scheremetjew M, Denise H, Potter S, Tarkowska A, Qureshi M, Salazar
664   GA, Pesseat S, Boland MA, Hunter FMI, Ten Hoopen P, Alako B, Amid C, Wilkinson DJ,
665   Curtis TP, Cochrane G, Finn RD. 2018. EBI Metagenomics in 2017: enriching the
666   analysis of microbial communities, from sequence reads to assemblies. Nucleic Acids
667   Res 46:D726–D735.

668 27. Breitwieser FP, Baker DN, Salzberg SL. 2018. KrakenUniq: confident and fast
669   metagenomics classification using unique k-mer counts. Genome Biol 19:198.

670 28. R Core Team. 2018. R: A Language and Environment for Statistical Computing. R
671   Foundation for Statistical Computing, Vienna, Austria.

672 29. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion
673   for RNA-seq data with DESeq2. Genome Biol 15:550.

674 30. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution
675   in R language. Bioinformatics 20:289–290.

676   31. Ding W, Baumdicker F, Neher RA. 2018. panX: pan-genome analysis and exploration.

677   Nucleic Acids Res 46:e5.

678   32. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y,

679   Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X,

680   Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka

681   G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N,

682   Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato

683   S, Sutton G, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Xenarios I, Yeh L-S, Young

684   S-Y, Mitchell AL. 2017. InterPro in 2017-beyond protein family and domain annotations.

685   Nucleic Acids Res 45:D190–D199.

686   33. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork

687   P. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by

688   eggNOG-Mapper. Mol Biol Evol 34:2115–2122.

689   34. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using

690   DIAMOND. Nat Methods 12:59–60.

691   35. Keck F, Rimet F, Bouchez A, Franc A. 2016. phylosignal: an R package to measure,

692   test, and explore the phylogenetic signal. Ecol Evol 6:2774–2780.

693   36. Snipen L, Liland KH. 2015. micropan: an R-package for microbial pan-genomics. BMC

694   Bioinformatics 16:79.

695   37. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and

696   other things): phytools: R package. Methods Ecol Evol 3:217–223.

697   38. Alex A, Antunes A. 2018. Genus-wide comparison of Pseudovibrio bacterial genomes

698   reveal diverse adaptations to different marine invertebrate hosts. PLoS One

699   13:e0194368.

700   39. Pasolli E, Schiffer L, Manghi P, Renson A, Obenchain V, Truong DT, Beghini F, Malik F,

701   Ramos M, Dowd JB, Huttenhower C, Morgan M, Segata N, Waldron L. 2017.

702   Accessible, curated metagenomic data through ExperimentHub. Nat Methods 14:1023–

703   1024.

704   40. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2.

705   Genome Biol 20:257.

706   41. Lu J, Breitwieser FP, Thielen P, Salzberg SL. 2017. Bracken: estimating species

707   abundance in metagenomics data. PeerJ Computer Science 3:e104.

708   42. de la Cuesta-Zuluaga J, Ley RE, Youngblut ND. 2020. Struo: a pipeline for building

709   custom databases for common metagenome profilers. Bioinformatics 36:2314–2315.

710   43. Griffith DM, Veech JA, Marsh CJ. 2016. cooccur : Probabilistic Species Co-Occurrence

711   Analysis in R. J Stat Softw 69:1–17.

712   44. Quinn TP, Richardson MF, Lovell D, Crowley TM. 2017. propr: An R-package for

713   Identifying Proportionally Abundant Features Using Compositional Data Analysis. Sci

714   Rep 7:16252.

715   45. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models

716   Using lme4. J Stat Softw 67:1–48.

717   46. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van

718   Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. 2014. Human genetics

719   shape the gut microbiome. Cell 159:789–799.

720   47. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohney RP, Li W, de Rinaldis E,

721   Bell JT, Venter JC, Nelson KE, Spector TD, Falchi M. 2019. Interplay between the

722   human gut microbiome and host metabolism. Nat Commun 10:4505.

723   48. Goodrich JK, Davenport ER, Clark AG, Ley RE. 2017. The Relationship Between the

724   Human Genome and Microbiome Comes into View. Annu Rev Genet 51:413–433.

725   49. Paul K, Nonoh JO, Mikulski L, Brune A. 2012. "Methanoplasmatales,"

726   Thermoplasmatales-related archaea in termite guts and other environments, are the

727   seventh order of methanogens. Appl Environ Microbiol 78:8245–8253.

728   50. Borrel G, O'Toole PW, Harris HMB, Peyret P, Brugère J-F, Gribaldo S. 2013.

729   Phylogenomic data support a seventh order of Methylotrophic methanogens and

730   provide insights into the evolution of Methanogenesis. Genome Biol Evol 5:1769–1780.

731   51. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN,

732   Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled

733   genomes substantially expands the tree of life. Nat Microbiol 2:1533–1542.

734   52. Borrel G, McCann A, Deane J, Neto MC, Lynch DB, Brugère J-F, O'Toole PW. 2017.

735   Genomics and metagenomics of trimethylamine-utilizing Archaea in the human gut

736   microbiome. ISME J 11:2059–2074.

737   53. Quinn TP, Erb I, Richardson MF, Crowley TM. 2018. Understanding sequencing data as

738   compositions: an outlook and review. Bioinformatics 34:2870–2878.

739   54. Fennema D, Phillips IR, Shephard EA. 2016. Trimethylamine and Trimethylamine N-

740   Oxide, a Flavin-Containing Monooxygenase 3 (FMO3)-Mediated Host-Microbiome

741   Metabolic Axis Implicated in Health and Disease. Drug Metab Dispos 44:1839–1850.

742   55. Begley M, Gahan CGM, Hill C. 2005. The interaction between bacteria and bile. FEMS

743   Microbiol Rev 29:625–651.

744   56. Hovey R, Lentes S, Ehrenreich A, Salmon K, Saba K, Gottschalk G, Gunsalus RP,

745   Deppenmeier U. 2005. DNA microarray analysis of Methanosarcina mazei Gö1 reveals

746   adaptation to different methanogenic substrates. Mol Genet Genomics 273:225–239.

