



# Genomic Insights into Adaptations of Trimethylamine-Utilizing Methanogens to Diverse Habitats, Including the Human Gut

 Jacobo de la Cuesta-Zuluaga,<sup>a</sup>  Tim D. Spector,<sup>b</sup>  Nicholas D. Youngblut,<sup>a</sup>  Ruth E. Ley<sup>a</sup>

<sup>a</sup>Department of Microbiome Science, Max Planck Institute for Developmental Biology, Tübingen, Germany

<sup>b</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

**ABSTRACT** *Archaea* of the order *Methanomassiliicoccales* use methylated amines such as trimethylamine as the substrates for methanogenesis. They form two large phylogenetic clades and reside in diverse environments, from soil to the human gut. Two genera, one from each clade, inhabit the human gut: *Methanomassiliicoccus*, which has one cultured representative, and “*Candidatus Methanomethylophilus*,” which has none. Questions remain regarding their distribution across biomes and human populations, their association with other taxa in the gut, and whether host genetics correlate with their abundance. To gain insight into the *Methanomassiliicoccales* clade, particularly its human-associated members, we performed a genomic comparison of 72 *Methanomassiliicoccales* genomes and assessed their presence in metagenomes derived from the human gut ( $n=4,472$ , representing 22 populations), non-human animal gut ( $n=145$ ), and nonhost environments ( $n=160$ ). Our analyses showed that all taxa are generalists; they were detected in animal gut and environmental samples. We confirmed two large clades, one enriched in the gut and the other enriched in the environment, with notable exceptions. Genomic adaptations to the gut include genome reduction and genes involved in the shikimate pathway and bile resistance. Genomic adaptations differed by clade, not habitat preference, indicating convergent evolution between the clades. In the human gut, the relative abundance of *Methanomassiliicoccales* spp. correlated with trimethylamine-producing bacteria and was unrelated to host genotype. Our results shed light on the microbial ecology of this group and may help guide *Methanomassiliicoccales*-based strategies for trimethylamine mitigation in cardiovascular disease.

**IMPORTANCE** *Methanomassiliicoccales* are less-known members of the human gut archaeome. Members of this order use methylated amines, including trimethylamine, in methane production. This group has only one cultured representative; how its members adapted to inhabit the mammalian gut and how they interact with other microbes is largely unknown. Using bioinformatics methods applied to DNA from a wide range of samples, we profiled the abundances of these *Archaea* spp. in environmental and host-associated microbial communities. We observed two groups of *Methanomassiliicoccales*, one largely host associated and one largely found in environmental samples, with some exceptions. When host associated, these *Archaea* have smaller genomes and possess genes related to bile resistance and aromatic amino acid precursors. We did not detect *Methanomassiliicoccales* in all human populations tested, but when present, they were correlated with bacteria known to produce trimethylamine. Due to their metabolism of trimethylamine, these intriguing *Archaea* may form the basis of novel therapies for cardiovascular disease.

**KEYWORDS** *Methanomassiliicoccales*, archaea, comparative genomics, human gut, metagenomes, microbiome


**Citation** de la Cuesta-Zuluaga J, Spector TD, Youngblut ND, Ley RE. 2021. Genomic insights into adaptations of trimethylamine-utilizing methanogens to diverse habitats, including the human gut. *mSystems* 6:e00939-20. <https://doi.org/10.1128/mSystems.00939-20>.

**Editor** Seth Bordenstein, Vanderbilt University

**Copyright** © 2021 de la Cuesta-Zuluaga et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ruth E. Ley, [rey@tuebingen.mpg.de](mailto:rey@tuebingen.mpg.de).

 A meta-analysis of vertebrate and environmental microbiomes shows trimethylamine (TMA)-utilizing methanogens are adapted to the gut and correlate with TMA-producing bacteria

**Received** 17 September 2020

**Accepted** 14 January 2021

**Published** 9 February 2021

**A**rchaea spp. generally make up a tenth or less of the biomass of the human gut microbiota; however, they are widely prevalent and occupy a unique metabolic niche, utilizing byproducts of bacterial metabolism as the substrates for methanogenesis (1). Members of *Methanobacteriales* are the dominant species of the human gut archaeome (1, 2). These include *Methanobrevibacter smithii*, which uses CO<sub>2</sub>, formate and H<sub>2</sub> as the substrates for methane production (3), and *Methanosphaera stadtmanae*, which consumes methanol and H<sub>2</sub> (4). Through methanogenesis, *Archaea* decrease partial pressures of H<sub>2</sub>, potentially increasing the energetic efficiency of primary fermenters and the production of short-chain fatty acids (5).

A second archaeal lineage, the order *Methanomassiliococcales*, is also found within the human gut, yet its members are less well characterized than those of *Methanobacteriales*. Members of this order, including human-derived *Methanomassiliococcus luminyensis*, "*Candidatus Methanomassiliococcus intestinalis*," and "*Candidatus Methanomethylophilus alvus*," perform H<sub>2</sub>-dependent methylotrophic methanogenesis for their sole energy source (6–8). Their genomes encode several methyltransferases and associated proteins that reduce methylamines, methanol, and methylated sulfides to methane (9). Studies based on 16S rRNA and *mcrA* gene diversity analysis indicate that the order *Methanomassiliococcales* is made up of two large clades, which mostly group species that have either a free-living (FL) or host-associated (HA) lifestyle (10, 11). Based on analyses of the genomes from three human-derived species from both clades, Borrel et al. (9) suggested that each clade colonized the mammalian gut independently. Members of the HA clade, including the human-associated "*Ca. M. alvus*," might be expected to show adaptations similar to those of other methanogens from the gut microbiota (12, 13). How members of the FL clade, including the human-associated *M. luminyensis* and "*Ca. M. intestinalis*," have converged on the gut niche remains to be explored.

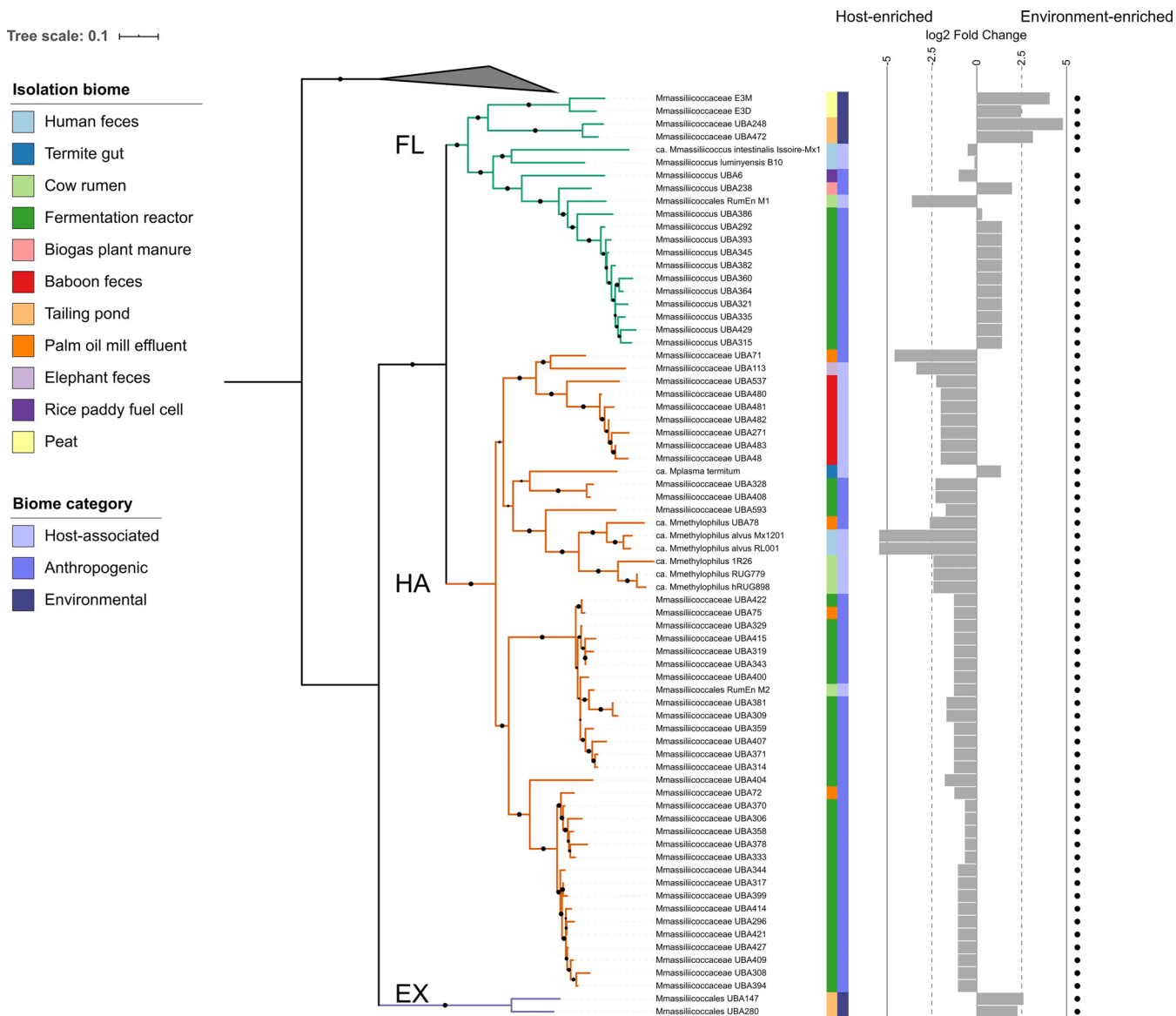
A better understanding of the ecology of *Methanomassiliococcales* may be of interest to human health, as they can utilize mono-, di-, and trimethylamine (TMA) as the substrates for methanogenesis in the gut (14). TMA, a byproduct of bacterial metabolism of carnitine, choline, and other compounds, is transformed in the liver into trimethylamine N-oxide (TMAO) (15). Circulating TMAO inhibits cholesterol transport and promotes its accumulation in macrophages, inducing the formation of atherosclerotic plaques (16). Decreasing TMA levels in the gut and reducing circulating TMAO levels have been proposed as a therapeutic strategy for cardiovascular disease (17). One way to use the gut microbiome to this end would be to boost levels of *Methanomassiliococcales* (18). To accomplish this goal requires a deeper understanding of its ecology.

