

# A comprehensive analysis of the global human gut archaeome from a thousand genome catalogue

Cynthia Maria Chibani<sup>1,8</sup>, Alexander Mahnert<sup>2,8</sup>, Guillaume Borrel<sup>3</sup>, Alexandre Almeida<sup>4,5</sup>, Almut Werner<sup>1</sup>, Jean-François Brugère<sup>6</sup>, Simonetta Gribaldo<sup>3</sup>, Robert D. Finn<sup>4</sup>, Ruth A. Schmitz<sup>1,9</sup>, Christine Moissl-Eichinger<sup>2,7,9</sup>

<sup>1</sup> Institute for Microbiology, Christian-Albrechts-University Kiel, Kiel, Germany

<sup>2</sup> Diagnostic & Research Institute of Hygiene, Microbiology and Environmental Medicine, Medical University Graz, Neue Stiftingtalstraße 6, 8010 Graz, Austria

<sup>3</sup> Department of Microbiology, Unit Evolutionary Biology of the Microbial Cell, Institut Pasteur, Paris, France

<sup>4</sup> European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom.

<sup>5</sup> Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom.

<sup>6</sup> Institut Universitaire de Technologie Clermont Auvergne, Université Clermont Auvergne, CNRS, UMR 6023 Laboratoire Microorganismes: Genome et Environnement, Clermont-Ferrand, France

<sup>7</sup> BioTechMed, Graz, Austria

<sup>8</sup> These authors contributed equally

<sup>9</sup> Lead contact

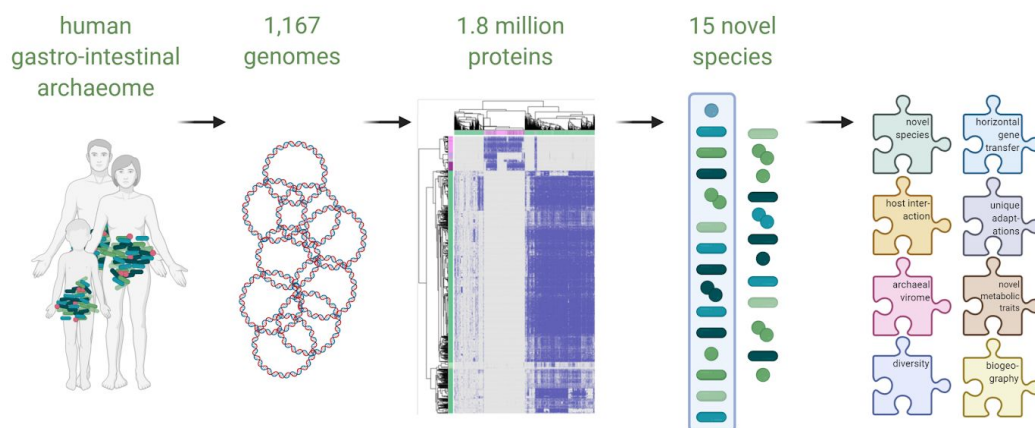
\* Correspondence: [christine.moissl-eichinger@medunigraz.at](mailto:christine.moissl-eichinger@medunigraz.at), [rschmitz@ifam.uni-kiel.de](mailto:rschmitz@ifam.uni-kiel.de)

## SUMMARY

The human gut microbiome plays an important role in health and disease, but the archaeal diversity therein remains largely unexplored. Here we report the pioneering analysis of 1,167 non-redundant archaeal genomes recovered from human gastrointestinal tract microbiomes across countries and populations. We identified three novel genera and 15 novel species including 52 previously unknown archaeal strains. Based on distinct genomic features, we warrant the split of the *Methanobrevibacter smithii* clade into two separate species, with one represented by the novel Candidatus *M. intestini*.

Patterns derived from 1.8 million proteins and 28,851 protein clusters coded in these genomes showed substantial correlation with socio-demographic characteristics such as age and lifestyle. We infer that archaea are actively replicating in the human gastrointestinal tract and are characterized by specific genomic and functional adaptations to the host. We further demonstrate that the human gut archaeome carries a complex virome, with some viral species showing unexpected host flexibility. Our work furthers our current understanding of the human archaeome, and provides a large genome catalogue for future analyses to decipher its role and impact on human physiology.

## GRAPHICAL ABSTRACT



## HIGHLIGHTS

- The human gut archaeome analysis reveals a previously unseen active diversity
- The most abundant methanogen, *Methanobrevibacter smithii*, splits into two species
- Archaeal protein catalogue can predict geography, demographics and health aspects
- Host-associated and environmental archaea show distinct genomic & functional traits

## INTRODUCTION

The human microbiome is increasingly recognized as a key player in human health (Duvall et al. 2017). While most research has focused on the bacterial and bacteriophages component, and to some extent unicellular eukaryotes (including fungi) and their viruses, the Archaea have been largely overlooked, mainly due to methodological reasons (Moissl-Eichinger et al. 2018; Mahnert et al. 2018; Bang and Schmitz 2018; Pausan et al. 2019; Borrel et al. 2020).

Recent analyses of metagenome-assembled genomes (MAGs) have revealed a huge unexplored human microbiome diversity, increasing the number of genome-represented bacterial species to over 2,000 (Nayfach et al. 2019; Pasolli et al. 2019; Almeida et al. 2019; 2020). Although archaea were reported in these datasets (Pasolli et al. 2019; Almeida et al. 2020), their diversity and functional potential remains largely unexplored.

Archaea are prokaryotes with distinct biology compared to bacteria and have been known to be present in the human gut for several decades (Borrel et al. 2020). Due to their regulatory function at the end of the microbial food chain, they are considered metabolic keystone species in the gastrointestinal microbial network. Some archaea also carry specific adaptive traits for survival and colonization of the human gut environment, such as bile salt hydrolases (Gaci et al. 2014), or adhesin-like proteins (Hansen et al. 2011; Borrel et al. 2017). Archaea can remove deleterious bacterial metabolites such as TMA (Brugère et al. 2014; Borrel et al. 2017) and can induce host immune responses (Bang et al. 2014; Vierbuchen et al. 2017). However, their role in human disease remains unclear (Mahnert et al. 2018; Borrel et al. 2020).

Here, we present a comprehensive analysis of a set of 1,167 archaeal genomes derived from the human gastrointestinal tract. We identified a set of genomes representing known and unknown archaeal strains ( $n=98$ ) and species ( $n=27$ ) from the human gastrointestinal tract. These genomes span methanogenic and halophilic archaea with a remarkably high number of as-yet undescribed representatives of the human microbiome (52 representing novel strains, and 15 novel species). In total, we identified 1.8 million coded proteins, of which ~50% remain un-annotated (0.98 million hypothetical proteins). This protein catalogue correlated with socio-demographic characteristics of the human carriers, such as age and lifestyle. The extensive resource also revealed novel taxonomic insights, such as the obvious split of *Methanobrevibacter smithii* into two separate species. Moreover, we could show that the archaeal microbiome component is actively replicating within the human gastrointestinal tract and is characterized by a specific genomic and functional adaptation towards the human host. The human gut archaeome carries a complex virome, which revealed an unexpected host flexibility for some viruses. Despite the newly available information, the archaeome still remains undersampled, indicating an even larger archaeal diversity to be expected.

This genome set will now serve as a foundation for future comparative genomics, species-directed cultivation attempts, and for further elucidating the role of the human archaeome within the microbiome and in human health and disease.

## RESULTS

### Over 1,000 unique archaeal genomes recovered from human gastrointestinal samples

To explore the diversity of archaea in human gastrointestinal samples, we compiled publicly available genomes from four recent metagenome-assembled genomes (MAGs) collections (Nayfach et al. 2019; Pasolli et al. 2019; Almeida et al. 2019; 2020) together with genomes from cultured archaea available in the NCBI (Kitts et al. 2016), PATRIC (Wattam et al. 2017) and IMG (Chen et al. 2019) repositories. To complete the dataset, we included genomes of *Candidatus* Methanomethylophilus alvus (Borrel et al. 2012), *Candidatus* Methanomassiliicoccus intestinalis (Borrel et al. 2013), of the human isolate *Methanobrevibacter arboriphilus* ANOR1 (Khelaifia et al. 2014b), and of Methanomassiliicoccales Mx02, Mx03, Mx06, and *Candidatus* M. intestinalis (Borrel et al. 2017). Henceforth, unless otherwise stated, “genomes” refers to genome sequences from both isolates and MAGs. For a detailed overview on the methodology please refer to Supplementary Fig 1.

To ensure high quality of the dataset, a threshold of >50% genome completeness and <5% contamination was used, following previously published methodology (Almeida et al. 2020). This procedure yielded 1,167 non-redundant archaeal genomes (Mash distance threshold of 0.001, 99.9% nucleotide identity; Ondov et al. 2016; Supplementary Table 1) which were further sub-grouped into a “strain”-list (<99% ANI similarity; Supplementary Table 1; 98 genomes), and a “species”-list (<95% ANI similarity; Supplementary Table 1; 27 genomes) (Fig. 1A). For this, the best quality genome (genome completeness, minimal contamination, strain heterogeneity and assembly continuity based on the N50 value) from each cluster was selected as representative, or whenever an isolate was available it was preferred and used for further analysis. More details on genome length, number of contigs, N50, G/C content, number of tRNAs and completeness are provided in Fig. 1, B-C.

The 1,167 genomes span a wide taxonomic diversity, and include members of the Methanobacteriales (87.15%), Methanomassiliicoccales (12.43%), Methanomicrobiales (0.26%) and Halobacteriales (0.17%; Fig. 1; Fig. 2, B-C). Most genomes were taxonomically affiliated with the known genus *Methanobrevibacter* (996 genomes; 85%), in agreement with earlier reports (Borrel et al. 2020). Other genomes were affiliated to *Methanomethylophilus* (38; 3.3%), *Methanomassiliicoccus* (29; 2.5%), *Methanosphaera* (20; 1.7%), and *Methanocorpusculum* (3; 0.3%). *Methanobacterium*, *Haloferax* and *Halorubrum* were only represented by one genome each. Of the 1,167 genomes, 78 (6.7%) could not be assigned to any previously described genus and 101 genomes (8.7%) did not match any known species. Those genomes are mostly Methanomassiliicoccales representatives.