747   57. Kamke J, Kittelmann S, Soni P, Li Y, Tavendale M, Ganesh S, Janssen PH, Shi W,

748   Froula J, Rubin EM, Attwood GT. 2016. Rumen metagenome and metatranscriptome

749   analyses of low methane yield sheep reveals a Sharpea-enriched microbiome

750   characterised by lactic acid formation and utilisation. Microbiome 4:56.

751 58. LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M. 2013.

752   Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr Opin

753   Biotechnol 24:160–168.

754 59. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, Pollard KS,

755   Sakharova E, Parks DH, Hugenholtz P, Segata N, Kyrpides NC, Finn RD. 2020. A

756   unified catalog of 204,938 reference genomes from the human gut microbiome. Nat

757   Biotechnol 490:55.

758 60. Sridharan GV, Choi K, Klemashevich C, Wu C, Prabakaran D, Pan LB, Steinmeyer S,

759   Mueller C, Yousofshahi M, Alaniz RC, Lee K, Jayaraman A. 2014. Prediction and

760   quantification of bioactive microbiota metabolites in the mouse gut. Nat Commun

761   5:5492.

762 61. Ruaud A, Esquivel-Elizondo S, de la Cuesta-Zuluaga J, Waters JL, Angenent LT,

763   Youngblut ND, Ley RE. 2020. Syntrophy via interspecies H2 transfer between and

764   underlies their global cooccurrence in the human gut. MBio 11:e03235–19.

765 62. Ng F, Kittelmann S, Patchett ML, Attwood GT, Janssen PH, Rakonjac J, Gagic D. 2016.

766   An adhesin from hydrogen-utilizing rumen methanogen Methanobrevibacter

767   ruminantium M1 binds a broad range of hydrogen-producing microorganisms. Environ

768   Microbiol 18:3010–3021.

769 63. Podar M, Gilmour CC, Brandt CC, Soren A, Brown SD, Crable BR, Palumbo AV,

770   Somenahally AC, Elias DA. 2015. Global prevalence and distribution of genes and

771   microorganisms involved in mercury methylation. Sci Adv 1:e1500675.

772 64. Horz H-P. 2015. Archaeal Lineages within the Human Microbiome: Absent, Rare or

773   Elusive? Life 5:1333–1345.

774   65. Feldewert C, Lang K, Brune A. 2020. The hydrogen threshold of obligately methyl-

775   reducing methanogens. FEMS Microbiol Lett fnaa137.

776   66. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, Spector TD,

777   Bell JT, Clark AG, Ley RE. 2016. Genetic Determinants of the Gut Microbiome in UK

778   Twins. Cell Host Microbe 19:731–743.

779   67. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A,

780   Le Roy CI, Raygoza Garay JA, Finnicum CT, Liu X, Zhernakova DV, Bonder MJ,

781   Hansen TH, Frost F, Rühlemann MC, Turpin W, Moon J-Y, Kim H-N, Lüll K, Barkan E,

782   Shah SA, Fornage M, Szopinska-Tokov J, Wallen ZD, Borisevich D, Agreus L,

783   Andreasson A, Bang C, Bedrani L, Bell JT, Bisgaard H, Boehnke M, Boomsma DI, Burk

784   RD, Claringbould A, Croitoru K, Davies GE, van Duijn CM, Duijts L, Falony G, Fu J, van

785   der Graaf A, Hansen T, Homuth G, Hughes DA, Ijzerman RG, Jackson MA, Jaddoe

786   VWV, Joossens M, Jørgensen T, Keszthelyi D, Knight R, Laakso M, Laudes M, Launer

787   LJ, Lieb W, Lusis AJ, Masclee AAM, Moll HA, Mujagic Z, Qibin Q, Rothschild D, Shin H,

788   Sørensen SJ, Steves CJ, Thorsen J, Timpson NJ, Tito RY, Vieira-Silva S, Völker U,

789   Völzke H, Võsa U, Wade KH, Walter S, Watanabe K, Weiss S, Weiss FU, Weissbrod O,

790   Westra H-J, Willemsen G, Payami H, Jonkers DMAE, Vasquez AA, de Geus EJC,

791   Meyer KA, Stokholm J, Segal E, Org E, Wijmenga C, Kim H-L, Kaplan RC, Spector TD,

792   Uitterlinden AG, Rivadeneira F, Franke A, Lerch MM, Franke L, Sanna S, D'Amato M,

793   Pedersen O, Paterson AD, Kraaij R, Raes J, Zhernakova A. 2020. Genetics of human

794   gut microbiome composition. biorxiv;2020.06.26.173724v1. Genetics. bioRxiv.

795   68. Roehe R, Dewhurst RJ, Duthie C-A, Rooke JA, McKain N, Ross DW, Hyslop JJ,

796   Waterhouse A, Freeman TC, Watson M, Wallace RJ. 2016. Bovine Host Genetic

797   Variation Influences Rumen Microbial Methane Production with Best Selection Criterion

798   for Low Methane Emitting and Efficiently Feed Converting Hosts Based on

799   Metagenomic Gene Abundance. PLoS Genet 12:e1005846.

800   69. Difford GF, Plichta DR, Løvendahl P, Lassen J, Noel SJ, Højberg O, Wright A-DG, Zhu

801   Z, Kristensen L, Nielsen HB, Guldbrandtsen B, Sahana G. 2018. Host genetics and the

802   rumen microbiome jointly associate with methane emissions in dairy cows. PLoS Genet

803   14:e1007580.

804   70. Hania WB, Ballet N, Vandeckerkove P, Ollivier B, O'Toole PW, Brugère J-F. 2017.

805   Archaebiotics: Archaea as Pharmabiotics for Treating Chronic Disease in Humans?, p.

806   42–62. *In* Sghaier, H, Najjari, A, Ghedira, K (eds.), Archaea - New Biocatalysts, Novel

807   Pharmaceuticals and Various Biotechnological Applications. InTech.

808   71. Douglas AE. 2020. The microbial exometabolome: ecological resource and architect of

809   microbial communities. Philos Trans R Soc Lond B Biol Sci 375:20190250.