Here, we conducted a comparative analysis of 71 *Methanomassiliococcales* genomes, together with an additional metagenome-assembled genome (MAG) corresponding to a strain of "*Ca. M. alvus*" that we retrieved by metagenome assembly of gut samples from subjects of the United Kingdom Adult Twin Registry (TwinsUK) cohort (19). We used 305 publicly available metagenomes to assess the prevalence of taxa across various habitat types. While the two large clades grouping host-associated (HA) and free-living (FL) taxa are generally enriched in host-associated and environmental metagenomes, a few exceptions stand out. Our results showed that the repertoire of adhesion proteins encoded by the genomes of taxa from each clade tended to differ. Genes involved in bile resistance and the shikimate pathway are likely involved in the adaptation to the gut environment of members of the HA clade, but not for the FL clade. Thus, gut-adapted members converged on life in the gut using different genomic adaptations. *Methanomassiliococcales* genera present in the human gut positively correlate with TMA-producing bacteria.

## RESULTS

### Genome-based phylogeny confirms two large *Methanomassiliococcales* clades.

Based on whole-genome phylogenetic analysis, the order *Methanomassiliococcales* forms two clades with robust support (Fig. 1). This phylogeny is in agreement with



**FIG 1** The order *Methanomassiliococcales* forms two large clades that loosely follow the source of isolation. Maximum-likelihood phylogeny of concatenated single-copy marker genes. The gray triangle corresponds to *Thermoplasma acidophilum*, *Picrophilus oshimae*, *Ferroplasma acidarmanus*, *Acidiplasma aeolicum*, and *Cuniculiplasma divulgatum*, which are outgroup taxa from class *Thermoplasmata*. Black circles indicate bootstrap values of >80 (of 1,000 bootstrap permutations), branch color represents the clade, and the scale bar represents the number of amino acid substitutions per site. Colored strips show the source of isolation of each of the included genomes and the general category to which the source belongs. Bar plots show the genome abundance enrichment in gut metagenome samples compared to environmental samples, calculated using DESeq2; dots indicate taxa with significant enrichment in either host or environmental biomes (adj.  $P < 0.05$ ). Mnassiliococcaceae, *Methanomassiliococceae*; Mmassiliococcus, *Methanomassiliococcus*; Mplasma, *Methanoplasma*; Mmethylophilus, *Methanomethylophilus*; Mmassiliococcales, *Methanomassiliococcales*.

previously reported phylogenies based on 16S rRNA and *mcrA* genes (10, 20, 21). A third distal clade was formed by two closely related MAGs generated in a recent massive metagenome assembly effort (22), which we labeled external (EX) (Fig. 1). We use the terminology of Borrel et al. (23), as follows: the clade including *Methanomassiliococcus* is labeled free living (FL), and the clade containing “*Candidatus* Methanomethylophilus” is labeled host associated (HA).

As observed previously (10), the reported source of the genomes was not always consistent with the clade in which it was grouped. For instance, while publicly available genomes originally retrieved from human, baboon, elephant, and cow gastrointestinal tracts were related to “*Candidatus* Methanomethylophilus” (HA), this clade also contained MAGs derived from digestors and reactors (Fig. 1) reportedly not treating

animal waste (see Table S1A in the supplemental material). Moreover, MAGs retrieved from pit mud of solid-state fermentation reactors used for the production of Chinese liquor were present in both the HA and FL clades (Table S1A). Similarly, “*Ca. M. intestinalis*” Issoire-Mx1, *M. luminyensis* B10, and *Methanomassiliicoccales* archaeon RumEn M1, all retrieved from mammal hosts, grouped in the FL clade.

**Abundance of *Methanomassiliicoccales* clades differs in gastrointestinal and environmental samples.** We assessed the abundance of species-level representative *Methanomassiliicoccales* taxa in publicly available metagenomes that included 145 samples from gastrointestinal tracts of nonhuman animals, such as cats, pigs, elks, cows, mice, white-throated woodrats, trout, chickens, and geese, and 160 environmental samples from sediment, ice, and diverse water and soil sources (Table S1B).

Taxa from all three clades were detected in a wide range of metagenomes from environmental and gut origin. We observed differences in environmental preference by clade. Abundance of taxa from clade EX was highest in environmental metagenomes ( $0.001\% \pm 0.0012\%$ ) (Fig. 1). These were also detected in gut samples ( $0.0002\% \pm 0.0005\%$ ), albeit with a very low abundance, and in fecal ( $0.0003\% \pm 0.0005$ ), large intestine ( $0.0001\% \pm 0.0002\%$ ), and stomach ( $0.0009\% \pm 0.0006\%$ ) metagenomes (Fig. 2). Given their low abundances, further analysis is focused on the FL and HA clades.

The aggregated abundance of clades FL and HA differed across biomes (Fig. 2). In agreement with their names, HA clade members were more abundant in host-associated samples, and FL in non-host-associated samples (Fig. 2). The prevalence and abundance of *Methanomassiliicoccales* taxa varied across animal hosts, yet the overall abundance patterns were consistent across hosts and sample types (see Fig. S2 in the supplemental material).

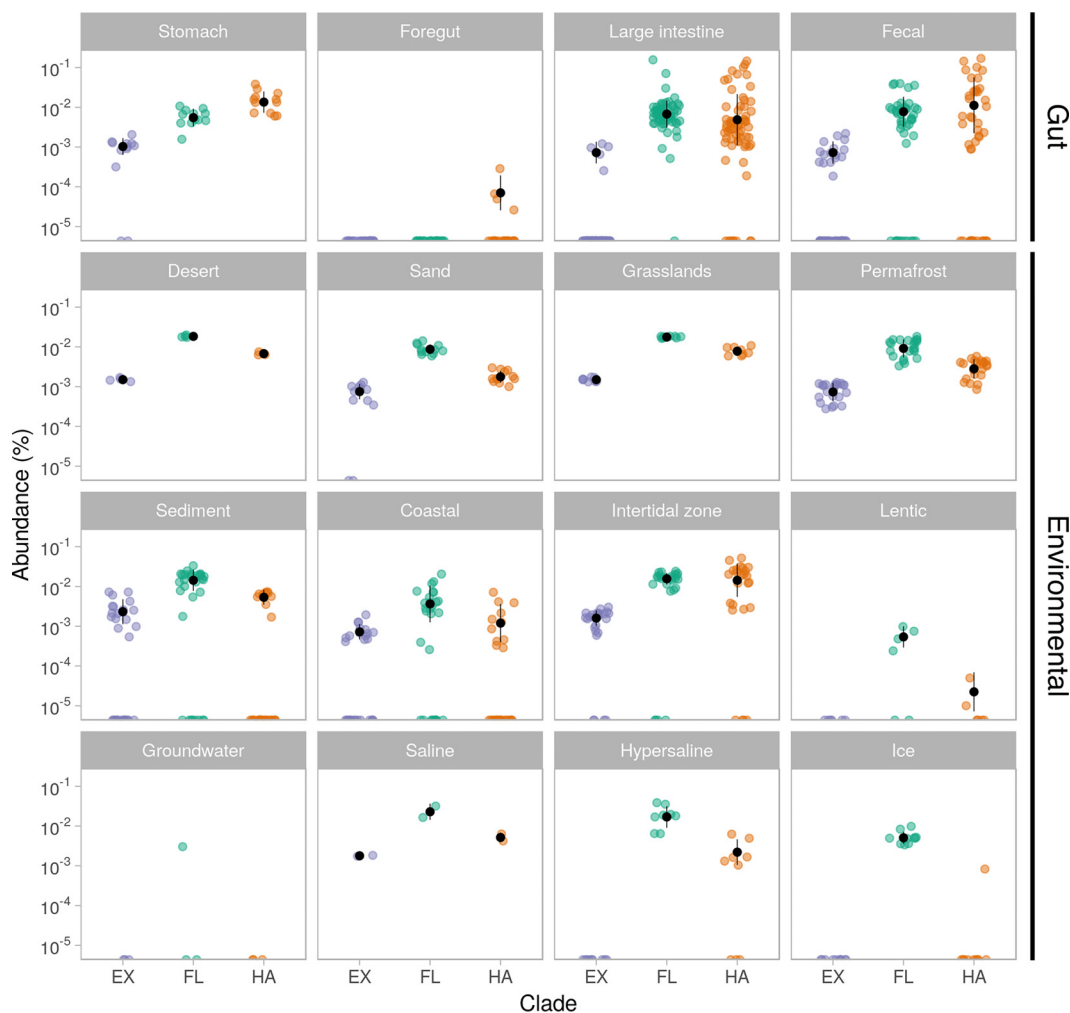
The combined abundance of members of clade FL was higher in samples from environmental biomes ( $0.01\% \pm 0.008\%$ ), although nonzero abundances were observed in digestive system metagenomes ( $0.008\% \pm 0.015\%$ ), with some samples containing levels comparable to that of clade HA (Fig. 2).

The mean abundance of clade HA in aggregate was higher in metagenomes from gut samples ( $0.014\% \pm 0.03\%$ ) compared to those from environmental biomes ( $0.004\% \pm 0.008\%$ ). However, among the environmental biomes, nonzero abundances of clade HA were detected in freshwater ( $0.002\% \pm 0.003\%$ ), marine ( $0.006\% \pm 0.011\%$ ), saline and alkaline ( $0.002\% \pm 0.002\%$ ), and soil ( $0.004\% \pm 0.003\%$ ) samples.

We further validated the differences in clade abundances across biomes by generating a dendrogram of *Methanomassiliicoccales* taxa using the DESeq2-based log fold change of individual taxa on gut versus environmental biomes (i.e., the effect size of the test as a measure of enrichment on a given environment). We then compared the structure of this dendrogram with that of the phylogenomic tree and found that they were positively correlated (cophenetic correlation = 0.67;  $P < 0.01$ ).