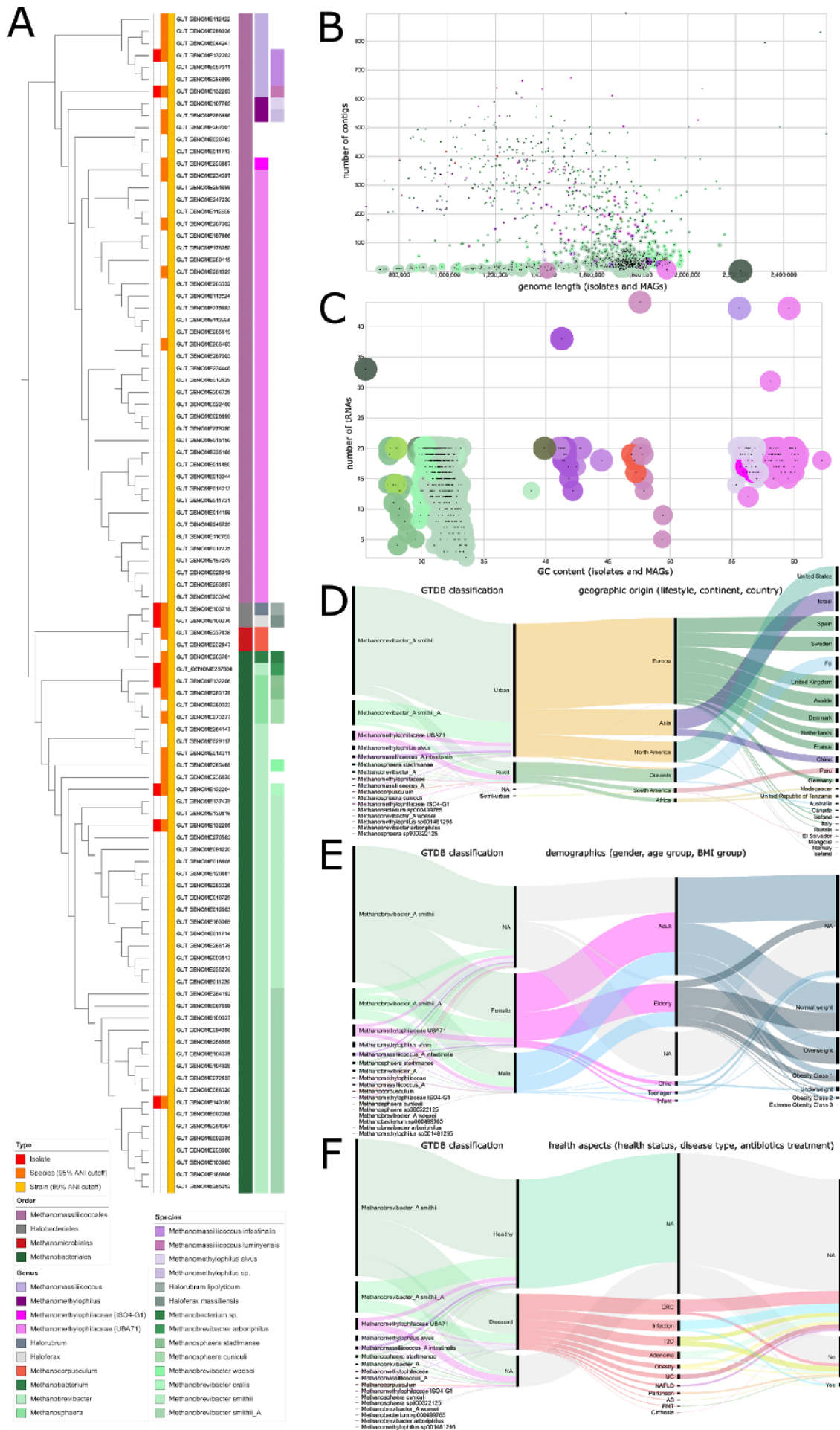
---

**Fig. 1: 1,167 archaeal genomes from the human gastrointestinal tract reveal taxonomic expansion of the archaeome and correlation with cohort metadata.** Note: *Methanobrevibacter smithii*\_A is later referred to as *Cand. M. intestinalis*.

**A)** ANI-based tree of genomes clustered at 99% similarity (“strains”). Isolates, species and strains are highlighted next to the tree topology, taxonomic affiliation is displayed next to the genome names.

**B,C)** Genome characteristics and numerical metadata (genome length, number of contigs, N50, G/C content, number of tRNAs and completeness) are displayed as scatterplots (colored radius refers to N50 in B and completeness [%] in C).

**D-F)** Categorical metadata is grouped in three alluvial diagrams referring to geographic origin (lifestyle, continent, country), demographics (gender, age group, BMI group) and health aspects (health status, disease type, antibiotics treatment) (NA = no data available).



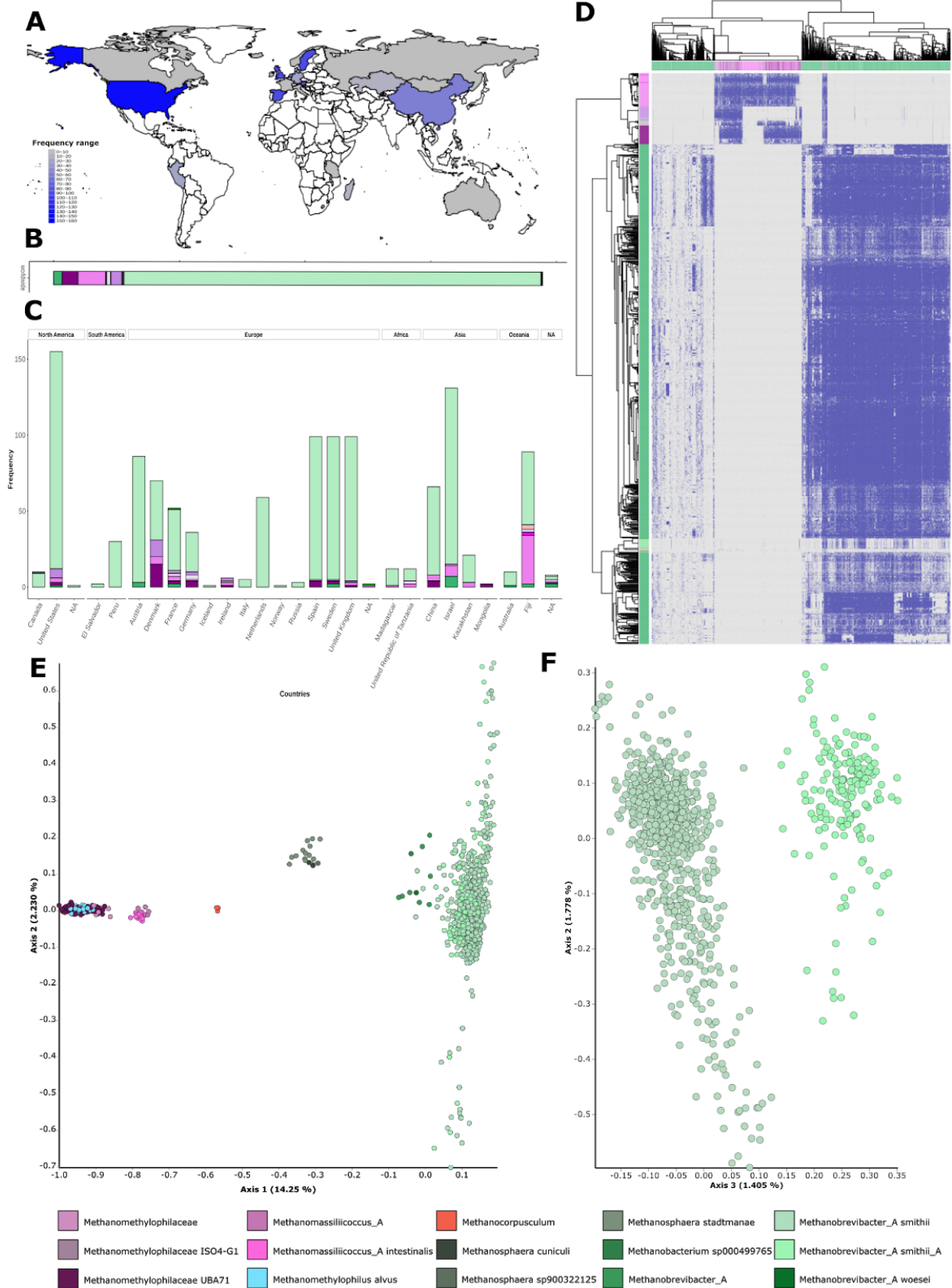
## Archaeal protein profile correlates with geographic and demographic parameters

The entire data set (Supplementary Table 1) covers samples from rural and urban populations of 24 countries (five continents; Fig. 2, A-C) and includes more than 800 genomes from infants, children, teenagers, adults and elderly subjects spanning all BMI (Body Mass Index) groups (underweight up to extreme obesity class 3; details on available metadata are given in Star\*Methods). Three-hundred subjects showed a diseased phenotype (e.g. colorectal cancer, type 2 diabetes, infection), and received antibiotic treatment. However, most genomes (3rd quartile, 75% of all values) were obtained from healthy female adults of normal weight, living in urban areas of Europe (Fig. 1, D-F).