810 **Figure and table legends**

811 *Figure legends*

812 **Figure 1. The order *Methanomassiliicoccales* forms two large clades that**

813 **loosely follow the source of isolation.** A maximum likelihood phylogeny of

814 concatenated single-copy marker genes. The gray triangle corresponds to

815 *Thermoplasma acidophilum, Picrophilus oshimae, Ferroplasma acidarmanus,*

816 *Acidiplasma aeolicum* and *Cuniculiplasma divulgatu*m; outgroup taxa from class

817 *Thermoplasmata*. Black circles indicate bootstrap values of > 80 (of 100 bootstrap

818 permutations), and branch color represents the clade. Colored strips show the

819 source of isolation of each of the included genomes and the general category to

820 which the source belongs. Bar plots show the genome abundance enrichment in gut

821 metagenome samples compared to environmental samples calculated using

822 DESeq2; dots indicate taxa with significant enrichment in either host or

823 environmental biome (Adj. P < 0.05). The scale bar represents the number of amino

824 acid substitutions per site.

825

826 **Figure 2. *Methanomassiliicoccales* clades are widespread but not abundant**

827 **across a range of environments and animal hosts.** Combined abundance of

828 representative genomes of the EX (purple), FL (green), HA (orange) clades on

829 metagenome samples from diverse biomes: stomach (n = 12), foregut (23), large

830 intestine (66), fecal (44), desert (4), sand (12), grasslands (8), permafrost (22),

831   sediment (31), coastal (28), intertidal zone (25), lentic (6), groundwater (3), saline

832   (2), hypersaline (9) and Ice (10). Abundances calculated for individual genomes

833   using KrakenUniq and aggregated by clade. Y-axis in logarithmic scale, black points

834   indicate mean relative abundance in percentage, black bars indicate standard

835   deviation.

836

837   **Figure 3. Genome characteristics and adhesion protein repertoire of**

838   ***Methanomassiliicoccales* reflect division of the order into clades, although**

839   **members of the Clade FL not enriched in environmental biomes resemble**

840   **those of the Clade HA.** The phylogeny is the same as shown in Figure 1. The

841   colored strip summarizes the biome enrichment analysis. Heatmaps show genome

842   features including genome GC content (GC; range: 41.26 %, 62.74 %), genome

843   length (Len; 969.311 bp, 2.620.233 bp), and number of predicted genes (Genes;

844   1057, 2607) (blue scale); or repertoire of adhesion proteins: Sel1 containing proteins

845   (Sel1; 0, 29), Listeria-Bacteroides repeat containing proteins (List-Bact; 0, 26),

846   tetratricopeptide repeats (TPR; 7, 40), Ankyrin repeats (ANK; 0, 3), Leucine-rich

847   repeats (LRR; 0, 9), Fibronectin type III (FN3; 0, 20) domains, Bacterial Ig-like

848   domains (Ig-like; 0, 12), YadA-like domain (YadA; 0, 1) and adhesin-like proteins

849   (ALP; 0, 12) (gray scale; columns ordered by hierarchical clustering).On both

850   heatmaps the color intensity of each feature is relative to the maximum value of each

851   category. Scale bar represent the number of amino acid substitutions per site.

852

853 **Figure 4. Ordination of gene content of *Methanomassiliicoccales* group taxa by**

854 **phylogenetic clade rather than by biome enrichment.** Principal component

855 analysis of the gene cluster presence of taxa from clades FL (green), HA (orange)

856 and EX (purple). Highlighted points correspond to outliers: taxa either not

857 significantly enriched in environmental or gut biomes, or with enrichment opposite to

858 the expectation given their clade.

859

860 **Figure 5. Members of *Methanomassiliicoccales* are rare members of the human**

861 **gut microbiota.** Scatter plots of the genera A) *ca.* Methanomethylophilus and B)

862 *Methanomassiliicoccus* show that their prevalence and mean abundance is low

863 across most studies and populations (n = 4472; 35 datasets) with subjects from

864 Australia (AUT), China (CHN), Denmark (DNK), Ethiopia (ETH), Fijo (FJI), Great

865 Britain (GBR), Ghana (GHA), Israel (ISR), Madagascar (MDG), Mongolia (MNG),

866 The Netherlands (NDL), El Salvador (SLV), Sweden (SWE), Tanzania (TZA) and the

867 United States (USA).

868

869 **Figure 6. Coabundance networks of *Methanomassiliicoccus* (green node, dark**

870 **edges) and "*ca.* Methanomethylophilus" (orange node, light edges) in the**

871 **human gut largely overlap.** Both *Methanomassiliicoccales* genera are significantly

872 co-abundant (cyan edge). Their abundances are also coordinated with another

873   archaeon (blue node) and TMA-producing bacterial taxa (red nodes).

874

875   ***Table legends***

876   **Table 1.** InterPro, eggNOG and Prokka annotations of gene clusters significantly

877   enriched on clade HA compared to Clade FL. Four gene clusters with no annotation

878   were omitted.

879

880   ***Supplementary figure legends***

881   **Figure S1. *Methanomassiliicoccales* taxa from all clades are widespread but**

882   **not abundant across a range of environments and animal hosts.** The

883   abundance of members of the FL and HA clades is comparable within similar

884   biomes, in particular, animal derived metagenomes. Abundance of each

885   representative genome on diverse metagenome and environmental metagenome

886   samples colored by clade (FL: green, HA: orange, EX: purple). Abundances

887   calculated for individual genomes using KrakenUniq and aggregated by clade. Note

888   that the Y-axis is in logarithmic scale and each plot has a different scale. Black

889   points indicate mean relative abundance in percentage, black bars indicate standard

890   deviation. Metagenome samples from stomach (n=12), foregut (23), large intestine

891   (66), fecal (44), desert (4), sand (12), grasslands (8), permafrost (22), sediment (31),

892   coastal (28), intertidal zone (25), lentic (6), groundwater (3), saline (2), hypersaline

893   (9) and Ice (10).

894

895 **Figure S2. Small clusters of unknown function dominate the pangenome of the**

896 **order *Methanomassiliicoccales*.** Gene cluster frequency spectrum of the order

897 Methanomassiliicoccales separated by (A) unknown or (B) known function. (C)

898 Fraction of gene clusters belonging to each COG category per clade. Core clusters

899 were defined as present in ≥ 80% of genomes of a clade; for the complete order,

900 gene clusters were present in ≥ 80% of the included genomes and at least one

901 member of each clade. The proportion of clusters of unknown functions in the core

902 genome of each clade was large and varied between clades, ranging from 23.0 % in

903 Clade HA to 38.5 % in Clade EX. The proportion of unknown clusters was lowest in

904 the complete taxonomic order, where it only accounted for 14.7 % of gene clusters.