Overall, we observed a low abundance of individual *Methanomassiliicoccales* taxa across all samples, ranging from 0 to 0.15% (Fig. 2 and Fig. S1). Their prevalence across hosts differed; they were prevalent in animals such as elks, pigs, poultry, and cattle, while in others, such as trout and geese, they were largely absent (see Fig. S3 in the supplemental material). The enrichment analysis of individual taxa from clade FL from diverse biomes showed that while most were significantly enriched in environmental metagenomes (adjusted [adj.]  $P < 0.1$ ), some taxa showed the opposite enrichment. *M. luminyensis* and *Methanomassiliicoccus* sp. UBA386 were not significantly enriched in gut or environmental biomes. “*Ca. M. intestinalis*” Issoire-Mx1, *Methanomassiliicoccales* archaeon RumEn M1, and *Methanomassiliicoccus* sp. UBA6 were significantly enriched in gut biomes (Fig. 1), although they were also present in multiple environmental biomes (fig. S1).

When assessed on a per-taxon basis, the vast majority of clade HA taxa were significantly enriched in gut samples (adj.  $P < 0.1$ ), with the exception of “*Candidatus* *Methanoplasma termitum*,” which was highly abundant in soil samples from grasslands and water samples from intertidal zones (Fig. 1).



**FIG 2** *Methanomassiliococcales* clades are widespread but not abundant across a range of environments and animal hosts. Combined abundance of representative genomes of the EX (purple), FL (green), and HA (orange) clades in metagenome samples from diverse biomes, as follows: stomach ( $n = 12$ ), foregut ( $n = 23$ ), large intestine ( $n = 66$ ), fecal ( $n = 44$ ), desert ( $n = 4$ ), sand ( $n = 12$ ), grasslands ( $n = 8$ ), permafrost ( $n = 22$ ), sediment ( $n = 31$ ), coastal ( $n = 28$ ), intertidal zone ( $n = 25$ ), lentic ( $n = 6$ ), groundwater ( $n = 3$ ), saline ( $n = 2$ ), hypersaline ( $n = 9$ ), and ice ( $n = 10$ ). Abundances were calculated for individual genomes using KrakenUniq and aggregated by clade. The y axis is in logarithmic scale; black points indicate mean relative abundance in percentage, and black bars indicate standard deviation.

**Genome characteristics and core genes functions differ between *Methanomassiliococcales* clades.** Given the tendency of clades FL and HA to be enriched in environmental or animal metagenomes, respectively, we searched for genes and genome features linked to putative adaptations of *Methanomassiliococcales* to an animal gut. For this, we compared 72 genomes from *Methanomassiliococcales* taxa retrieved from humans, non-human animals, and environmental sources.

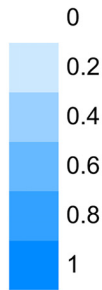
We observed that genomes were more similar to others closely located on the phylogeny in terms of genome GC content, genome length and total gene count (local indicator of phylogenetic association [LIPA] adj.  $P < 0.01$  in all cases) (Fig. 3). To determine whether these features differed between clades, while accounting for the autocorrelation due to evolutionary history, we performed a phylogenetic analysis of variance (ANOVA). Clade FL taxa had significantly larger genomes (mean  $\pm$  standard deviation [SD],  $1,985.1 \pm 245.1$  kb) than either clade HA ( $1,318.3 \pm 187.3$  kb) or clade EX ( $1,872.2 \pm 173.8$  kb) (phylogenetic ANOVA adj.  $P = 0.028$ ). In accordance with this, clade FL also had the highest gene count (FL,  $2,153.1 \pm 233.7$  genes; HA,  $1,377.7 \pm 187.7$  genes; EX,  $1,567.0 \pm 90.5$  genes; adj.  $P = 0.025$ ). While this was nonsignificant, clades HA

Tree scale: 0.1

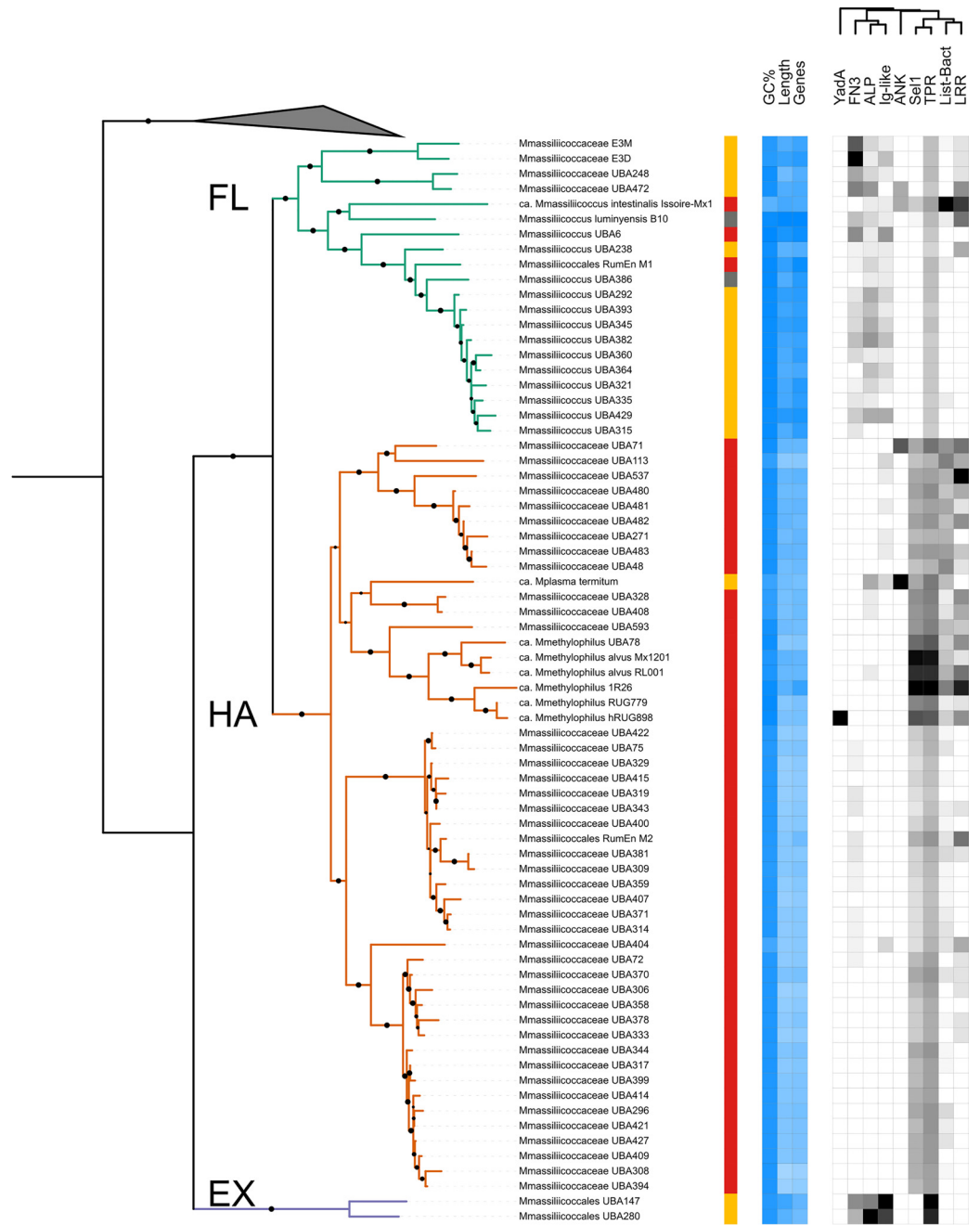
**Biome Enrichment**

- Gut
- Non-significant
- Environmental

**Genome Characteristics**



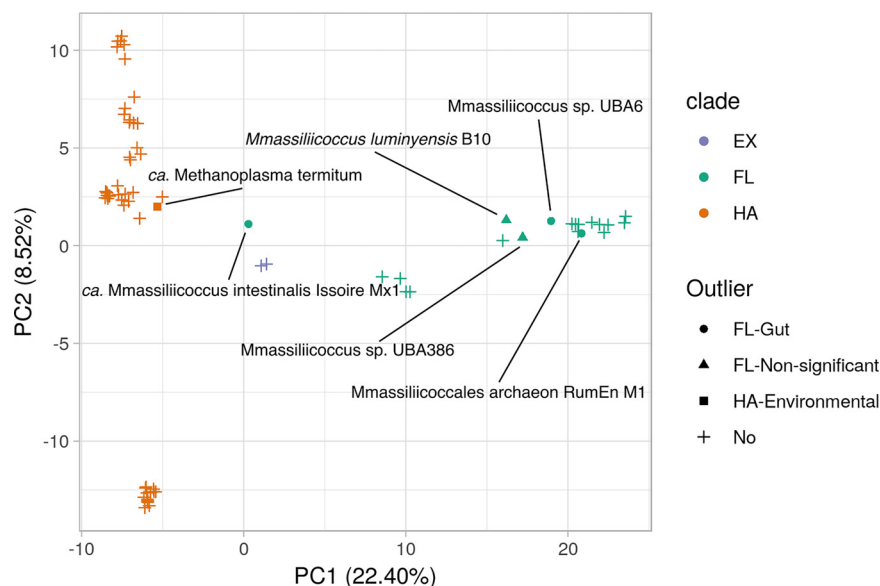
**Adhesion proteins**



**FIG 3** Genome characteristics and adhesion proteins of *Methanomasilliococcales* reflect division of the order into clades. Note that members of clade FL not enriched in environmental biomes resemble those of clade HA. The phylogeny is the same as that shown in Fig. 1. The colored strip summarizes the biome enrichment analysis. Heatmaps show genome features, including genome GC content (“GC%,” range, 41.26% to 62.74%), genome length (“Length,” 969.311 bp to 2.620.233 bp), and number of predicted genes (“Genes,” 1,057 to 2,607) (blue scale), or a repertoire of eukaryote-like proteins: YadA-like domain (YadA; 0, 1), fibronectin type III (FN3; 0, 20) domains, bacterial Ig-like domains (Ig-like; 0, 12), ankyrin repeats (ANK; 0, 3), Sel1-containing proteins (Sel1; 0, 29), tetratricopeptide repeats (TPR; 7, 40), *Listeria-Bacteroides* repeat-containing proteins (List-Bact; 0, 26), and leucine-rich repeats (LRR; 0, 9) (gray scale shows columns ordered by hierarchical clustering); and adhesion-like proteins (ALP; 0, 12). On both heatmaps, the color intensity of each feature is relative to the maximum value of each category.

and EX taxa tended to have a lower GC content than clade FL taxa (FL, 59.1% ± 4.8%; HA, 55.8% ± 2.8%; EX, 54.4% ± 0.5%; adj. *P* = 0.6).