In total, 1.8 million proteins were identified, 54% of which lacked any functional annotation. A protein catalogue of all 1,167 archaeal genomes was generated by clustering the genes predicted across all genomes (>50% amino acid identity and >80% coverage) (Supplementary Table 2). 3,050 proteins (thereof 58% hypothetical proteins) were found to be shared among more than 50 genomes in our dataset, mirroring the taxonomic distance of the two most abundant orders, Methanomassiliicoccales and Methanobacteriales (Fig. 2D). To overcome biases introduced by potential residual MAGs contamination issues, we focused our analyses on patterns observed in two or more genomes, unless stated otherwise. Additionally, we explored protein diversity patterns and their functional characterization among isolated genomes to corroborate those observed in MAGs (Supplementary Fig. 4).

Using the protein catalogue information, the predictive potential of these metadata categories was investigated. We found a high predictive potential of the unified archaeal protein matrix composition according to genome characteristics (completeness:  $R = 0.97$ ,  $p = 1.2 \cdot 10^{-131}$ , GC content:  $R = 0.99$ ,  $p = 7.3 \cdot 10^{-224}$ ), taxonomic classification (baseline accuracy 100%) (Fig. 2, E-F), or certain demographics such as subject age ( $R = 0.7$ ,  $p = 2.8 \cdot 10^{-19}$ , age groups baseline accuracy 73.2%), lifestyle (baseline accuracy 93.3%) or treatment with antibiotics (baseline accuracy 95.7%). To a lesser extent, other metadata categories such as the subjects health status (baseline accuracy 81.0%), their sex (baseline accuracy 73.2%), the BMI ( $R = 0.4$ ,  $p = 2.9 \cdot 10^{-5}$ ), the geographic origin (continent: baseline accuracy 78.7%, country: baseline accuracy 62.3%) or individual disease types (baseline accuracy 67.2%) were useful to describe the composition of the unified archaeal protein matrix (Supplementary Fig. 2). The created protein catalogue can now serve as basis for future functional characterizations of the archaeome, and to identify proteins potentially involved in human health and disease.







**Fig. 2: Archaeal genomes from the human gut microbiome distribution and the corresponding unified protein catalogue.** Note: *Methanobrevibacter smithii*\_A is later referred to as *Cand. M. intestini*.

**A)** Worldwide distribution of the archaeal genomes, colored according to abundance (white indicates data not available). Barplot representing **B)** the overall proportion of taxa and **C)** the number per country, of the archaeal genomes isolated from the human gut.

**D)** Unified human archaeal protein catalogue based on the inter-species pan-genome of all 1,167 archaeal genomes. Heatmap depicts the presence of 3,050 proteins (found in more than 50 genomes, columns) across the 1,167 archaeal genomes (rows).

**E)** The taxonomic distinction of Methanomassiliicoccales, Halobacteriales and Methanobacteriales based on the protein profile (Fig. 2D), displayed in a PCoA plot based on Bray-Curtis distances at a depth of 623 archaeal proteins. The PCoA showed five distinct clusters referring to Methanomethylophilaceae, *Methanomassiliicoccus*, *Methanocorpusculum*, *Methanosphaera* and *Methanobacteriaceae*.

Notably, the clade of *Methanobacteriaceae* was subdivided into *Methanobacterium* and a heterogeneous cluster of *Methanobrevibacter* (**F**).

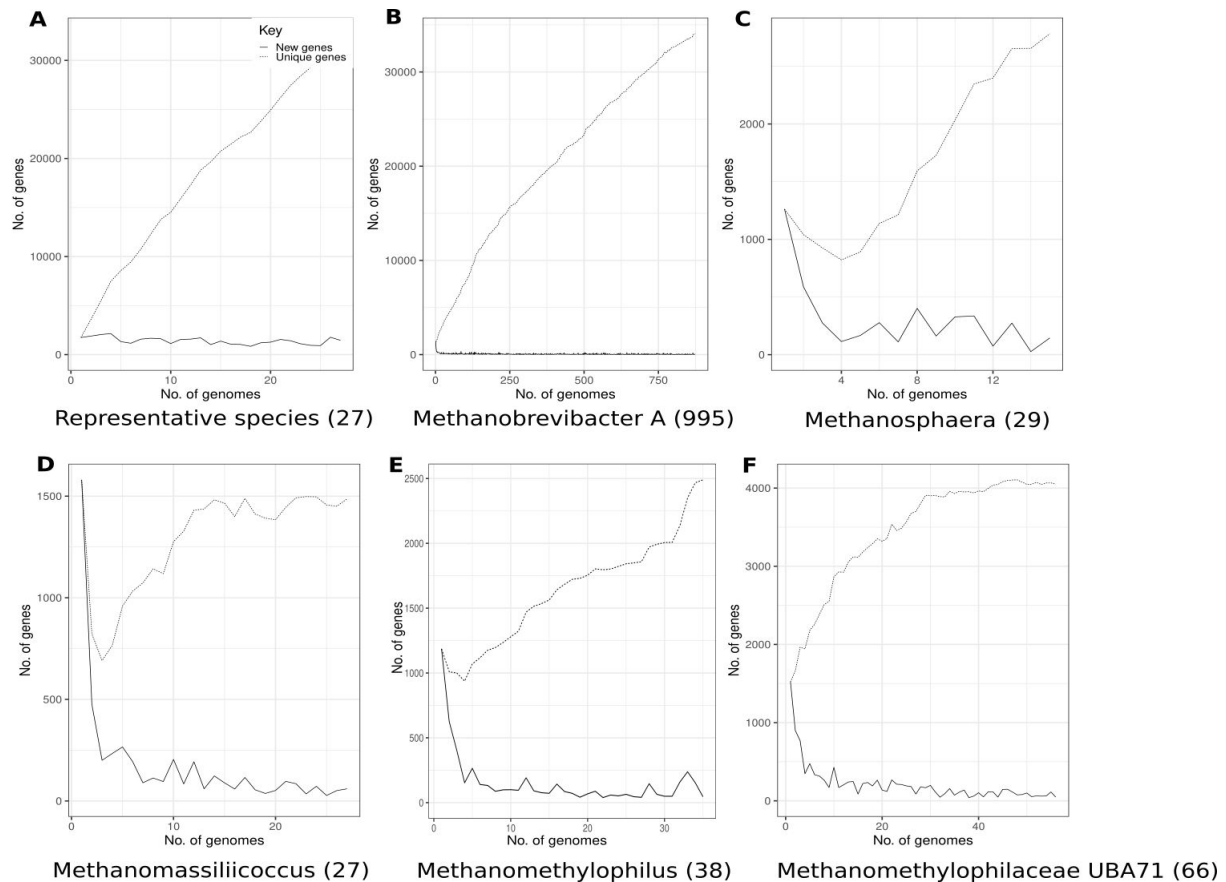
---

### The genetic diversity of the gut archaeome remains undersampled

To understand the level of sampling completeness of the human archaeome, we computed the genetic diversity across the 27 representative archaeal genome species, and within those genera that had over 10 genome representatives (details given in Supplementary Table 1; Fig. 3). We excluded singletons to avoid observations from potential contaminants of a MAG. We found that more genomes are needed to have a saturated overview of the genetic diversity of the genera *Methanobrevibacter* (995 genomes,  $\alpha=0.63$ ), and of the Methanomethylophilaceae UBA71 clade (66 genomes,  $\alpha=0.95$ ) (indicated by an  $\alpha < 1$  which means each added genome contributed new genes). However, additional species of the genera *Methanosphaera* (29 genomes,  $\alpha=2$ ), *Methanomassiliicoccus* (27 genomes,  $\alpha=1.07$ ), and *Methanomethylophilus* (38 genomes,  $\alpha=1.3$ ) would not contribute new genes (for more details please refer to Star\*Methods) (Tettelin et al. 2008). Observations were further confirmed by analysis at the family level excluding singletons (Supplementary Fig. 4).

Nonetheless, excluding singletons meant excluding pangenes of species which are less abundant per genus. Therefore, in analysis that included singletons, a different picture was obtained. Under these conditions, we found that more species are needed to have a saturated overview of the genetic diversity of the genera *Methanosphaera* ( $\alpha=0.64$ ), of *Methanomassiliicoccus* ( $\alpha=0.61$ ), of *Methanomethylophilus* ( $\alpha=0.49$ ).

Due to the nature of MAGs, it is important to isolate more genomes of the species that are found scarcely in the dataset in order to assess the true genetic diversity of a genus in the GIT.



**Fig. 3: Pan-core-genome analysis of the representative MAGs species and for the genera that harbored over 10 genomes.** The number of gene clusters for unique genes (dashed line) and new genes (solid line) results from the successive addition of genomes. The curve shows mean values of multiple iterations (> 100) where genomes were randomly added and the pan- and core-genome was computed sequentially after each addition. **A)** 27 representative MAGs species, **B)** *Methanobrevibacter\_A*, **C)** *Methanosphaera*, **D)** *Methanomassiliicoccus*, **E)** *Methanomethylophilus*, and **F)** *Methanomethylophilaceae* UBA71.

### The dataset reveals novel members of the human gastrointestinal archaeome, particularly in Asian and Oceanian populations

We obtained 20 genomes affiliated with *Methanosphaera*, including three genomes from isolates (Miller and Wolin 1985; Hansen et al. 2011). Taxonomically, human-associated *Methanosphaera* genomes were affiliated to three distinct species-level clades (Supplementary Fig. 5). Among those, *M. stadtmanae* was the most commonly retrieved, with 17 genomes (14 MAGs). Two MAGs (ANI 98.5%) clustered within *M. cuniculi*, and were retrieved from Asian healthy subjects living in an urban environment. The *M. cuniculi* type strain was originally isolated from the intestinal tract of a rabbit (Biavati, Vasta, and Ferry 1988) and has not been reported thus far in human hosts. One additional MAG belonging to the genus of *Methanosphaera* (Fig. 1), was binned from a gut metagenome of a diseased (colorectal cancer) European male (BMI 21, age: 64, urban environment). This genome clustered together with RUG761, a genome recovered from cattle intestines (Stewart et al. 2018); ANI 99.0%; Supplementary Fig. 5).