905 COG functional classification descriptions by groups. Information Storage and

906 processing: (B) Chromatin structure and dynamic, (J) Translation, ribosomal

907 structure and biogenesis, (K) Transcription, (L) Replication, recombination and

908 repair. Cellular processes and signaling: (D) Cell cycle control, cell division,

909 chromosome partitioning, (M) Cell wall/membrane/envelope biogenesis, (N) Cell

910 motility, (O) Post translational modification, protein turnover, chaperone, (T) Signal

911 transduction mechanisms, (U) Intracellular trafficking, secretion, and vesicular,

912 transport, (V) Defense mechanisms, (Z) Cytoskeleton. Metabolism: (C) Energy

913 production and conversion, (E) Amino acid transport and metabolism, (F) Nucleotide

914 transport and metabolism, (G) Carbohydrate transport and metabolism, (H)

915   Coenzyme transport and metabolism, (I) Lipid transport and metabolism, (P)

916   Inorganic ion transport and metabolism, (Q) Secondary metabolites biosynthesis,

917   transport and catabolism. Poorly characterized: (X) No annotation retrieved, (S)

918   Function unknown.

919

920   ***Supplementary table legends***

921   **Table S1.** CBI assembly accession, genome characteristics, study information and

922   source of isolation of 71 publicly available genomes from the order

923   Methanomassiliicoccales retrieved from NCBI in June 2018, plus the ca. M. alvus

924   MAG here reported. Study accession and title of UBA genomes obtained from

925   supplementary tables of Parks et al., 2017 (doi: 10.1038/s41564-017-0012-7),

926   otherwise, obtained from NCBI bioproject.

927

928   **Table S2.** SRA and MGnify accession information of publicly available metagenome

929   samples from gastrointestinal and environmental biomes

930

931   **Table S3.** SRA, study and country information of publicly available human gut

932   metagenome samples

933

934   **Table S4.** Prevalence and mean abundance of candidatus Methanomethylophilus

935   and Methanomassiliicoccus across multiple human populations.

936

937 **Table S5.** Intraclass correlation coefficients (ICC) of and FRD-adjusted P values of

938 relative abundances Methanomassiliicoccus and other control taxa on monozygotic