To compare gene presence and absence across clades, we performed a pangenome analysis. After identification of orthologous gene clusters based on sequence similarity using panX software, we obtained 13,695 clusters, of which 7,312 were present at least once in clade FL, 6,592 in clade HA, and 1,833 in clade EX. A large proportion of gene



**FIG 4** Ordination of gene content of *Methanomassiliicoccales* group taxa by phylogenetic clade rather than by biome enrichment. Principal-component analysis of the gene cluster presence of taxa from clades FL (green), HA (orange), and EX (purple). Highlighted points correspond to outliers, namely, taxa either not significantly enriched in environmental or gut biomes or with enrichment opposite to expectation given their clade.

clusters were of unknown function according to the clusters of orthologous genes (COG) functional classification ( $38.4\% \pm 4.3\%$ ); gene clusters of unknown function tended to be detected in one or two genomes (see Fig. S4A in the supplemental material).

Principal component analysis (PCA) of gene cluster presence/absence differentiated clades along principal component 1 (PC1) (Fig. 4). We defined outlier taxa as FL taxa enriched in gut biomes (*Methanomassiliicoccus* archaeon RumEn M1, *Methanomassiliicoccus* sp. UBA6, “*Ca. M. intestinalis*” Issoire-Mx1, *M. luminyensis* B10, and *Methanomassiliicoccus* sp. UBA386) and the HA taxon enriched in non-host biomes (“*Ca. M. termitum*”). Outliers mostly clustered with their close relatives, not with the taxa enriched in the same biome (Fig. 4), with the exception of “*Ca. M. intestinalis*” Issoire-Mx1, which did not cluster with either clade.

#### Gene clusters enriched in clade HA evidence adaptation to the gut environment.

Because of the small number of genomes that cluster within clade EX, and because these are largely absent from animal-associated samples, subsequent analyses focus on comparisons between clades FL and HA.

To identify gene clusters potentially involved in the adaptation of members of Clade HA to a host environment, we compared the gene cluster content between clades. The gene cluster frequency spectrum shows many clusters present in few genomes; 7,990 (58.3%) gene clusters were singletons, and 2,002 (14.6%) were doubletons (Fig. S4A and B). After removing rare gene clusters by filtering those with near-zero variance, we included 2,937 clusters, which we then used to perform in phylogenetic ANOVA. Results reveal 14 gene clusters significantly enriched in HA compared to FL (adj.  $P < 0.1$  in all cases) (Table 1). Three gene clusters are involved in detoxification and xenobiotic metabolism, namely, those encoding bile acid: sodium symporter, bleomycin resistance protein and HAD superfamily hydrolase. Two clusters are related to shikimate or chorismate metabolism, namely, those encoding chorismate mutase II and prephenate dehydratase. Other annotated clusters include the small unit of exonuclease VII, Holliday junction resolvase Hjc, nitrogen regulatory protein PII, xylose isomerase-like protein, and metal-binding domain containing protein; four had poor or no

**TABLE 1** InterPro, eggNOG, and Prokka annotations of gene clusters significantly enriched in clade HA compared to clade FL

InterPro accession no.	InterPro annotation	NOG accession no.	COG category <sup>a</sup>	Prokka gene name	Prokka annotation
IPR002657	Bile acid:sodium symporter/arsenical resistance protein Acr3	COG0385@NOG	S		Hypothetical protein
IPR029068	Glyoxalase/bleomycin resistance protein/dihydroxybiphenyl dioxygenase	-	X		Hypothetical protein
IPR006357	HAD superfamily hydrolase, subfamily IIA	COG0647@NOG	G	<i>gph</i>	Glyceraldehyde 3-phosphate phosphatase
IPR002701	Chorismate mutase II, prokaryotic-type	COG1605@NOG	E	<i>aroQ</i>	Chorismate mutase
IPR001086	Prephenate dehydratase	COG0077@NOG	E	<i>pheA</i>	Prephenate dehydratase
IPR003761	Exonuclease VII, small subunit	COG1722@NOG	L	<i>xseB</i>	Exodeoxyribonuclease 7 small subunit
IPR002732	Holliday junction resolvase Hjc	COG1591@NOG	L	<i>rutD</i>	Putative aminoacrylate hydrolase RutD
IPR015867	Nitrogen regulatory protein PII/ATP phosphoribosyltransferase, C-terminal	COG3323@NOG	S		Hypothetical protein
IPR013022	Xylose isomerase-like, TIM barrel domain	11IHC@NOG	L		Hypothetical protein
IPR019271	Protein of unknown function DUF2284, metal-binding	11RTN@NOG	S		Hypothetical protein

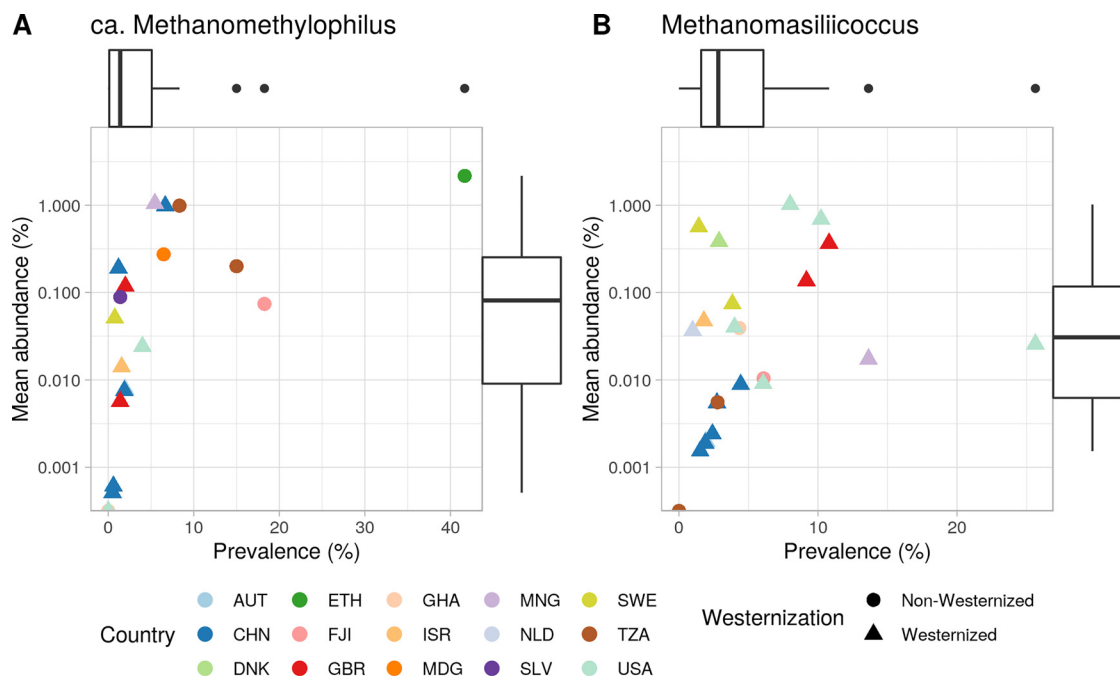
<sup>a</sup>COG functional classification descriptions: E, amino acid transport and metabolism; L, replication, recombination, and repair; G, carbohydrate transport and metabolism; S, function unknown; X, no annotation retrieved.

annotation (Table 1). Similar results were obtained when we performed this analysis without outlier taxa and when biome enrichment was used as independent variable (not shown), further indicating that genomic adaptations differ by clade, not habitat preference. Likewise, 89 clusters were enriched in clade FL compared to HA; these are presented in Table S1E.

**Genomic adaptations to the gut of members of the FL clade.** To determine whether outlier taxa belonging to clade FL had similar adaptations to the gut to those of members of clade HA, we explored gene clusters that were present in these outliers and in clade HA but were rare in other members of clade FL. We selected gene clusters present in the core genome of clade HA (i.e., present in at least 40 taxa or 80% of this clade, see Text S1 in the supplemental material) and present in less than half of the FL taxa. A total of 15 gene clusters were obtained, most of them encoded by only one of the outlier taxa. Two gene clusters, ferrous iron transport proteins A and B (InterPro accession numbers IPR030389 and IPR007167), were present in three of the outliers (*M. luminyensis*, “*Ca. M. intestinalis*” Issoire-Mx1, and *Methanomassiliicoccales* archaeon RumEn M1). Other clusters detected in more than one outlier included an uncharacterized membrane protein (InterPro number IPR005182, in *Methanomassiliicoccus* sp. UBA6 and *Methanomassiliicoccales* archaeon RumEn M1), a putative nickel-responsive regulator (InterPro number IPR014864, in *M. luminyensis* B10 and *Methanomassiliicoccus* sp. UBA386), and an ABC transporter (InterPro number IPR037294, in *M. luminyensis* B10 and *Methanomassiliicoccus* sp. UBA386). The remaining gene clusters, detected once, corresponded to transcriptional regulators or proteins of unknown function.