The dataset of human-associated Methanomassiliicoccales consisted of 145 genomes corresponding to 12 candidate species and one isolate (*M. luminyensis*) (Fig. 1, 2). The genomes were distributed into two families, the majority of them belonging to “host-associated” Methanomethylophilaceae (116 genomes), the other to Methanomassiliococcaceae (“free-living clade”; 29 genomes). Five of the candidate species corresponded to genomes that were previously found in human samples,

comprising 81% of the Methanomassiliicoccales from the current study. These included Methanomassiliicoccales Mx-06 sp. (Borrel et al. 2017; 44 genomes), *Methanomethylophilus alvus* (Borrel et al. 2012; 37 genomes) and *Methanomassiliicoccus intestinalis* (Borrel et al. 2013; 20 genomes), being the most prevalent Methanomassiliicoccales representative in human populations (Borrel et al. 2017). Mx-06 representatives were mostly present in young adults (32 years old, n = 34) from rural areas (80%; n = 40) in Oceania, Asia and Africa (65%, 13% and 7% respectively, n = 43). Together with its high prevalence (80%) in a population of 7-48 years old uncontacted Amerindians (Clemente et al. 2015; Borrel et al. 2017), it appears that this species is strongly linked with non-westernized populations. The young age of people having this species contrasts with previously reported positive correlation between age and methanogen prevalence. We found that several representatives of this species have the genetic potential to metabolize trimethylamine (TMA), a bacterial metabolite involved in trimethylaminuria and suspected in cardiometabolic, cardiovascular and renal diseases; they also have functional traits associated with adaptation to the human gut such as bile-salt hydrolase. This species is part of a well-supported clade that is separated from other Methanomethylophilaceae genus (*Methanomethylophilus*, *Methanogranum* and *Methanoplasma*) and belongs to the candidate genus “UBA71” following GTDB classification (Supplementary Table 1). We thus suggest that it represents a new genus and propose the name of *Candidatus Methanoprimatia hominis* for representatives of Mx-06 (representative MAG: GUT\_GENOME268463). In addition to the species previously identified through MAGs or culture approaches, we identified six novel Methanomassiliicoccales species, represented by 24 MAGs. One of those gathers 12 MAGs and was more often found among Asian people. We propose to name it “*Ca. Methanoprimatia macfarlanii*” (representative MAG: GUT\_GENOME 251929).

In addition, a number of archaeal taxa not yet described to be constituents of the human gastrointestinal tract were recovered from the MAG dataset.

This included one MAG affiliated to *Halorubrum lipolyticum* (Halobacteriaceae, GUT\_GENOME103718, Fig. 1), which showed 100% ANI similarity to the type strain originally isolated from a Chinese lake (Cui et al. 2006). Together with isolate *Haloferax massiliensis*, which was retrieved from human feces in 2018 (Khelaifia et al. 2018), they represent the only two genomes affiliated to halophilic archaea available from the human GIT. This observation is in contrast to the previously reported high prevalence of Haloarchaea of Asian cohorts (Kim et al. 2020). None of the two genomes revealed obvious adaptations towards the human gut ecosystem such as bile salt hydrolases and may therefore be transient residents.

Two genomes affiliated to Methanocorpusculaceae (Methanomicrobiales) (Fig. 1; ANI 98.6%), representing two strains and one species, were retrieved from subjects from the Fiji islands (Fig. 2). Closest relatives were *M. parvum* (Strain XII, type strain; ANI 70.32%) and *M. bavaricum* (DSM4179; 70.25%), which were originally isolated from digester and wastewater environments, respectively.

Several novel genomes were also detected within the *Methanobrevibacter* clade, including the cluster containing GUT\_GENOME014311 (Fig. 1, Fig. 4A). This cluster was represented by seven genomes (95% ANI cut-off), clustering into two distinctive strains and one species. All seven genomes were originally binned from samples obtained from Asia (China) and Oceania (Fiji) only; genomes were obtained from healthy and diseased, rural and urban subjects. The closest relatives identified were *Methanobrevibacter woesei* GCA 003111605T (71.54% ANI) and *Methanobrevibacter smithii* ATCC 35061 NC 009515T. Additional *Methanobrevibacter* genomes were recovered with larger distance to known isolates, such as GUT\_GENOME236870 and GUT\_GENOME237437 (99.67% ANI; both isolated from a Fiji cohort; Fig. 1), with a genome similarity of 84.13% to *M. gottschalkii* DSM11977 (ANI), or GUT\_GENOME237054 (also from a Fiji cohort), showing a similarity of 78.11% to *M. gottschalkii* (ANI). An overview on host association, geography, genome size and taxonomic association of known *Methanobrevibacter* species and genomes is given in Fig. 4, B-G). It shall be mentioned that no MAG corresponding to *Methanobrevibacter oralis* was identified in our dataset,

although this species from the human oral microbiome has been at least recovered once from human stool (Khelaifia et al. 2014a).

Signatures of the genus *Methanobacterium* have been frequently detected in amplicon-based archaeome studies using small intestine biopsies or samples from oral cavities, but have neither been isolated nor detected by genome-centric analyses (Faveri et al. 2011; Matarazzo et al. 2011; Koskinen et al. 2017). The presence of *Methanobacterium* species in the human GIT have been confirmed in our studies, as one *Methanobacterium* MAG (GUT\_GENOME283701) was obtained from a European male (age 65; colorectal cancer), with a genome size of 1.9 Mbp (98.13% completeness; 0% contamination). Based on ANI values, the closest described relative was *Methanobacterium formicum* DSM3637 (87.75%), a well-known, formate consuming methanogen in ruminants (Chellapandi et al. 2018). Formate consuming capacity was as well found for the *Methanobacterium* MAG identified in our study.

Undersampled populations obviously bear an unknown archaeal wealth which requires future dedicated studies.

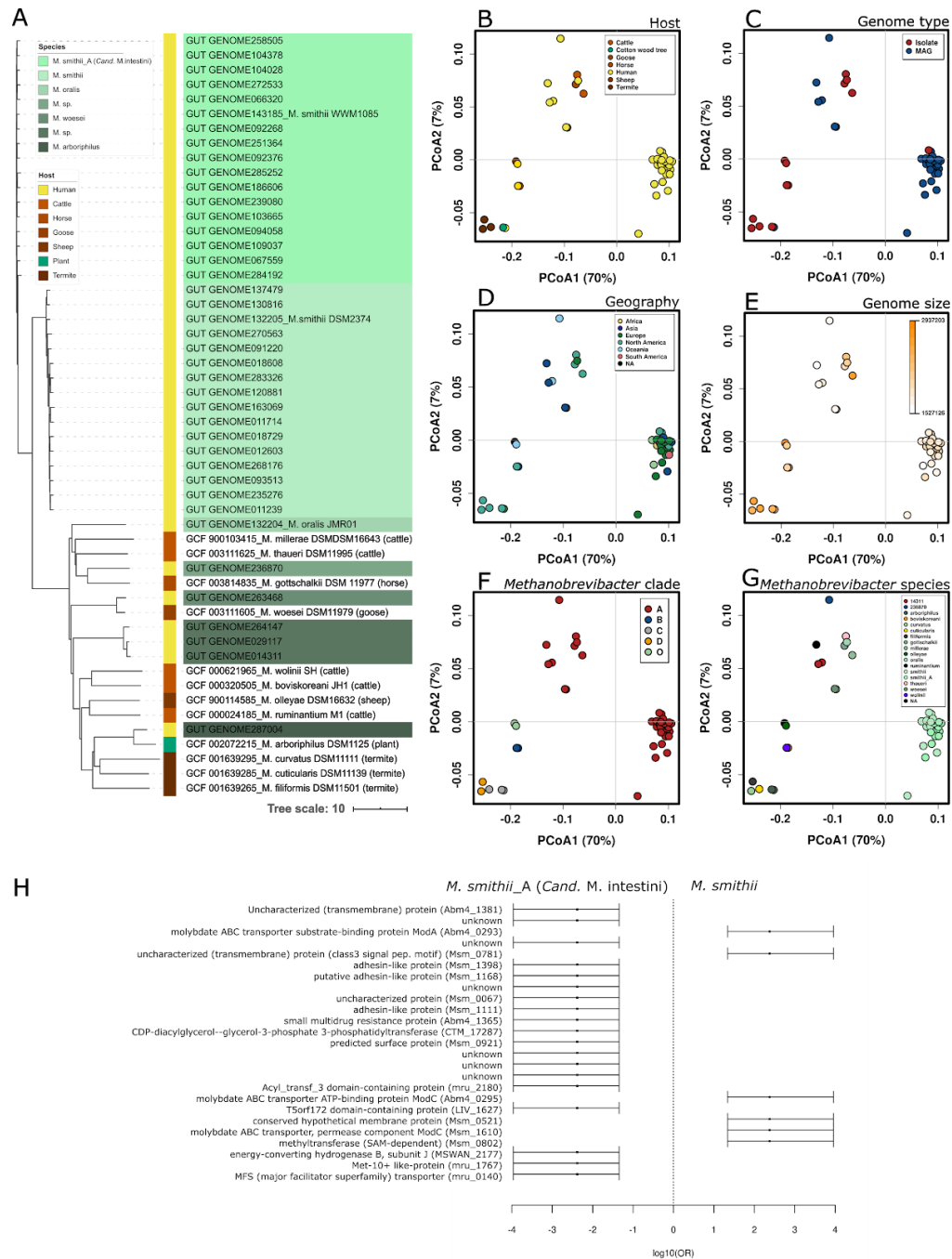
### The *Methanobrevibacter smithii* clade splits into two separate species

The *Methanobrevibacter smithii* clade was the largest taxonomic clade detected in our study (Fig. 2). Based on ANI similarity values, as well as information derived from the protein catalogue, the *Methanobrevibacter smithii* group was represented by two species-level clades (tentatively named “smithii” and “smithii\_A” according to GTDB classification; Fig. 2F, Fig. 4A; Supplementary Table 1 also Pasolli et al. 2019). *M. smithii\_A* was represented in our entire dataset 185 times (16% of entire dataset), whereas *M. smithii* was detected 797 times (68%), together representing 84% of all genomes in our dataset (Fig. 2, A-C). Notably, the two *M. smithii* groups differed significantly with respect to their protein profile (PERMANOVA,  $p=0.001$ ; Fig. 2F) and genome size ( $p<0.001$ , Wilcoxon rank sum test, two-sided, genome size corrected by completeness), with median genome sizes of 1.7 Mbp for *M. smithii* and 1.8 Mbp for *M. smithii\_A* (Supplementary Table 4; genome sizes for isolates: 1.7 Mbp (reference genome *M. smithii*) and 1.9 Mbp (isolate WWM1085, see below), respectively).