939 and dizygotic twins

940

941
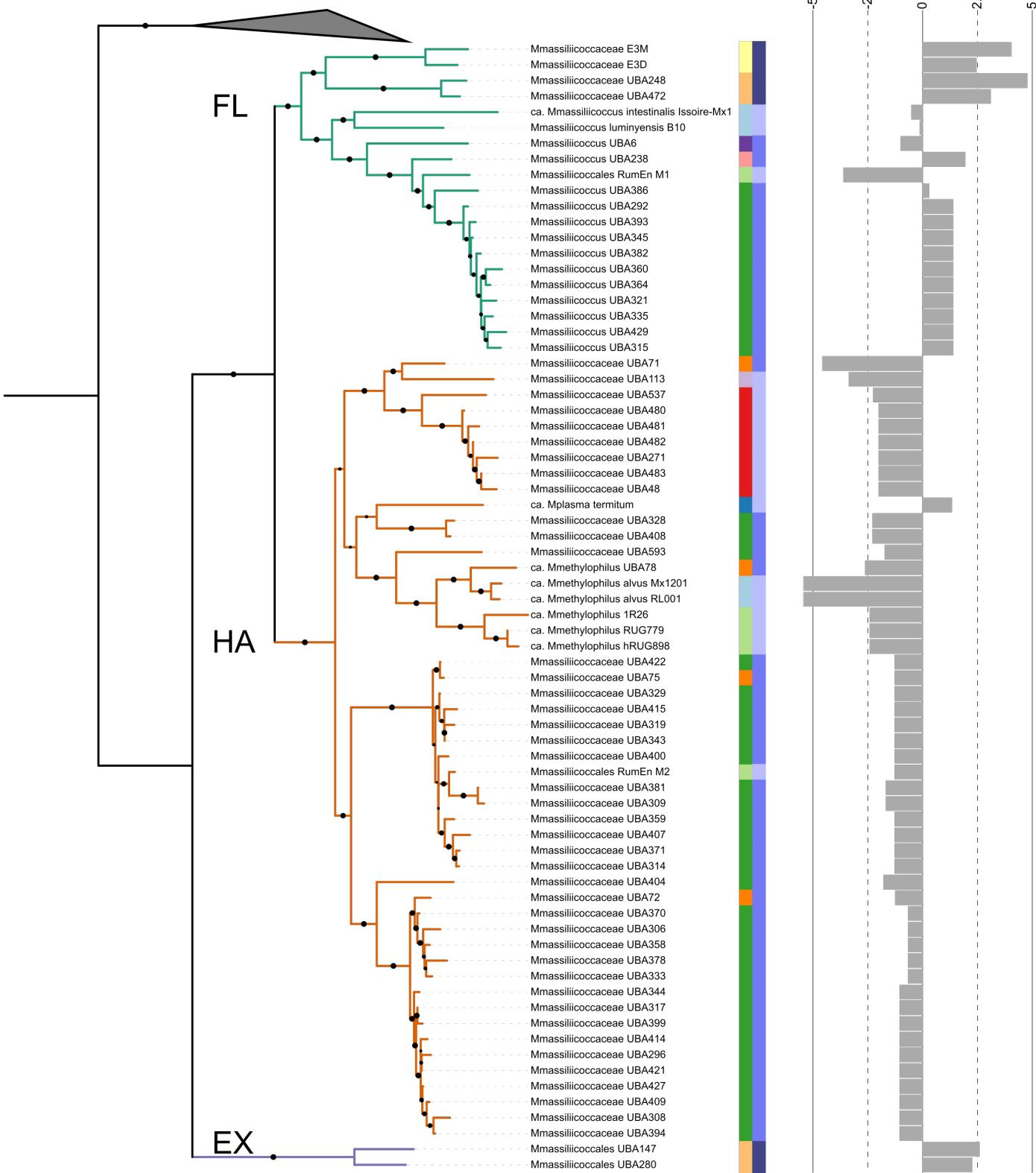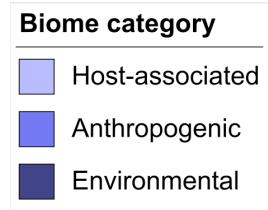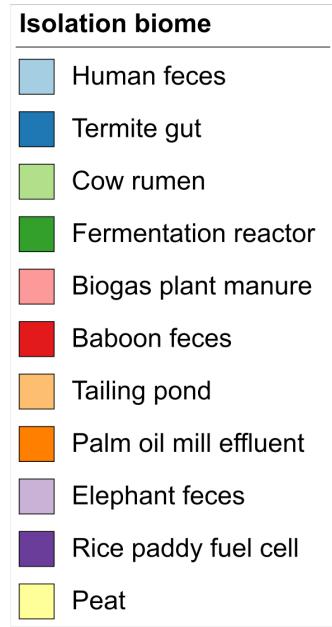