**The repertoire of adhesion proteins tended to differ between clades HA and FL.** We compared between FL and HA clades two large groups of membrane proteins involved in adhesion, namely eukaryote-like proteins (ELPs), a series of protein families involved in microbial adherence to the host (24), and adhesin-like proteins (ALPs), a class of proteins hypothesized to be involved in the microbe-microbe interactions of *Methanobacteriales* in the gut (13). We aggregated the counts of gene clusters annotated as the ALP and ELP classes and performed phylogenetic ANOVA. This analysis showed a trend toward differing repertoires of adhesion proteins by clade (Fig. 3), although we did not observe significant differences in the frequency of these factors (adj.  $P > 0.1$  in all cases). Taxa from clade HA tended to have higher mean counts of tetratricopeptide repeats (TPR) (mean  $\pm$  SD counts: HA,  $16.30 \pm 6.56$ , and FL,  $9.55 \pm 1.70$ ), Sel1-containing repeats (Sel1) (HA,  $9.32 \pm 5.69$ , and FL,  $0.35 \pm 1.35$ ), *Listeria-Bacteroides* repeats (List-Bact) (HA,  $3.68 \pm 3.76$ , and FL,  $1.65 \pm 5.78$ ), and leucine-rich repeats (LRR) (HA,  $1.5 \pm 2.15$ , and FL,  $1.1 \pm 2.02$ ) than FL taxa. Conversely, adhesin-like proteins (ALPs) (FL,  $2.25 \pm 1.48$ , and





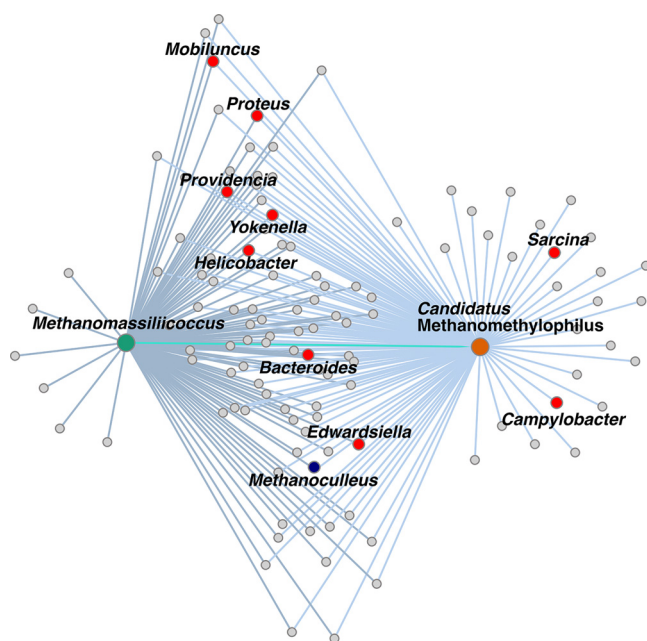
**FIG 5** *Methanomassiliicoccales* are rare members of the human gut microbiota. Scatterplots of the genera (A) “*Ca. Methanomethylophilus*” and (B) *Methanomassiliicoccus* show that their prevalence and mean abundance is low across most studies and populations (35 data sets) with subjects ( $n=4,472$ ) from Austria (AUT), China (CHN), Denmark (DNK), Ethiopia (ETH), Fijo (FJI), Great Britain (GBR), Ghana (GHA), Israel (ISR), Madagascar (MDG), Mongolia (MNG), The Netherlands (NLD), El Salvador (SLV), Sweden (SWE), Tanzania (TZA), and the United States (USA).

HA,  $0.14 \pm 0.61$ ), Ig-like domains (FL,  $1.55 \pm 1.32$ , and HA,  $0.20 \pm 0.53$ ) and fibronectin type III (FN3) domains (FL,  $4.55 \pm 5.09$ , and HA,  $0.32 \pm 0.62$ ) tended to be more abundant in the genomes of members of clade FL. We did not detect invasion protein B (IaIB) in any of the analyzed genomes.

Hierarchical clustering based on the presence or absence of adhesion factors largely grouped *Methanomassiliicoccales* taxa by clade (see Fig. S5 in the supplemental material). Additionally, all adhesion factors, with the exceptions of ankyrin repeats (ANK) and the *Yersinia* adhesin A-like domain (YadA), showed a significant phylogenetic signal (adj.  $P < 0.05$  in all cases), further highlighting that closely related taxa had similar counts.

Interestingly, outlier taxa from clade FL had gene counts of several of the adhesion factors higher than the mean of their own clade and more characteristic of clade HA. In some cases, the gene counts were higher than the mean for clade HA. These included *Listeria-Bacteroides* repeats (gene cluster counts: *M. luminyensis*, 2; “*Ca. M. intestinalis*” Issoire-Mx1, 26; *Methanomassiliicoccales* archaeon RumEn M1, 2), Sel1 repeats (*M. luminyensis*, 1; “*Ca. M. intestinalis*” Issoire-Mx1, 6), and leucine-rich repeats (*M. luminyensis*, 5; “*Ca. M. intestinalis*” Issoire-Mx1, 7).

***Methanomassiliicoccales* taxa cooccur with each other, with other Archaea, and with TMA-producing bacteria in the human gut.** We characterized the distribution of *Methanomassiliicoccales* spp. across a collection of human gut metagenomes derived from 34 studies. Together, the combined 4,472 samples represented people from 22 countries, resulting in 35 unique data sets (i.e., study-country combinations). Across the whole set, we detected just two genera, *Methanomassiliicoccus* (clade FL) and “*Ca. Methanomethylophilus*” (clade HA), both rare members of the human gut microbiota (Fig. 5). “*Ca. Methanomethylophilus*” was detectable in 19 out of 35 data sets; in these 19 data sets, it had a prevalence ranging from 0.5% to 41.7%, and mean abundance ranged from  $4.8 \times 10^{-6}\%$  to  $2.2 \times 10^{-2}\%$ . Similarly, *Methanomassiliicoccus* was detectable in 22 of the 35 data sets; in the 22 data sets, it had a prevalence range of 1% to 25.7% and a mean abundance range of  $1.5 \times 10^{-5}\%$  to  $1.0 \times 10^{-2}\%$  (Table S1D).



**FIG 6** Coabundance networks of *Methanomassiliicoccus* (green node, dark edges) and “*Ca. Methanomethylophilus*” (orange node, light edges) in the human gut largely overlap. Both *Methanomassiliicoccales* genera are significantly coabundant (cyan edge). Their abundances are also coordinated with those of another archaeon (blue node) and TMA-producing bacterial taxa (red nodes).

We tested associations of these two genera with age, sex, and Westernization status of the subjects using linear mixed models that included the data set and country as random effects. Subjects from non-Westernized countries had a significantly higher prevalence of “*Ca. Methanomethylophilus*” (mean prevalence  $\pm$  SD: non-Westernized = 8.9%  $\pm$  28.5%, Westernized = 1.1%  $\pm$  10.3%; adj.  $P = 0.002$ ). Westernized individuals were more likely to harbor higher prevalences of *Methanomassiliicoccus*, although differences were not significant (non-Westernized = 3.9%  $\pm$  19.4%, Westernized, 5.0%  $\pm$  21.7%; adj.  $P > 0.1$ ). The age and sex of the individuals did not explain variance in the prevalence or abundance of either genus (adj.  $P > 0.1$  in all cases).

To identify other microbial taxa positively associated with members of *Methanomassiliicoccales* in the human gut, we calculated a network of positively associated microorganisms (i.e., coabundant taxa) across samples ( $\rho > 0.1$  in all cases) (25). In addition, we determined which taxa were present with members of *Methanomassiliicoccales* at a greater prevalence than that expected by chance (i.e., cooccurring taxa) relative to a permuted null model (26). Results showed that both “*Ca. Methanomethylophilus*” and *Methanomassiliicoccus* were part of the same coabundance network, together with a third archaeal genus, *Methanoculleus* (order *Methanomicrobiales*). We did not find evidence of positive or negative abundance associations of either *Methanomassiliicoccales* genus with *Methanobrevibacter*. Cooccurrence analysis showed a random association pattern between these taxa ( $P > 0.05$  for both “*Ca. Methanomethylophilus*” and *Methanomassiliicoccus*), indicating that their ecological niches do not overlap that of *Methanobrevibacter*.

Analysis of the combined network of “*Ca. Methanomethylophilus*” and *Methanomassiliicoccus* revealed a large overlap between taxa associated with either genus (Fig. 6 and Table S1F): out of 119 taxa in the network, 86 (72.3%) were associated with both. Moreover, 51 taxa (42.9%) also had a significant positive cooccurrence pattern with both genera (adj.  $P < 0.05$  in all cases). Most bacterial members of this network had low relative abundances; only *Bacteroides* and *Parabacteroides* had a mean relative abundance above 1% (range, 22.7% to 0.0005%). Interestingly, they included taxa that can potentially produce TMA, since their genomes contain genes encoding enzymes involved in its

synthesis; these taxa included *Bacteroides*, *Campylobacter*, *Yokenella*, *Mobiluncus*, *Proteus*, *Providencia*, and *Edwardsiella* (27).

**Abundance of *Methanomassiliicoccales* species is not concordant in monozygotic or dizygotic human twins.** To evaluate whether host genetics influences the abundance of *Methanomassiliicoccales* in the human gut, we compared the intraclass correlation coefficient (ICC) of their abundances at the genus level using a set of 153 monozygotic (MZ) and 200 dizygotic (DZ) twin pairs from the TwinsUK cohort. As a control, we first compared the mean ICC across all taxa between MZ and DZ twins and found that  $ICC_{MZ}$  (0.1) was significantly higher than  $ICC_{DZ}$  (0.03) ( $P < 0.01$ ). In addition, we assessed the ICC values of bacterial (*Christensenella*, *Faecalibacterium*, and *Bifidobacterium*) and archaeal (*Methanobrevibacter*) genera, and consistently found a higher correlation for MZ compared to DZ twins (Table S1G). We were only able to assess ICC values of *Methanomassiliicoccus*, as it was the only *Methanomassiliicoccales* taxon detected in the twins with a prevalence (8.64%) above the 5% cutoff (see Materials and Methods). We did not detect a significant concordance between the abundances of *Methanomassiliicoccus* in MZ ( $ICC_{MZ} = 0.004$ ; adj.  $P = 0.59$ ) or in DZ twins ( $ICC_{DZ} = 0.017$ ; adj.  $P = 0.71$ ). Given the low abundance of *Methanomassiliicoccales* taxa, we performed a sensitivity analysis using samples with a high sequencing depth ( $>12$  million reads/sample); however, we did not observe differences in the abundance and prevalence of the *Methanomassiliicoccales* genera or in the ICC estimates (data not shown).

## DISCUSSION

While the source of the members of the *Methanomassiliicoccales* has been noted in previous surveys of single markers such as 16S rRNA and *mcrA* genes (10, 11), here, we searched metagenomes from host-associated and environmental samples for their relative abundances. Overall, the HA taxa were enriched in host-associated samples and the FL taxa were enriched in environmental samples; intriguingly, all taxa, regardless of clade, were detected in both biomes. This suggests that members of the order *Methanomassiliicoccales* are generalists with an overall habitat preference according to clade, although there were some exceptions to the general pattern. We show that members of *Methanomassiliicoccales* use many of the same adaptations to the gut as other methanogens. These adaptations include genome reduction and genes involved in the shikimate pathway and bile resistance. In addition, gut-enriched taxa tend to have a distinct repertoire of genes encoding adhesion factors. We observed that potential adaptations to the gut differed by clade, not preferred habitat, indicating convergence on a shared niche through different genomic solutions. In the human gut, *Methanomassiliicoccales* taxa correlated with TMA-producing bacteria, rather than host genetics or other host factors.