All *M. smithii* strains carried the *modA* gene, which was not detected in any of the “smithii\_A” genomes (Supplementary Table 4). This gene is involved in molybdate transport, and responsible for substrate binding (Self et al. 2001). In addition, amongst the top 25 discriminative proteins (Fig. 4H), the molybdate ABC transporter permease component, as well as the molybdate ABC transporter ATP-binding protein was identified in 94% of all *M. smithii* genomes, but in none of the *M. smithii\_A* genomes. This indicates a different pathway for molybdate acquisition in the *M. smithii\_A* clade. The *M. smithii\_A* genomes were further characterized by additional unique membrane/cell-wall associated proteins, such as adhesin-like proteins, surface proteins, and a number of uncharacterized membrane proteins (Fig. 4H).

Based on the extent of discriminative features, and an average nucleotide identity of 93.95% between the two representative genomes of *M. smithii* and *M. smithii\_A*, we propose to rename the smithii\_A clade, represented by isolate WWM1085 (GUT\_GENOME143185, Jennings et al. 2017), into *Candidatus Methanobrevibacter intestini* to further emphasize the presence of two predominant, distinctive *Methanobrevibacter* clades in the human intestinal tract. *Cand. M. intestini* and *M. smithii* cannot be distinguished on 16S rRNA gene sequence, which might be the reason for missing this separation in two distinctive clades before. However, analysis of the *mcrA* gene revealed a consistent difference between the two clades, with an average of 2.15 % difference in amino acid sequence (1.82-2.22%; Supplementary Material 1).

In future studies it will be important to further investigate the distribution and the potential interplay of *M. smithii* and *Cand. M. intestini*, in order to understand their biological role(s) in the human body.



**Fig. 4: Characteristics of the *Methanobrevibacter* genomes.** **A)** Phylogenetic tree of the *Methanobrevibacter* clade based on ANI distance. Twelve representative genomes from other sources than humans were included for comparison (Supplementary Table 3). Genomes (strain-level) from the human gastrointestinal tract are highlighted in green colors (taxon label). The bar on the left displays the origin: human (yellow bar), animals (shades of red) and plant (green). **B-G)** PcoA plots of protein profiles according to host (**B**), genome source type (**C**), geography (**D**), genome size (**E**), *Methanobrevibacter* clade according to GTDB (**F**) and species (**G**). **H)** Forest plot showing the outcome of the Wilcoxon rank test comparison of genomes from *M. smithii\_A (Cand. M. 16S rRNAintestini)* vs. *M. smithii* (only TOP 25 proteins are shown; FDR<0.000005); bar displays the odds ratio (OR) (Supplementary Table 4).

## The human archaeome is actively replicating

In order to address the question whether the human gastrointestinal archaea are actively replicating in their host environment, we performed iRep analysis (see Star\*Methods). Based on iRep analysis of 26 genomes with fully available metagenomic data resources, 10-50% of the archaeal population was found to be actively replicating at the time point of sampling with a mean replication rate of  $1.4 \pm 0.1$ . Replication rates differed between *M. smithii*, *M. smithii\_A* (*Cand. M. intestini*), *Methanomassiliicoccus*, and Methanomethylophilaceae UBA71 clade. The highest replication rate was observed for *Methanobrevibacter*, the lowest was observed for the *Methanomethylophilaceae* representatives; this finding, however, was not statistically significant (Kruskal-Wallis test on all groups:  $p = 0.4$ , pairwise Wilcoxon rank sum test on selected taxa: mean FDR corrected  $p$  value = 0.615). Also, iRep values did not reveal any significant correlations with geographic origin, demographics or health aspects (Supplementary Fig. 3). In case of unfiltered iRep values from 69 genomes, the mean replication rate increased to  $1.9 \pm 0.9$ . According to this data set the highest replication rate was observed again for *Methanobrevibacter*, but here the lowest was observed for *Methanomassiliicoccus*. Interestingly, both subsets revealed higher replication rates of *Methanobrevibacter smithii\_A* (*Cand. M. intestini*) than for *Methanobrevibacter smithii* (Supplementary Fig. 3).



## The human archaeome carries a complex, previously unseen virome

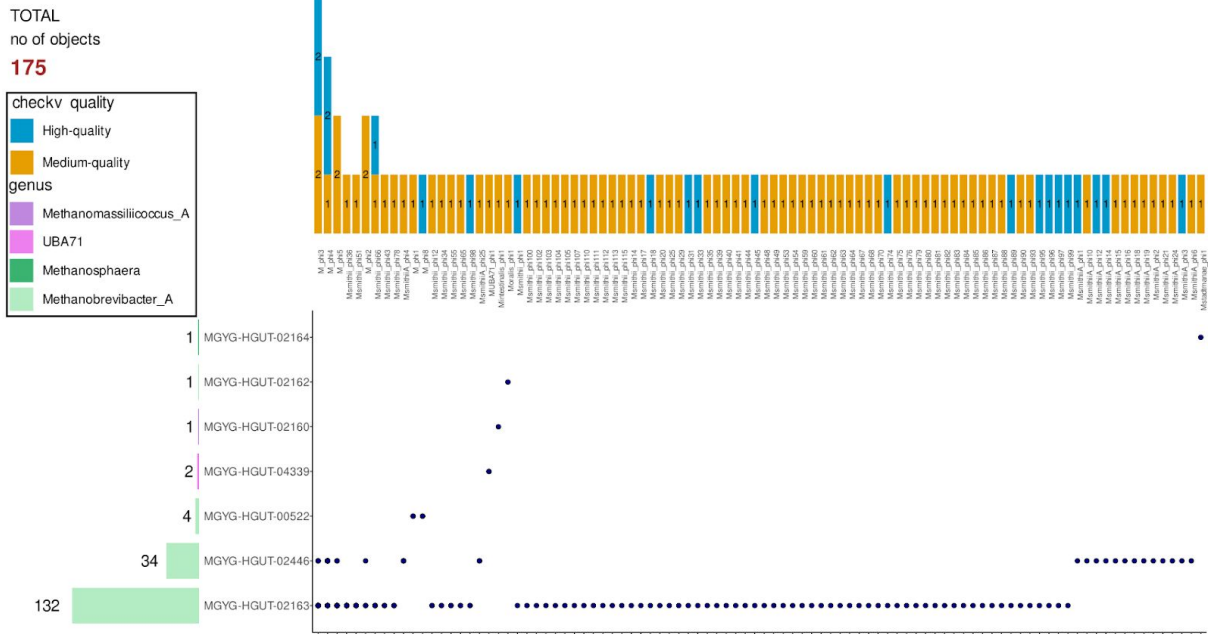
The virome associated with gut archaea is largely unknown. We identified up to 94 viral populations in our genome datasets by using CheckV (Nayfach, Camargo, et al. 2020). This number is the result of clustering 45 high-quality (>90% completeness) and 130 medium quality (50-90% completeness) archaeal proviruses at 95% identity and 80% coverage. The selected cut-off is repetitively used and is shown to be consistent with viral species definition (Duhaime and Sullivan 2012; Brum et al. 2015; Duhaime et al. 2017; Bobay and Ochman 2018; Roux et al. 2019; Gregory et al. 2016; 2020) (Fig. 5; Supplementary Table 5). Of the identified high-quality (HQ) proviruses, 91 viral species representatives were found to be specific for *Methanobrevibacter*, one for *Methanomassiliicoccus*, one for *Methanosphaera*, and one for Methanomethylphilaceae UBA71.

Notably, the identified proviruses seem to infect i) the same host regardless of geographical location which suggests that those viruses can be widespread, and ii) hosts of different species (specifically, prophages infecting both *M. smithii*, *Cand. M. intestini*, Supplementary Table 5) which suggests a broad host range.

While archaeal viruses in extreme environments have been discovered in the early 1970s (Torsvik and Dundas 1974; 1980), little is known about non-extremophilic viruses in the highly abundant mesophilic environments and only a few non-extremophilic archaeal viruses have been isolated so far (review and overview papers of archaeal viruses Snyder et al. 2015; Prangishvili et al. 2006; 2017). Only a handful of archaeal viruses have been described to infect the **same archaea genera** widespread in both the environment and the human host. In more detail, we found no similarities between the identified proviruses and the lytic phages Drs3 infecting *Methanobacterium formicicum* (Wolf et al. 2019), six viruses infecting various *Halorubrum* species (Pietilä et al. 2012), the provirus Msmi-Pro1 infecting *Methanobrevibacter smithii* ATCC 35061 (Krupovič, Forterre, and Bamford 2010) and the provirus  $\phi$ mru infecting *Methanobrevibacter ruminantium* (Ouwerkerk, Gilbert, and Klieve 2011). However, to the best of our knowledge, no viruses/proviruses have been identified in the past infecting *Methanomassiliicoccales* and *Methanobacteriales* members of the human gut.