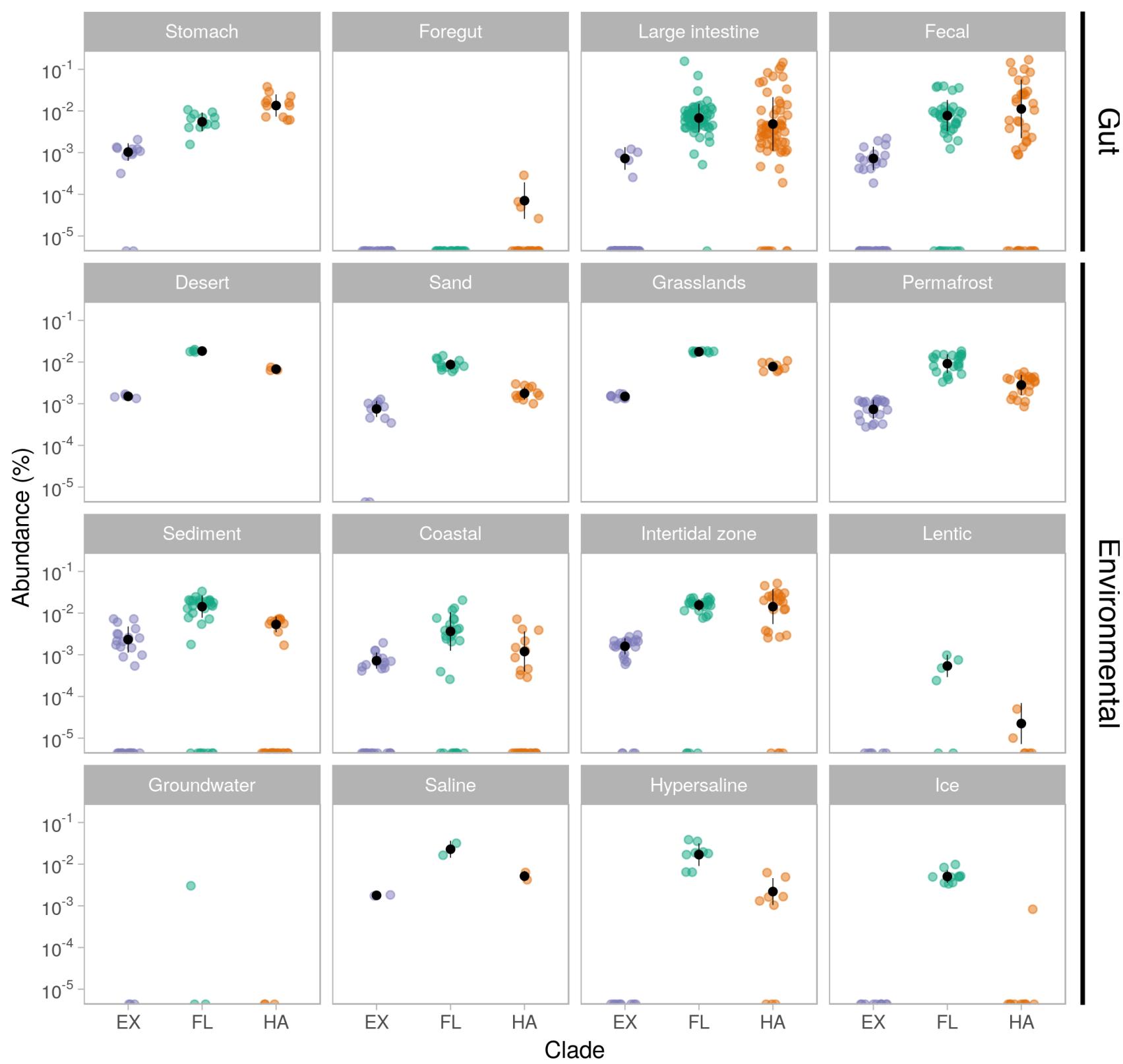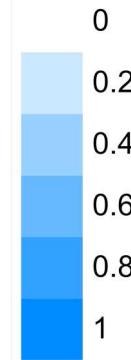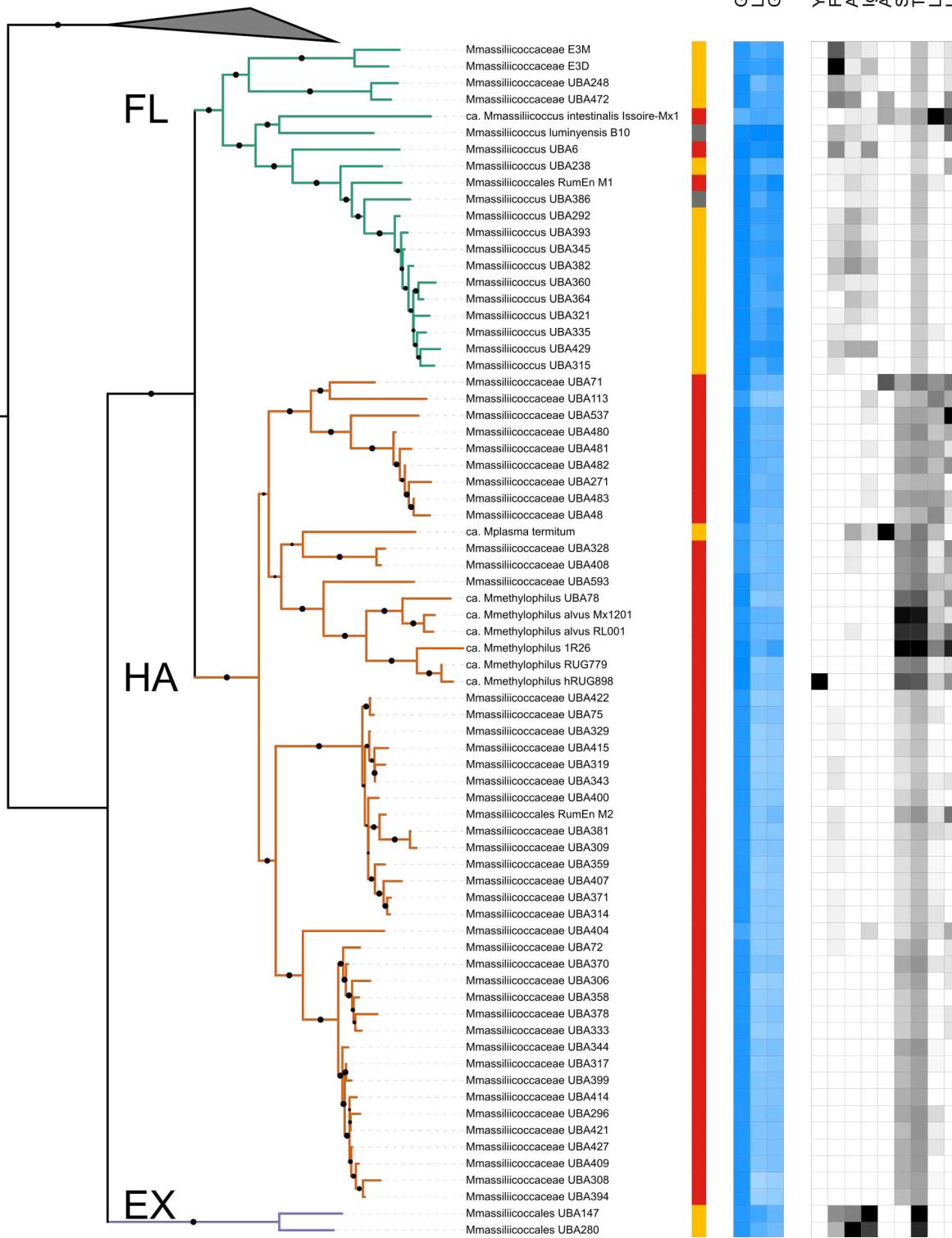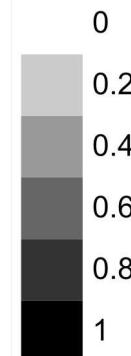