For members of the HA clade, adaptations to life in the gut included an enrichment of genes involved in bile acid transport, efflux pumps, and hydrolases, which play a role in tolerance to these compounds in the gastrointestinal tract (28). This adaptation is also shared with other members of the gut microbiota, including *Methanobacteriales* taxa; *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* are resistant to bile salts (3, 4). Other gene clusters with known functions enriched in clade HA are involved in metabolism of shikimate and chorismate. The shikimate pathway is involved in the synthesis of aromatic amino acids in plants and microbes, but it is absent in mammals. Shikimate metabolism is carried out by archaeal (29) and bacterial (30, 31) members of the animal gut microbiota and was reported as one of the most conserved metabolic modules in a large-scale gene catalogue from the human gut (32). In turn, aromatic amino acids can be transformed by the gut microbiota into active metabolites, which are involved in diverse physiological processes (33) and conditions such as cardiovascular disease (34). Indeed, plasma concentrations of microbial derivatives of tryptophan have even been shown to negatively correlate with atherosclerosis (35). It remains to be elucidated whether *Methanomassiliicoccales* are involved in human health through the metabolism of aromatic amino acids and associated compounds.

We observed that each clade tended to encode different adhesion factors, although without statistical significance. These factors are involved in the maintenance of syntrophic relationships of the methanogens with bacterial (12, 36) or eukaryotic (37) microorganisms. Two groups of adhesion factors, proteins containing Sel1 domains and *Listeria-Bacteroides* repeats, have been previously studied in *Methanomassiliicoccales* taxa retrieved from the gut (9, 23). Our assessment of these factors in the broader context of the order *Methanomassiliicoccales* showed that these two groups are more likely to be higher in clade HA than clade FL taxa, with the exception of the outlier taxa. Indeed, the repertoire of ELPs and ALPs was similar between species inhabiting the gut, regardless of their clade. This emphasizes the potential involvement of these proteins in the adaptation to the intestinal environment, although the exact mechanisms are yet to be elucidated.

In contrast, members of clade FL appear to be generalists that colonized the animal gut independently from the HA clade. It has been previously noted that *M. luminyensis*, an outlier from clade FL, could have a facultative association to the animal gut. It possesses genes involved in nitrogen fixation, oxidative stress (9), and mercury methylation (23), which are common in soil microorganisms but rare in members of the gut microbiota (38). In accordance with this, we observed that members of clade FL are widespread and abundant in soil, water, and gut metagenomes, with a preference for environmental biomes. Similarities in ELP content between gut-dwelling taxa from both clades indicate that interaction with the host or other members of the gut microbiota might be a key factor in the adaptation of these methanogens.

Analysis of the gene content of outlier taxa from clade FL showed that they tended to be more similar to members of their own clade than to taxa from clade HA, with the exception of “*Ca. M. intestinalis*” Issoire-Mx1, which was distinct from either clade FL and HA. In addition, there was little overlap in gene clusters commonly observed in clade HA and outlier taxa from clade FL, with the exception of the adhesion factors discussed above. These observations support the hypothesis that colonization of animal guts by members of *Methanomassiliicoccales* occurred in two independent events (9, 23), and suggests that there is not one solution to life in the gut for these archaea, as members from two clades seem to have solved the problem with a different set of adaptations.

Characterization of the abundance of *Methanomassiliicoccales* taxa across human populations showed members of this group are rare in the microbiota of healthy adults. We did not detect them in all the studied populations, and, when detected, they had low prevalence and abundance. Our extensive analysis of human gut samples corroborates estimates of *Methanomassiliicoccales* prevalence (up to 11%) (23, 39, 40) and mean abundance (below 1%) (23, 41). Differences in *Methanomassiliicoccales* carriage between Westernized and non-Westernized populations remain to be explained, and may be due to diet. While Westernized diets are richer in TMA precursors than non-Western diets (42), intake varies across populations (43).

Our analysis allowed us to assess whether *Archaea* in the human gut are mutually exclusive. We observed positive correlations of “*Ca. Methanomethylophilus*” and *Methanomassiliicoccus* with each other and with *Methanoculleus*, another rare archaeal member of the gut microbiota (45). We did not find evidence of association between members of *Methanomassiliicoccales* and *Methanobrevibacter*, positive or otherwise, confirming the previous report that these methanogens are not mutually exclusive (39); abundance of H<sub>2</sub> in the gut, together with differences in other substrate utilization, might result in nonoverlapping niches (46).

While genus *Methanobrevibacter* was consistently found to have a moderate heritability in the TwinsUK (19, 39, 47) and other cohorts (48, 49), this was not the case for members of *Methanomassiliicoccales*. Similarly to humans, methane production (50) and abundance of *Methanobrevibacter* (51) are also heritable in bovine cattle, but *Methanomassiliicoccales* taxa are not (51). Thus, host genetics might be linked to particular taxa and methanogenesis pathways, not to all *Archaea* or to methane production as a whole.

Genera “*Ca. Methanomethylophilus*” and *Methanomassiliococcus* cooccur with TMA-producing bacteria (27), further supporting their potential use as a way of targeting intestinal TMA (52). The exact nature of the ecological relationships each of these taxa establishes with other members of the microbiome remains to be elucidated. In a facilitation scenario between the methanogens and H<sub>2</sub> and TMA producers, freely available TMA and H<sub>2</sub> required for methylotrophic methanogenesis could be utilized by *Methanomassiliococcales* taxa (53) without cost to the producer. Alternatively, the methanogens could establish syntrophic interactions with other microorganisms, whereby the consumption of these metabolites is also beneficial to the producer (53).

The present study extends our understanding of the order *Methanomassiliococcales* by revealing genomic adaptations to life in the gut by members of both clades that make up this group. Furthermore, the positive correlation between the relative abundances of these TMA-utilizing *Archaea* with TMA-producing bacteria in the gut is a first step toward understanding how they may be harnessed for therapeutic management of gut TMA levels in the context of cardiovascular disease.

## MATERIALS AND METHODS

For a detailed description of the methods, see Text S1 in the supplemental material.

**Genome annotation and phylogenomic tree reconstruction.** We used 71 substantially complete genomes (completeness,  $\geq 70\%$ ) with low contamination (contamination,  $< 5\%$ ) retrieved from the NCBI Assembly database (<https://www.ncbi.nlm.nih.gov/assembly>), plus an additional high-quality metagenome-assembled genome (MAG) corresponding to “*Candidatus Methanomethylophilus alvus*.” Gene calling was performed using Prokka (54). A maximum-likelihood phylogenomic tree was constructed using PhyloPhlAn (55) with the 72 *Methanomassiliococcales* genomes plus members of the order *Thermoplasmatales* as an outgroup. We used interactive Tree Of Life (iTOL) (56) to visualize the tree.

**Abundance of *Methanomassiliococcales* in environmental and animal gastrointestinal metagenomes.** We retrieved 305 publicly available gastrointestinal and environmental metagenome samples (57) (see Table S1B in the supplemental material). To avoid multiple mapping of reads, we dereplicated the 72 genomes at a species level (95% average nucleotide identity [ANI]) using dRep (58), resulting in 29 representative genomes. We quantified the abundance of dereplicated *Methanomassiliococcales* genomes in the metagenomes using KrakenUniq (59). We estimated the enrichment of each representative *Methanomassiliococcales* taxon in host or environmental metagenomes using DESeq2 with the Wald test (60) on sequence counts and classifying metagenome samples as either host derived or environmental.

**Comparative genomics.** We grouped the predicted genes into gene clusters using panX (61) and used InterProScan (62) and eggNOG-mapper (63) for annotation. Phylogenetic signal of genome characteristics and gene cluster presence was tested using the *phyloSignal* R package with the local indicator of phylogenetic association (LIPA) (64). The R package *micropan* (65) was used to create a pangenome principal-component analysis (PCA). We performed phylogenetic ANOVA using the R package *phytools* (66) to determine clusters enriched in clades FL or HA. We adjusted *P* values for multiple comparisons with the Benjamini-Hochberg method. Due to the exploratory nature of this work, tests were considered significant if they had an adjusted *P* value (adj. *P*) of  $< 0.1$ ; a false-discovery-rate-adjusted *P* value cutoff of 0.1 implies that 10% of significant tests will result in false-positives. In cases where adjusting *P* values was not necessary, raw *P* values are provided.

We assessed the presence of eukaryote-like proteins (ELPs) (24) by combining the counts of gene clusters classified as Sel1-containing proteins (Sel1), *Listeria-Bacteroides* repeat-containing proteins (List-Bact), tetratricopeptide repeats (TPR), ankyrin repeats (ANK), leucine-rich repeats (LRR), fibronectin type III (FN3) domains, laminin G domains, bacterial Ig-like domains, *Yersinia* adhesin A-like domain (YadA), TadE-like domain, or invasion protein B (IaIB). Likewise, we characterized the presence of parallel beta-helix-repeat-containing proteins, also known as adhesin-like proteins (ALPs).

**Characterization of *Methanomassiliococcales* distribution across human populations.** We retrieved and quality controlled 4,472 publicly available human gut metagenomes from 34 independent studies (Table S1C). Reads were classified using Kraken (67) and Bracken (68) with custom databases (69). Taxa with  $< 100$  reads in a given sample were considered absent. To determine the cooccurrence patterns of *Methanomassiliococcales* in the human gut, we used the *cooccur* package (26); to determine their coabundance patterns, we calculated the proportionality of taxa abundance ( $\rho$ ) with the *propr* R package (70). The *lme4* and *lmerTest* R packages (71) were used to fit linear mixed effects models to test differences of *Methanomassiliococcales* genera abundance by Westernization status, age, and gender. We employed binomial linear mixed models to test differences in genera prevalence. Lists of potential TMA (27, 72, 73) and methanol (74) (see Text S1 in the supplemental materials) producers were compiled from the available literature.