We additionally explored the novelty of these proviruses by comparing them with the latest comprehensive human Gut Virome Database (GVD) (Gregory et al. 2020), and the Viral Refseq Database using the network-based viral classification tool vConTACT2 (Bin Jang et al. 2019). No clustering resulted between the high- and medium- quality identified viruses and both databases. Due to the lack of similar archaeal viral genomes in the reference databases, the classification and further characterization of discovered archaeal viruses through metagenomic approaches remains challenging.

Taken together, these results reveal that archaeal viruses probably have a currently underestimated diversity and likely ecological importance in the human gut microbiome. Consequently, the current and ongoing identification of archaeal viral species provide unique opportunities for future in-depth functional characterizations, taxonomic classification, estimation of the prevalence of each viral species, isolation and follow-up studies of virus-archaea dynamics.



**Fig. 5:** UpSet plot showing presence of viral species representatives color coded by the assigned quality (yellow for high-quality, and blue for medium-quality) in genome species where a member of the viral cluster was identified. Viral clusters representative were named as follows: if the cluster is of viruses infecting hosts of the same genus the name is Mgenus\_phi\_Number and if the cluster is of viruses infecting hosts of different genera the name is M\_phi\_Number.

## Human-associated archaea exhibit a lower proportion of horizontal gene transfer than animal-associated archaea

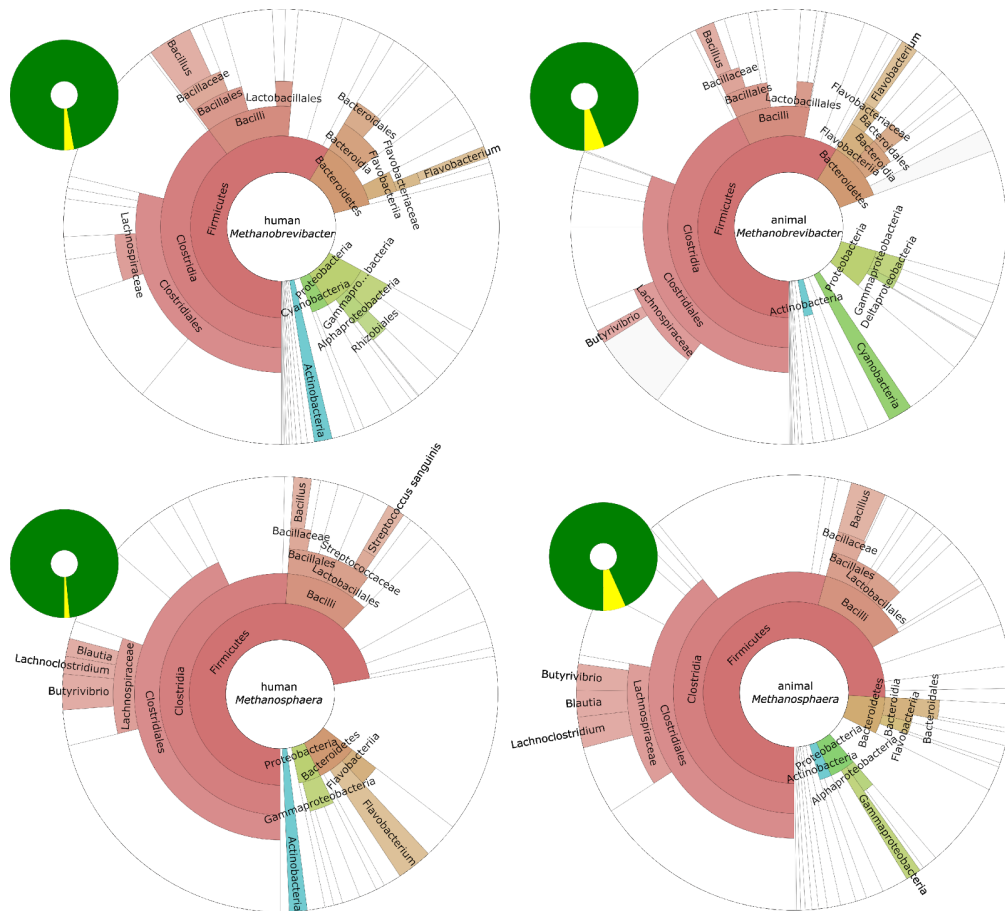
The adaptation of archaea to the human gut environment may have been favored by specific acquisition of genes from the resident bacterial community providing new functions. To assess this possibility, we compared the retrieved *Methanosphaera* and *Methanobrevibacter* genomes (strain list, only isolates and MAGs with 0% contamination) with isolates and genomes derived from animal sources (Supplementary Table 6; Fig. 6).

Human-associated methanogens revealed a significantly lower proportion of non-archaeal genes, irrespective of whether we considered isolates only or both isolates and MAGs. Human-associated *Methanobrevibacter* species carried on average approx. 2.84% genes annotated as of non-archaeal origin, which was significantly lower than the proportion of non-archaeal genes in animal-associated *Methanobrevibacter* species (6.09%; Mann-Whitney U test,  $p=0.00308$ ; genomes from isolates only: 6.36%). In particular *Methanobrevibacter smithii*/*smithii\_A* (*Cand. M. intestini*) representatives revealed a very low contribution of non-archaeal genes (2.11%; genomes from isolates only: 1.8%). Human-associated *Methanosphaera* species carried on average a proportion of 1.45% of genes of bacterial annotation (genomes from isolates only: 0.68%). Animal-associated *Methanosphaera*, however, contained a significantly higher proportion of bacterial genes (6.74%;  $p=0.00452$ , Mann-Whitney U test; genomes from animal isolates only: 5.31%).

An extraordinary high contribution of bacterial genes were observed for Methanomassiliicoccales, revealing a proportion of archaeal-annotated genes of 74.96% only (mean of 9 MAGs (0% contamination, high quality genomes; strain list) and 2 isolates). The proportion ranged from 65.72-96.88% in the MAGs, the genomes from isolates revealed an archaeal-gene annotation of 67.35 and 69.65%. This might reflect however an incomplete annotation of the Methanomassiliicoccales genomes based on the low number of available genomes and isolates.

In all cases, the largest contribution was observed from Firmicutes (Clostridia and bacilli; example bile salt hydrolases see below) with a lower contribution of Bacteroidetes, and Proteobacteria. Those clades represent the most abundant bacterial microbiome components, increasing the probability for HGT towards the archaeome (Lurie-Weinberger et al. 2012) (Fig. 6; Supplementary Material 3).

Our results indicate that adaptation towards the human host is not necessarily reflected by a higher proportion of genes derived from the human gastrointestinal bacteriome.



**Fig. 6: Contribution of bacterial-annotated genes** in human- (left) and animal- (right) associated *Methanobrevibacter* and *Methanosphaera* species: proportion (small circles, yellow refers to proportion of bacterial annotation) and potential bacterial origin (taxa as displayed in the large circles). Unclassified taxa are whitened out. Only MAGs with 0% contamination and of high quality (taken from “strain list”) and genomes from isolates were analyzed.

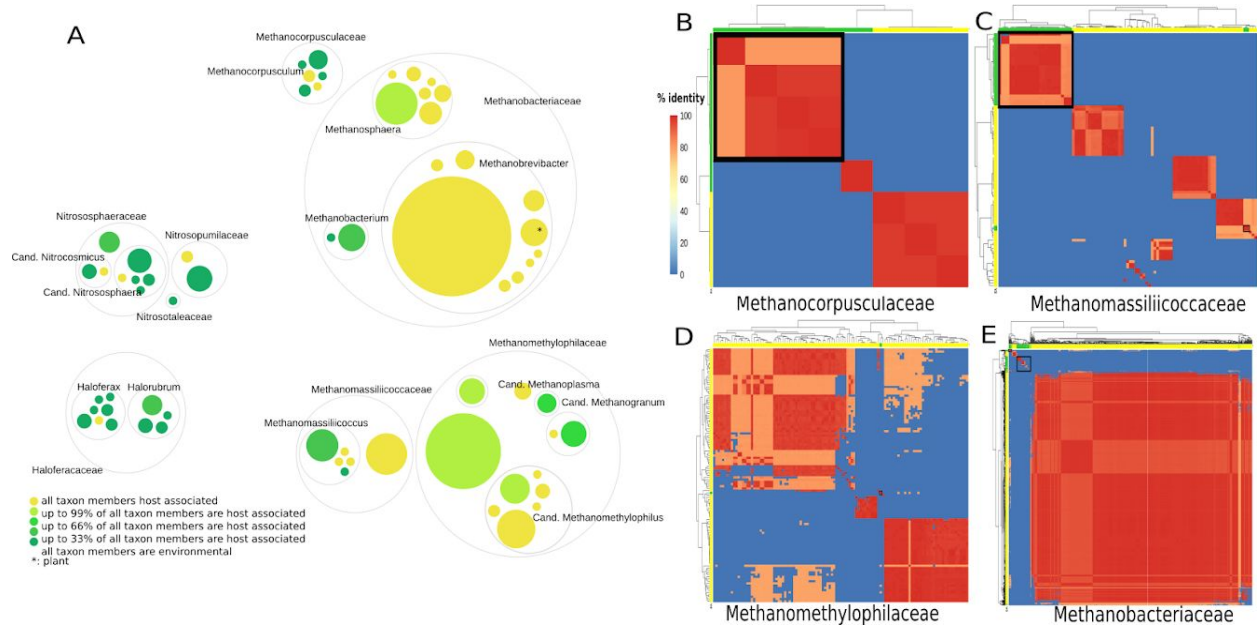
## Host-associated archaea diverge from environmental relatives

We reasoned that host-associated archaea diverge taxonomically and functionally due to the characteristics of their individual host environments.