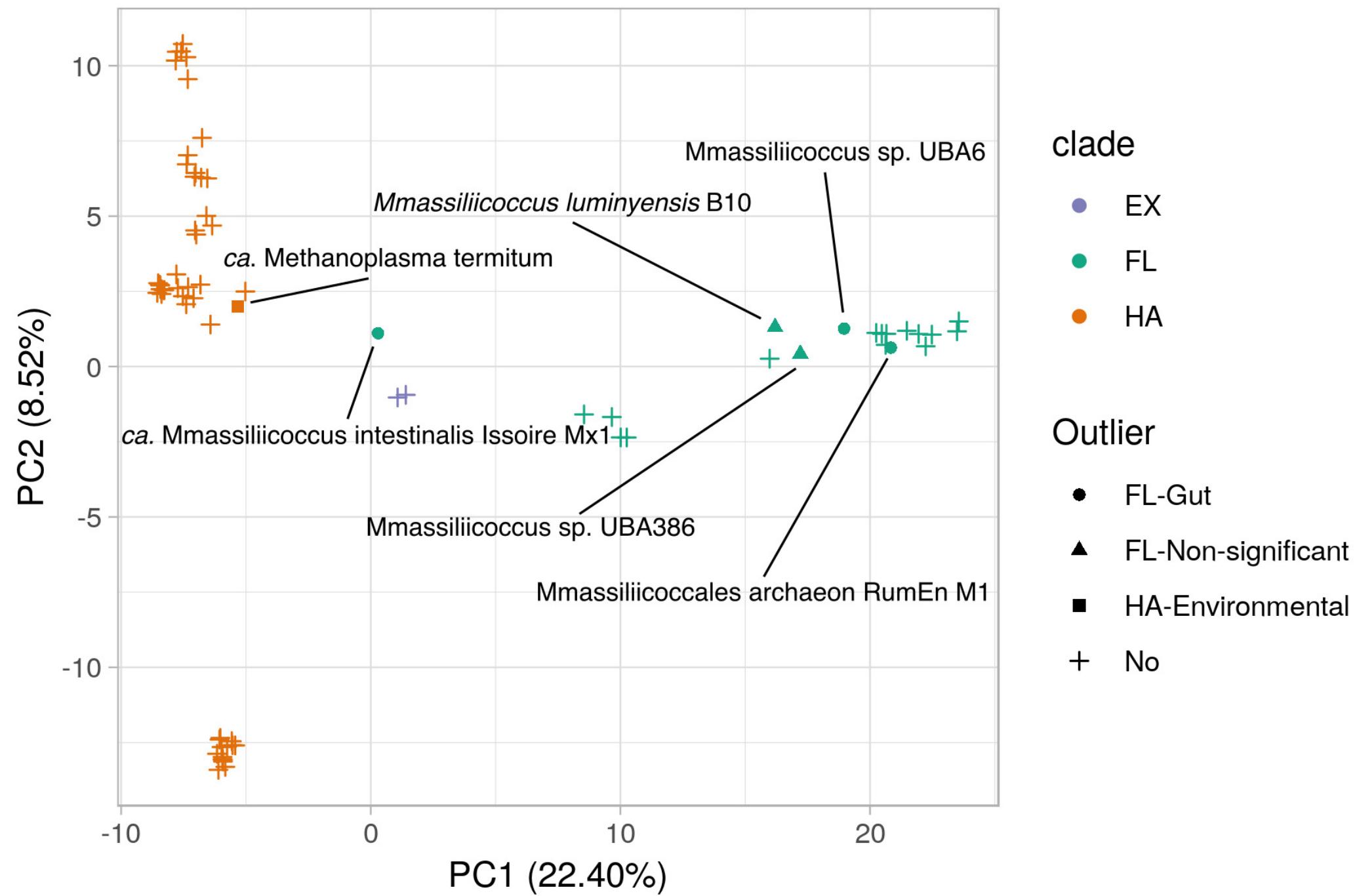
A **ca. Methanomethylophilus**
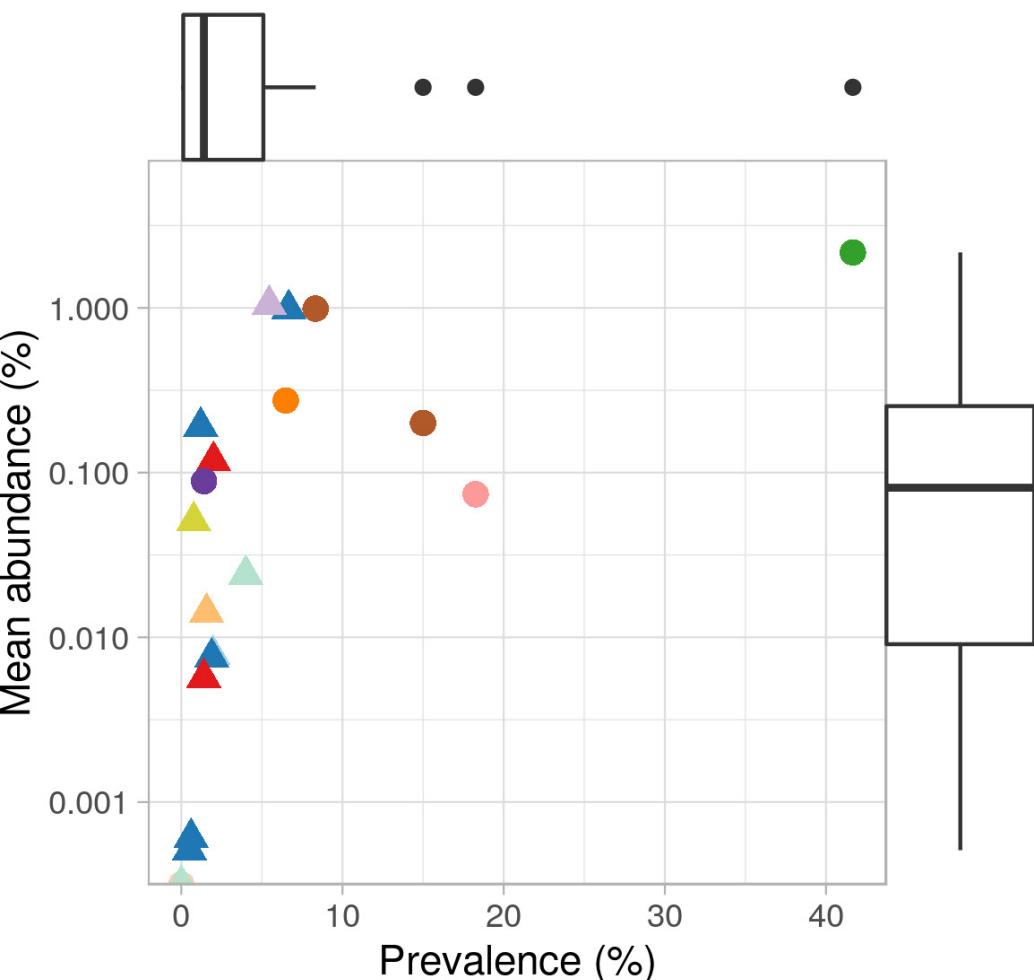
B **Methanomasiliicoccus**