Heritability of *Methanomassiliococcales* taxa was assessed by comparing relative abundances of taxa within 153 monozygotic (MZ) and 200 dizygotic (DZ) twin pairs from the United Kingdom Adult Twin Registry (TwinsUK) (19, 39, 75). Absolute read counts were transformed using the Yeo-Johnson transformation and adjusted by body mass index (BMI), sex, and sequencing depth (19, 39). We calculated the intraclass correlation coefficient (ICC) in MZ and DZ twins with the *irr* R package, and adjusted *P* values using the Benjamini-Hochberg method. We compared the mean ICC across all taxa between MZ and DZ

twins using the Mann-Whitney test and by assessing the ICC of taxa previously reported as heritable (*Methanobrevibacter*, *Faecalibacterium*, *Christensenella*, and *Bifidobacterium*) (39, 48).

**Data availability.** The raw sequence data are available from the European Nucleotide Archive under study accession number PRJEB40256. Jupyter notebooks are available at <https://github.com/leylabmpi/Methanomassilii>. The “*Candidatus* Methanomethylophilus” MAG generated here can be found at <http://ftp.tue.mpg.de/ebio/projects/Mmassilii/>.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**TEXT S1**, DOCX file, 0.02 MB.

**FIG S1**, TIF file, 2 MB.

**FIG S2**, TIF file, 0.5 MB.

**FIG S3**, TIF file, 2.1 MB.

**FIG S4**, TIF file, 0.5 MB.

**FIG S5**, TIF file, 0.9 MB.

**TABLE S1**, XLSX file, 0.2 MB.

## ACKNOWLEDGMENTS

This work was supported by the Max Planck Society. The study also received support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. We thank EMBO and the organizers and participants of the Bioinformatics and Genome Analyses course held at the Fondazione Edmund Mach in San Michele all’Adige, Italy, for sponsoring the attendance of J.D.L.C.-Z. and for their feedback.

We are also grateful to Daphne Welter, Jessica Sutter, and Albane Ruaud for the fruitful discussions and comments.

We declare no competing interests.

## REFERENCES

- Borrel G, Brugère J-F, Gribaldo S, Schmitz RA, Moissl-Eichinger C. 2020. The host-associated archaeome. *Nat Rev Microbiol* 18:622–636. <https://doi.org/10.1038/s41579-020-0407-y>.
- Moissl-Eichinger C, Pausan M, Taffner J, Berg G, Bang C, Schmitz RA. 2018. Archaea are interactive components of complex microbiomes. *Trends Microbiol* 26:70–85. <https://doi.org/10.1016/j.tim.2017.07.004>.
- Miller TL, Wolin MJ. 1982. Enumeration of *Methanobrevibacter smithii* in human feces. *Arch Microbiol* 131:14–18. <https://doi.org/10.1007/BF00451492>.
- Miller TL, Wolin MJ. 1985. *Methanosphaera stadtmaniae* gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Arch Microbiol* 141:116–122. <https://doi.org/10.1007/BF00423270>.
- Horz H-P, Conrads G. 2010. The discussion goes on: what is the role of *Euryarchaeota* in humans? *Archaea* 2010:967271. <https://doi.org/10.1155/2010/967271>.
- Borrel G, Harris HMB, Tottey W, Mihajlovski A, Parisot N, Peyretailade E, Peyret P, Gribaldo S, O’Toole PW, Brugère J-F. 2012. Genome sequence of “*Candidatus* Methanomethylophilus alvus” Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens. *J Bacteriol* 194:6944–6945. <https://doi.org/10.1128/JB.01867-12>.
- Borrel G, Harris HMB, Parisot N, Gaci N, Tottey W, Mihajlovski A, Deane J, Gribaldo S, Bardot O, Peyretailade E, Peyret P, O’Toole PW, Brugère J-F. 2013. Genome sequence of “*Candidatus* Methanomassiliicoccus intestinalis” Isoire-Mx1, a third thermoplasmatales-related methanogenic archaeon from human feces. *Genome Announc* 1:e00453-13.
- Dridi B, Fardeau M-L, Ollivier B, Raoult D, Drancourt M. 2012. *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int J Syst Evol Microbiol* 62:1902–1907. <https://doi.org/10.1099/ijs.0.033712-0>.
- Borrel G, Parisot N, Harris HMB, Peyretailade E, Gaci N, Tottey W, Bardot O, Raymann K, Gribaldo S, Peyret P, O’Toole PW, Brugère J-F. 2014. Comparative genomics highlights the unique biology of *Methanomassiliicoccales*, a *Thermoplasmatales*-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* 15:679. <https://doi.org/10.1186/1471-2164-15-679>.
- Söllinger A, Schwab C, Weinmaier T, Loy A, Tveit AT, Schleper C, Ulrich T. 2016. Phylogenetic and genomic analysis of *Methanomassiliicoccales* in wetlands and animal intestinal tracts reveals clade-specific habitat preferences. *FEMS Microbiol Ecol* 92:fiv149. <https://doi.org/10.1093/femsec/fiv149>.
- Speth DR, Orphan VJ. 2018. Metabolic marker gene mining provides insight in global diversity and, coupled with targeted genome reconstruction, sheds further light on metabolic potential of the. *PeerJ* 6:e5614. <https://doi.org/10.7717/peerj.5614>.
- Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, Latreille P, Kim K, Wilson RK, Gordon JL. 2007. Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci U S A* 104:10643–10648. <https://doi.org/10.1073/pnas.0704189104>.
- Hansen EE, Lozupone CA, Rey FE, Wu M, Guruge JL, Narra A, Goodfellow J, Zaneveld JR, McDonald DT, Goodrich JA, Heath AC, Knight R, Gordon JL. 2011. Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proc Natl Acad Sci U S A* 108 Suppl 1:4599–4606. <https://doi.org/10.1073/pnas.1000071108>.
- Söllinger A, Ulrich T. 2019. Methylophilic methanogens everywhere—physiology and ecology of novel players in global methane cycling. *Biochem Soc Trans* 47:1895–1907. <https://doi.org/10.1042/BST20180565>.
- Brown JM, Hazen SL. 2018. Microbial modulation of cardiovascular disease. *Nat Rev Microbiol* 16:171–181. <https://doi.org/10.1038/nrmicro.2017.149>.
- Geng J, Yang C, Wang B, Zhang X, Hu T, Gu Y, Li J. 2018. Trimethylamine *N*-oxide promotes atherosclerosis via CD36-dependent MAPK/JNK pathway. *Biomed Pharmacother* 97:941–947. <https://doi.org/10.1016/j.biopha.2017.11.016>.
- Wang Z, Roberts AB, Buffa JA, Levison BS, Zhu W, Org E, Gu X, Huang Y, Zamanian-Daryoush M, Culley MK, DiDonato AJ, Fu X, Hazen JE, Krajcik D, DiDonato JA, Lusa AJ, Hazen SL. 2015. Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell* 163:1585–1595. <https://doi.org/10.1016/j.cell.2015.11.055>.
- Brugère J-F, Borrel G, Gaci N, Tottey W, O’Toole PW, Malpuech-Brugère C. 2014. Archaeobiotics: proposed therapeutic use of archaea to prevent