In 16S rRNA gene-based analyses (details given in Star\*Methods), we found that members of genera *Methanobrevibacter* and *Methanosphaera*, as well as *Cand. Methanomethylophilus* belonged almost exclusively to taxa from host-associated (animal, human, plant) sources, whereas *Methanocorpusculum*, Nitrososphaeria and Haloferaceae were more related to environmental strains (Fig. 7A). ANI-based analyses of the families Methanobacteriaceae, Methanocorpusculaceae, Methanomethylophilaceae and Methanomassiliococcaceae revealed an overall clear separation between the MAGs of different origins (Fig. 7, B-E). More specifically, for Methanomassiliococcaceae (Fig. 7C), we observed that the two gut genomes classified as *M. luminyensis* (GUT\_GENOME132203 and GUT\_GENOME140888) cluster within an environmental archaeal MAGs clade whose members are taxonomically classified as uncultured *Methanomassiliococcus*; however, both human-derived genomes share a higher ANI value compared to the genomes of the sister clade. Concerning Methanobacteriaceae (Fig. 7E), GCA\_002509095.1 (Methanobacteriaceae archaeon UBA254) and GUT\_GENOME283178 (*Methanosphaera* sp900322125) shared an ANI value of 96.09% while every other pairwise comparison between human and environmental archaeal MAG had an ANI value less than 88.09%.

Based on taxonomic classification resolved at the species level, most species identified in the human gut are indeed host associated (reviewed in Bang and Schmitz 2018; Moissl-Eichinger et al. 2018; Borrel et al. 2020). More specifically, based on the information on their respective biomes, the archaeal genomes of this study can be classified into three groups: i) exclusively found in the human gut, ii) host (human, animal, plant)-associated and iii) widespread in the environment, with the first two groups representing the majority (reviewed in Bang and Schmitz 2018; Moissl-Eichinger et al. 2018; Borrel et al. 2020). Following this classification and based on the current availability of genomes and metadata, *H. massiliensis*, *M. oralis*, *M. smithii*, *M. smithii\_A* (now: *Cand. M. intestini*), *M. stadtmanae*, *M. intestinalis* and *M. alvus* can be considered to group (i). Species belonging to group (ii) include *M. woesei*, and *M. cuniculi*. Species of group (iii) are represented by *H. lipolyticum* (Cui et al. 2006), *M. arboriphilus* (Zeikus and Henning 1975; Khelaifia et al. 2014b) and *M. luminyensis* (Dridi et al. 2012; Borrel et al. 2017), which are widespread in various kinds of environments.

It should be noted that for certain clades only a limited number of genomes is currently available, and thus their host/environmental tropism remains to be precisely determined (e.g. *Methanomethylophilus* sp001481295, *Methanosphaera* sp900322125, *Methanobacterium* sp000499765, *H. massiliensis*, *Methanomassiliococcus*). Notably, in both types of analyses, the genus *Methanobrevibacter* was overwhelmingly affiliated with the human and other hosts. This is a unique trait amongst all analyzed archaeal genera (Fig. 7A).



**Fig. 7: Comparison of host-associated and environmental relatives.**

**A)** Circle packing plot, displaying the environmental or host-associated nature of specific taxa. Analysis performed on 16S rRNA gene level. Number of sequences analyzed per taxon are indicated by circle size; colors indicate the proportion of host-associated signatures (see legend). The largest contribution was observed from *Methanobrevibacter smithii* sequences (n= 192). Note, that the yellow color (“host-associated”) also includes human, animal and plant (\*; only *M. arboriphilus*)- associated taxa.

**B-C-D-E)** Average nucleotide identity (ANI) percent identity heatmap visualization. ANI analysis based on MinHash sequence mapping was performed using fastANI and visualized using the pheatmap library in R. ANI values represented range from 75–80% ANI colored in light orange, 80-90% ANI colored in darker orange, and over 95% ANI in red. Heatmap for genomes assigned to the taxonomic family of **B)** Methanocorpusculaceae, **C)** Methanomassiliicoccaceae, **D)** Methanomethylophilaceae, and **E)** Methanobacteriaceae. Genomes isolated from the human gut microbiome (labeling on the x and y-axes in yellow) can be separated from the genomes isolated from the environment (labeling on the x and y-axes in green; Supplementary Table 7). Environmental archaeal genome clustering is marked with a black square.



## Metabolic and functional interaction of the archaeome with the gut environment

We analyzed specific features that could indicate the advanced interaction of the human-associated GIT archaea with their gut environment, i.e. with host and non-archaeal microbiome components (Fig. 8).

Loss of genes involved in dealing with oxidative stress are considered as a trait of host-association, as environmental strains have to face non-permanently, strict anaerobic conditions, while this is not the case for strains inhabiting the GIT. We therefore specifically analyzed the presence of genes associated with oxygen resistance (catalase, superoxide dismutase, peroxiredoxin, rubredoxin and thioredoxin; Lyu and Lu 2018). Catalase was detected in some Methanomassiliicoccales (mainly *Methanomassiliicoccus* representatives) and Haloarchaea, and in *Methanobrevibacter arboriphilus* and *Methanobacterium*. The presence of a superoxide dismutase was rarely detected, namely in *Haloferax* and *Halorburum* members. None of the *Methanobrevibacter* representatives, except *M. arboriphilus*, carried the peroxiredoxin gene. In contrast, thioredoxin and rubredoxin were detected in the majority of the genomes.

Additional functions of interest are adhesins and bile salt hydrolases (i.e. choloylglycine hydrolase, CGH). A number of adhesins was specifically annotated in the genomes of *Methanobrevibacter* and *Methanosphaera* species (Fig. 8A). CGH homologues were detected in 11 out of 28 of the archaeal species, including the five most prevalent ones (*M. smithii*, *Cand. M. intestini*, *M. stadtmanae*, *M. alvus* and *Cand. M. hominis*). CGH genes were not detected in any of the Methanomassiliicoccus genomes and in the Haloferaceae, indicating their importance for specialization toward the human gut. It should be noted that the CGH genes detected in Methanomassiliicoccales, Methanomicrobiales and Methanobacteriales formed separate clusters within Firmicutes bile salt hydrolases gene tree (Supplementary Fig. 6), indicating their potential acquisition from different events of HGTs.

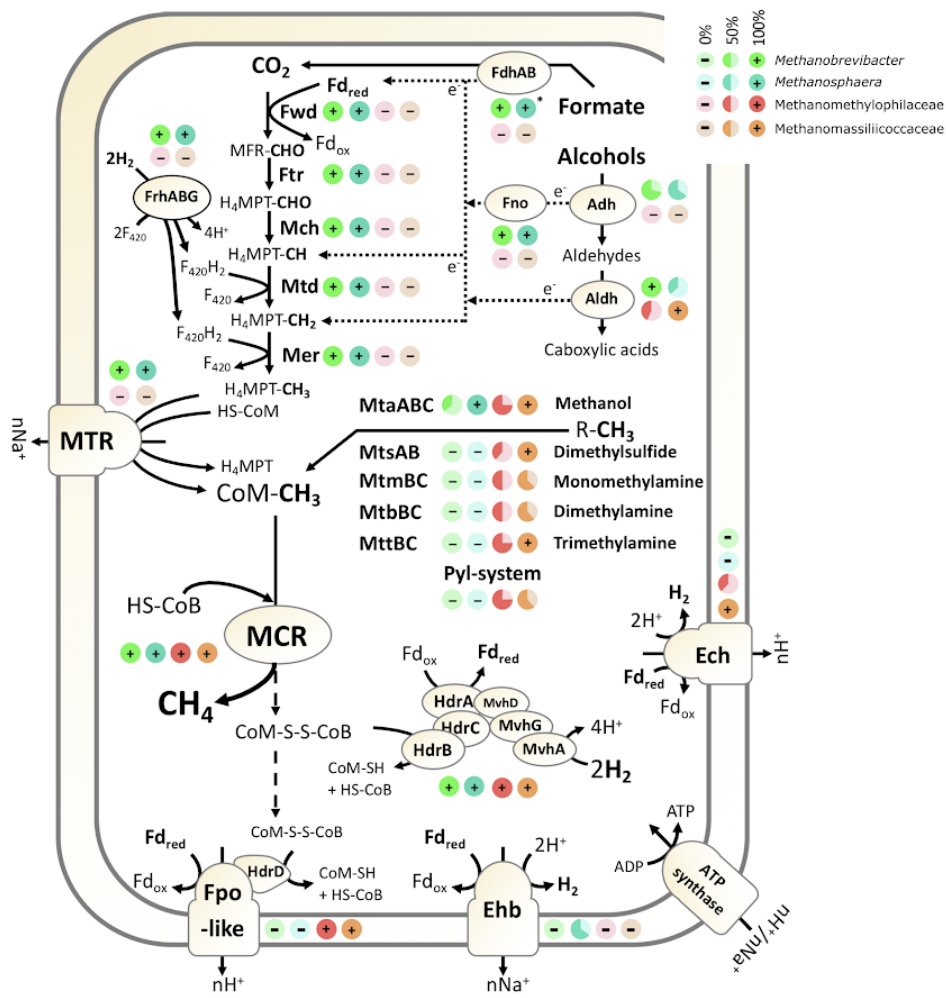
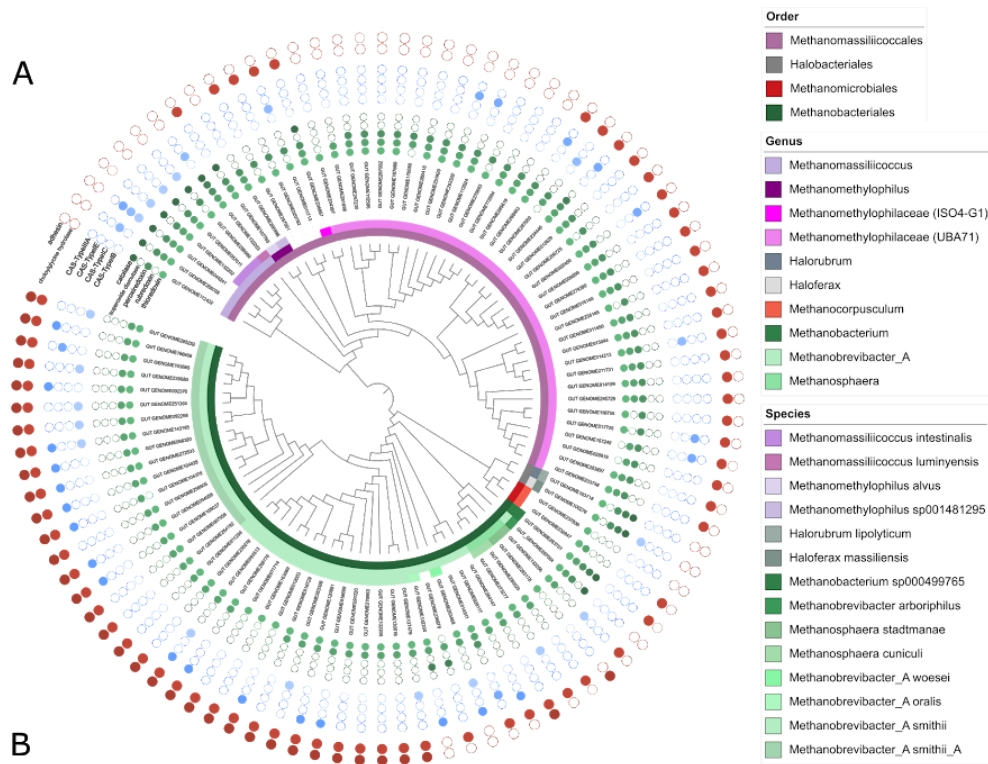
Additional adaptations were observed at the metabolism level. Apart from key components of methanogenesis, MCR and Hdr/Mvh complexes, the main gut methanogens (Methanobacteriales and Methanomassiliicoccales) possess very distinct methanogenesis pathways (Fig. 8B; Supplementary Table 8). For example, differently to all Methanomassiliicoccales, all human gut *Methanobrevibacter* have the genetic potential for formate and H<sub>2</sub>/CO<sub>2</sub> utilisation. However, both Methanobacteriales and Methanomassiliicoccales representatives carry the *mtaABC* genes, providing the genetic potential to use methanol. In total, 83% of the MAGs have these genes, suggesting that methanol utilization might provide a selective advantage in the human gut. However, the conditions under which *Methanobrevibacter* use methanol and whether this compound is a methanogenesis substrate or enters an anabolic pathway remain unknown.