Country: AUT, CHN, DNK, ETH, FJI, GBR, GHA, ISR, MDG, MNG, NLD, SLV, SWE, TZA, USA
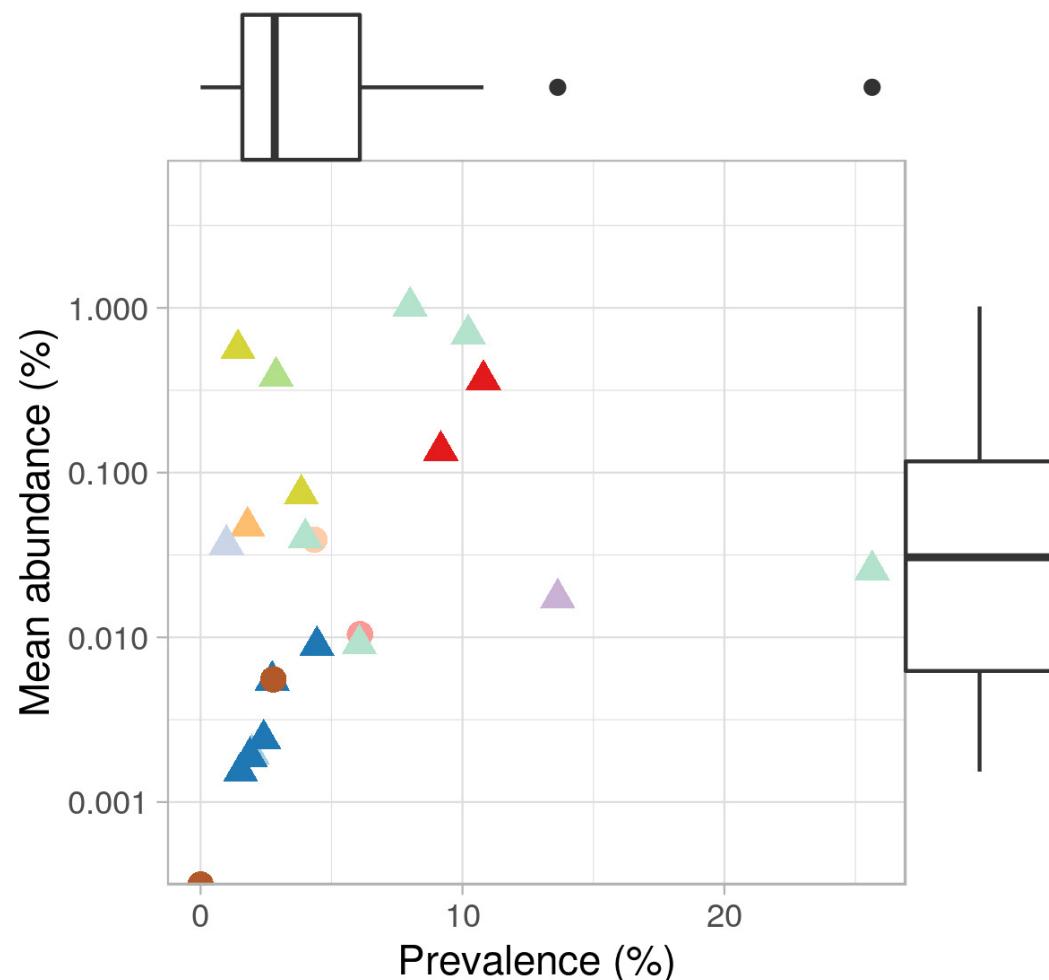
Westernization: Non-Westernized (●), Westernized (▲)
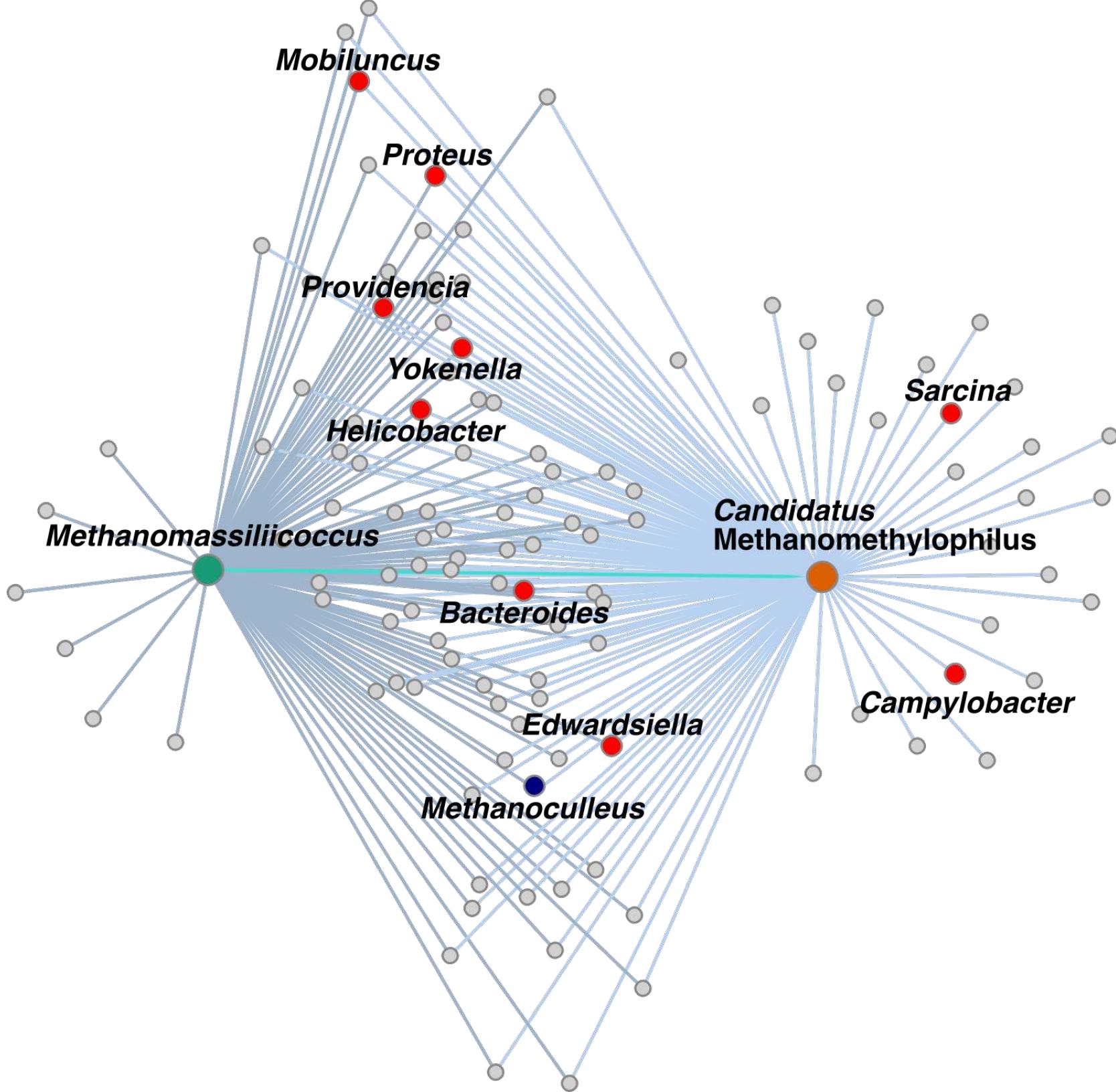
Interpro accession	Interpro annotation	NOG accession	COG category	Prokka gene name	Prokka annotation
IPR001086	Prephenate dehydratase	COG0077@NOG	E	pheA	Prephenate dehydratase
IPR002657	Bile acid:sodium symporter/arsenical resistance protein Acr3	COG0385@NOG	S	-	hypothetical protein
IPR002701	Chorismate mutase II, prokaryotic-type	COG1605@NOG	E	aroQ	Chorismate mutase
IPR002732	Holliday junction resolvase Hjc	COG1591@NOG	L	rutD	Putative aminoacrylate hydrolase RutD
IPR003761	Exonuclease VII, small subunit	COG1722@NOG	L	xseB	Exodeoxyribonuclease 7 small subunit
IPR006357	HAD-superfamily hydrolase, subfamily IIA	COG0647@NOG	G	gph	Glyceraldehyde 3-phosphate phosphatase
IPR013022	Xylose isomerase-like, TIM barrel domain	11IHC@NOG	L	-	hypothetical protein
IPR015867	Nitrogen regulatory protein PII/ATP phosphoribosyltransferase, C-terminal	COG3323@NOG	S	-	hypothetical protein
IPR019271	Protein of unknown function DUF2284, metal-binding	11RTN@NOG	S	-	hypothetical protein
IPR029068	Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase	-	X	-	hypothetical protein