- trimethylaminuria and cardiovascular disease. *Gut Microbes* 5:5–10. <https://doi.org/10.4161/gmic.26749>.
19. Xie H, Guo R, Zhong H, Feng Q, Lan Z, Qin B, Ward KJ, Jackson MA, Xia Y, Chen X, Chen B, Xia H, Xu C, Li F, Xu X, Al-Aama JY, Yang H, Wang J, Kristiansen K, Wang J, Steves CJ, Bell JT, Li J, Spector TD, Jia H. 2016. Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst* 3:572–584.e3. <https://doi.org/10.1016/j.cels.2016.10.004>.
  20. Paul K, Nonoh JO, Mikulski L, Brune A. 2012. “*Methanoplasmatales*,” *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl Environ Microbiol* 78:8245–8253. <https://doi.org/10.1128/AEM.02193-12>.
  21. Borrel G, O’Toole PW, Harris HMB, Peyret P, Brugère J-F, Gribaldo S. 2013. Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis. *Genome Biol Evol* 5:1769–1780. <https://doi.org/10.1093/gbe/evt128>.
  22. Parks DH, Rinke C, Chuvpochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>.
  23. Borrel G, McCann A, Deane J, Neto MC, Lynch DB, Brugère J-F, O’Toole PW. 2017. Genomics and metagenomics of trimethylamine-utilizing *Archaea* in the human gut microbiome. *ISME J* 11:2059–2074. <https://doi.org/10.1038/ismej.2017.72>.
  24. Alex A, Antunes A. 2018. Genus-wide comparison of *Pseudovibrio* bacterial genomes reveal diverse adaptations to different marine invertebrate hosts. *PLoS One* 13:e0194368. <https://doi.org/10.1371/journal.pone.0194368>.
  25. Quinn TP, Erb I, Richardson MF, Crowley TM. 2018. Understanding sequencing data as compositions: an outlook and review. *Bioinformatics* 34:2870–2878. <https://doi.org/10.1093/bioinformatics/bty175>.
  26. Griffith DM, Veech JA, Marsh CJ. 2016. cooccur: probabilistic species co-occurrence analysis in R. *J Stat Softw* 69:1–17.
  27. Fennema D, Phillips IR, Shephard EA. 2016. Trimethylamine and trimethylamine *N*-oxide, a flavin-containing monooxygenase 3 (FMO3)-mediated host-microbiome metabolic axis implicated in health and disease. *Drug Metab Dispos* 44:1839–1850. <https://doi.org/10.1124/dmd.116.070615>.
  28. Begley M, Gahan CGM, Hill C. 2005. The interaction between bacteria and bile. *FEMS Microbiol Rev* 29:625–651. <https://doi.org/10.1016/j.femsre.2004.09.003>.
  29. Hovey R, Lentjes S, Ehrenreich A, Salmon K, Saba K, Gottschalk G, Gunsalus RP, Deppenmeier U. 2005. DNA microarray analysis of *Methanosarcina mazei* Gö1 reveals adaptation to different methanogenic substrates. *Mol Genet Genomics* 273:225–239. <https://doi.org/10.1007/s00438-005-1126-9>.
  30. Kamke J, Kittelmann S, Soni P, Li Y, Tavendale M, Ganesh S, Janssen PH, Shi W, Froula J, Rubin EM, Attwood GT. 2016. Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a *Sharpea*-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4:56. <https://doi.org/10.1186/s40168-016-0201-2>.
  31. LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M. 2013. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr Opin Biotechnol* 24:160–168. <https://doi.org/10.1016/j.copbio.2012.08.005>.
  32. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, Pollard KS, Sakharova E, Parks DH, Hugenholtz P, Segata N, Kyrpides NC, Finn RD. 2020. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* 49:55.
  33. Lin R, Liu W, Piao M, Zhu H. 2017. A review of the relationship between the gut microbiota and amino acid metabolism. *Amino Acids* 49:2083–2090. <https://doi.org/10.1007/s00726-017-2493-3>.
  34. Liu Y, Hou Y, Wang G, Zheng X, Hao H. 2020. Gut microbial metabolites of aromatic amino acids as signals in host-microbe interplay. *Trends Endocrinol Metab* 31:818–834. <https://doi.org/10.1016/j.tem.2020.02.012>.
  35. Cason CA, Dolan KT, Sharma G, Tao M, Kulkarni R, Helenowski IB, Doane BM, Avram MJ, McDermott MM, Chang EB, Ozaki CK, Ho KJ. 2018. Plasma microbiome-modulated indole- and phenyl-derived metabolites associate with advanced atherosclerosis and postoperative outcomes. *J Vasc Surg* 68:1552–1562.e7. <https://doi.org/10.1016/j.jvs.2017.09.029>.
  36. Ruaud A, Esquivel-Elizondo S, de la Cuesta-Zuluaga J, Waters JL, Angenent LT, Youngblut ND, Ley RE. 2020. Syntrophy via interspecies H<sub>2</sub> transfer between and underlies their global cooccurrence in the human gut. *mBio* 11:e03235-19. <https://doi.org/10.1128/mBio.03235-19>.
  37. Ng F, Kittelmann S, Patchett ML, Attwood GT, Janssen PH, Rakonjac J, Gagic D. 2016. An adhesin from hydrogen-utilizing rumen methanogen *Methanobrevibacter ruminantium* M1 binds a broad range of hydrogen-producing microorganisms. *Environ Microbiol* 18:3010–3021. <https://doi.org/10.1111/1462-2920.13155>.
  38. Podar M, Gilmour CC, Brandt CC, Soren A, Brown SD, Crable BR, Palumbo AV, Somenahally AC, Elias DA. 2015. Global prevalence and distribution of genes and microorganisms involved in mercury methylation. *Sci Adv* 1:e1500675. <https://doi.org/10.1126/sciadv.1500675>.
  39. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. 2014. Human genetics shape the gut microbiome. *Cell* 159:789–799. <https://doi.org/10.1016/j.cell.2014.09.053>.
  40. Dridi B, Henry M, Richet H, Raoult D, Drancourt M. 2012. Age-related prevalence of *Methanomassiliicoccus luminyensis* in the human gut microbiome. *APMIS* 120:773–777. <https://doi.org/10.1111/j.1600-0463.2012.02899.x>.
  41. Vanderhaeghen S, Lacroix C, Schwab C. 2015. Methanogen communities in stools of humans of different age and health status and co-occurrence with bacteria. *FEMS Microbiol Lett* 362:fnv092. <https://doi.org/10.1093/femsle/fnv092>.
  42. Zeisel SH, Mar M-H, Howe JC, Holden JM. 2003. Concentrations of choline-containing compounds and betaine in common foods. *J Nutr* 133:1302–1307. <https://doi.org/10.1093/jn/133.5.1302>.
  43. Wiedeman AM, Barr SI, Green TJ, Xu Z, Innis SM, Kitts DD. 2018. Dietary choline intake: current state of knowledge across the life cycle. *Nutrients* 10:1513. <https://doi.org/10.3390/nu10101513>.
  44. Reference deleted.
  45. Horz H-P. 2015. Archaeal lineages within the human microbiome: absent, rare or elusive? *Life (Basel)* 5:1333–1345. <https://doi.org/10.3390/life5021333>.
  46. Feldewert C, Lang K, Brune A. 2020. The hydrogen threshold of obligately methyl-reducing methanogens. *FEMS Microbiol Lett* 367:fnaa137. <https://doi.org/10.1093/femsle/fnaa137>.
  47. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, Spector TD, Bell JT, Clark AG, Ley RE. 2016. Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* 19:731–743. <https://doi.org/10.1016/j.chom.2016.04.017>.
  48. Goodrich JK, Davenport ER, Clark AG, Ley RE. 2017. The relationship between the human genome and microbiome comes into view. *Annu Rev Genet* 51:413–433. <https://doi.org/10.1146/annurev-genet-110711-155532>.
  49. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, Le Roy CI, Raygoza Garay JA, Finnicum CT, Liu X, Zhernakova DV, Bonder MJ, Hansen TH, Frost F, Rühlemann MC, Turpin W, Moon J-Y, Kim H-N, Lüll K, Barkan E, Shah SA, Fornage M, Szopinska-Tokov J, Wallen ZD, Borisevich D, Agreus L, Andreasson A, Bang C, Bedrani L, Bell JT, Bisgaard H, Boehnke M, Boomsma DI, Burk RD, Claringbould A, Croitoru K, Davies GE, van Duijn CM, Duijts L, Falony G, Fu J, van der Graaf A, Hansen T, Homuth G, Hughes DA, Ijzerman RG, Jackson MA, Jaddoe VVW, Joossens M, Jørgensen T, Keszthelyi D, Knight R, Laakso M, Laudes M, et al. 2020. Genetics of human gut microbiome composition. *bioRxiv* <https://doi.org/10.1101/2020.06.26.173724>.
  50. Roehe R, Dewhurst RJ, Duthie C-A, Rooke JA, McKain N, Ross DW, Hyslop JJ, Waterhouse A, Freeman TC, Watson M, Wallace RJ. 2016. Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet* 12:e1005846. <https://doi.org/10.1371/journal.pgen.1005846>.
  51. Difford GF, Plichta DR, Løvendahl P, Lassen J, Noel SJ, Højberg O, Wright A-DG, Zhu Z, Kristensen L, Nielsen HB, Guldbandsen B, Sahana G. 2018. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genet* 14:e1007580. <https://doi.org/10.1371/journal.pgen.1007580>.
  52. Hania JB, Ballet N, Vandekerckove P, Olivier B, O’Toole PW, Brugère J-F. 2017. Archaeobiotics: archaea as probiotics for treating chronic disease in humans? p 42–62. *In* Sghaier H, Najjari A, Ghedira K (ed), *Archaea: new biocatalysts, novel pharmaceuticals and various biotechnological applications*. InTech Open, London, UK.
  53. Douglas AE. 2020. The microbial exometabolome: ecological resource and architect of microbial communities. *Philos Trans R Soc Lond B Biol Sci* 375:20190250. <https://doi.org/10.1098/rstb.2019.0250>.
  54. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
  55. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 4:2304. <https://doi.org/10.1038/ncomms3304>.
  56. Letunic I, Bork P. 2016. Interactive Tree Of Life (iTOL) v3: an online tool for

- the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245. <https://doi.org/10.1093/nar/gkw290>.
57. Mitchell AL, Scheremetjew M, Denise H, Potter S, Tarkowska A, Qureshi M, Salazar GA, Pesseat S, Boland MA, Hunter FMI, Ten Hoopen P, Alako B, Amid C, Wilkinson DJ, Curtis TP, Cochrane G, Finn RD. 2018. EBI metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Res* 46:D726–D735. <https://doi.org/10.1093/nar/gkx967>.
  58. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 11:2864–2868. <https://doi.org/10.1038/ismej.2017.126>.
  59. Breitwieser FP, Baker DN, Salzberg SL. 2018. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. *Genome Biol* 19:198. <https://doi.org/10.1186/s13059-018-1568-0>.
  60. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
  61. Ding W, Baumdicker F, Neher RA. 2018. panX: pan-genome analysis and exploration. *Nucleic Acids Res* 46:e5. <https://doi.org/10.1093/nar/gkx977>.
  62. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell AL. 2017. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res* 45: D190–D199. <https://doi.org/10.1093/nar/gkx1107>.
  63. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 34:2115–2122. <https://doi.org/10.1093/molbev/msx148>.
  64. Keck F, Rimet F, Bouchez A, Franc A. 2016. phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol Evol* 6:2774–2780. <https://doi.org/10.1002/ece3.2051>.
  65. Snipen L, Liland KH. 2015. micropan: an R-package for microbial pan-genomics. *BMC Bioinformatics* 16:79. <https://doi.org/10.1186/s12859-015-0517-0>.
  66. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 3:217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>.
  67. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>.
  68. Lu J, Breitwieser FP, Thielen P, Salzberg SL. 2017. Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science* 3:e104. <https://doi.org/10.7717/peerj-cs.104>.
  69. de la Cuesta-Zuluaga J, Ley RE, Youngblut ND. 2020. Struo: a pipeline for building custom databases for common metagenome profilers. *Bioinformatics* 36:2314–2315. <https://doi.org/10.1093/bioinformatics/btz899>.
  70. Quinn TP, Richardson MF, Lovell D, Crowley TM. 2017. propr: an R-package for identifying proportionally abundant features using compositional data analysis. *Sci Rep* 7:16252. <https://doi.org/10.1038/s41598-017-16520-0>.
  71. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48.
  72. Rath S, Heidrich B, Pieper DH, Vital M. 2017. Uncovering the trimethylamine-producing bacteria of the human gut microbiota. *Microbiome* 5:54. <https://doi.org/10.1186/s40168-017-0271-9>.
  73. Rath S, Rud T, Pieper DH, Vital M. 2019. Potential TMA-producing bacteria are ubiquitously found in Mammalia. *Front Microbiol* 10:2966. <https://doi.org/10.3389/fmicb.2019.02966>.
  74. Dorokhov YL, Shindyapina AV, Sheshukova EV, Komarova TV. 2015. Metabolic methanol: molecular pathways and physiological roles. *Physiol Rev* 95:603–644. <https://doi.org/10.1152/physrev.00034.2014>.
  75. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohney RP, Li W, de Rinaldis E, Bell JT, Venter JC, Nelson KE, Spector TD, Falchi M. 2019. Interplay between the human gut microbiome and host metabolism. *Nat Commun* 10:4505. <https://doi.org/10.1038/s41467-019-12476-z>.