The two dominant *Methanobrevibacter* species display the genetic potential to use alcohols (likely secondary alcohols and ethanol) as electron donors for methanogenesis. One of the *Methanosphaera* species may also have the genetic capacity to reduce methanol with ethanol for methanogenesis as described earlier (Hoedt et al. 2016), but this species was only encountered once in our analyses, and *Methanosphaera stadtmanae*, the main *Methanosphaera* species in human, cannot perform this pathway.

The majority (11/13) of the human-gut associated Methanomassiliicoccales species code for the MttBC methyltransferase and corrinoid protein needed for methanogenesis from TMA. This capacity would allow them to decrease the concentrations of this molecule produced by gut microbiota and involved in cardiovascular diseases (Brugère et al. 2014; Borrel et al. 2017). The presence of the *mttBC* genes was detected in a larger proportion of the Methanomassiliicoccales MAGs originating from Europe and North America (~60%) with respect to Africa and Asia (~40%) or Oceania (17%) (Supplementary Fig. 7). These variations may reflect different TMA-production capacity by bacteria in the microbiota across these populations and diet habits. One of the two Methanomethylophilaceae species lacking TMA-utilization capacity (*Cand. Methanoprimatia macfarlanii*) also lacks MtbBC and MtmBC methyltransferases and corrinoid proteins for dimethylamine and monomethylamine utilization, respectively. However, several strains of this species have the genes to encode Pyrrolysine synthesis (*pylSBCD*), a proteinogenic (UAG codon-encoded) amino-acid almost only present these

methylamine methyltransferases (Srinivasan, James, and Krzycki 2002; Brugère et al. 2018). The absence of detection of the methylamine methyltransferases in these MAGs, including MttBC for TMA utilization, is thus likely due to their incompleteness. The other species lacking methylamine methyltransferase, corresponding to Methanomassiliicoccales Mx02 sp. (Borrel et al. 2017), also lacks any other genes known to be involved in methyl-compound utilization or in any alternative methanogenesis pathways (Supplementary Table 8). The absence of these methanogenesis genes in all the MAGs of Methanomassiliicoccales Mx02 sp. and in previously obtained related MAGs, support assumptions (Borrel et al. 2017; Söllinger et al. 2016) on the presence of novel methanogenesis pathways likely based on unknown methyltransferases, or another metabolic route in the Methanomassiliicoccales. We propose the name of *Candidatus* Methanarchanum hacksteinii Mx02 for this potential new genus and species (representative MAG: GUT\_GENOME287001).

Our observations indicate that members of the human archaeome possess a variety of specific adaptations towards this specific environment, many of which are probably not yet discovered.



---

**Fig. 8: Adaptations of human gut archaea with respect to ecosystem and metabolism.** Note: *M. smithii\_A* corresponds to *Cand. M. intestini*.

**A)** Overview of the absence and presence of genes involved in host-interaction (red), oxygen resistance (green) and detected CAS systems (blue). Genomes (strain-list) were analyzed using MaGe Microscope and genes were counted as present when automatic annotation was positive (“putative” annotation was counted as positive).

**B)** Methanogenic pathways in 23 human gut-associated Methanobacteriales and Methanomassiliicoccales species. The proportion of species with a given protein or protein complex is indicated by pie charts for *Methanobrevibacter* (n=7), *Methanosphaera* (n=3), Methanomethylphilaceae (n=8) and *Methanomassiliicoccus* (n = 5). For clarity, the nature of the electron transporter and some intermediate steps in the electron transfers are not displayed for formate and alcohol utilization. R-CH<sub>3</sub> corresponds to methanol, dimethylsulfide, monomethylamine, dimethylamine or trimethylamine. Alcohol could be ethanol or secondary alcohols. The absence of certain enzymes may be due to the fact that the MAGs are incomplete.

## DISCUSSION

This work adds novel information on the biology of the human gastrointestinal archaeome, by characterizing an extended collection of 1,167 non-redundant archaeal genomes. This assessment of the genome collection and the catalogue of 1.8 million putative proteins can now serve as a reference base for extensive investigation of the hitherto unexplored gastrointestinal archaeome.

The comprehensive analysis of the globally distributed gut archaeome offers a unique resource to generate hypotheses for future studies. This includes aspects on the archaeal physiology, the interaction with the bacterial microbiome and the virome, the comparison of animal- and human-host association, and the archaeal cross-talk with the human host. Moreover, the provided resource can serve as an important database for targeted cultivation of members of the archaeome, as well as their unique virome.

One of the main remaining questions is whether archaeal pathogens do exist, since a per se pathogenic capacity of archaea has never been identified (Borrel et al. 2020). Therefore, our resource is an important basis for further exploring potential pathogenic traits, such as virulence factors.

As our insights are largely predicted from genomic information, additional work is urgently required to address the archaeome functional capacity using expression and cultivation-based assays. Furthermore, our dataset is based on publicly available samples, which were originally processed for the analysis of the bacterial component of the microbiome (Almeida et al. 2019; Nayfach et al. 2019; Pasolli et al. 2019; Nayfach, Roux, et al. 2020; Almeida et al. 2020). Therefore, archaeal species that require specialized methods (Mahnert et al. 2018) for cell lysis and DNA extraction may be missing from our current collection. Moreover, as several taxa in our collection are only represented by single genomes, additional conspecific strains will be needed to validate many of our genomic insights.

Future efforts should seek to extend the dataset beyond the gastrointestinal environment, to other human body sites (Koskinen et al. 2017) and hosts (Moissl-Eichinger et al. 2018). Moreover, incorporating both transcriptomics and proteomics data in the future will further reinforce the genomic predictions and improve our understanding of the regulation of archaea physiology and host adaptation.

Overall this work contributes substantially to the understanding of the microbiome as a complex multi-domain microbial network of the human gastrointestinal tract (Taffner et al. 2018; Nayfach, Roux, et al. 2020; Youngblut, Reischer, et al. 2020; Youngblut, Cuesta-Zuluaga, et al. 2020). We showed that despite the availability of more than 1,000 genomes, the human archaeome still remains undersampled. Future work aiming to expand the archaeal pangenome will enable the identification of specific archaeal functions/molecular mechanisms that may be causal in microbiome-associated human health conditions.

## ACKNOWLEDGEMENTS

CME received funding from the Austrian Science Fund FWF (P 32697, P 30796), which is highly appreciated. CME and AM gratefully acknowledge computational resources of the MedBioNode at the Medical University Graz and the Life Science Compute Cluster (LiSC) operated by the Computational Systems Biology group at the University of Vienna. We thank the Medical University Graz ZMF Galaxy Team: Core Facility Computational Bioanalytics, Medical University of Graz, funded by the Austrian Federal Ministry of Education, Science and Research, Hochschulraum-Strukturmittel 2016 grant as part of BioTechMed Graz. We gratefully acknowledge the discussion with Katharina Bick on taxonomic naming, and the scientific support provided by Zobia Hameed.

RS received partial funding from Kiel Marine science cluster at the CAU and the German ministry of education and science, BMBF (031B0851B), which is highly appreciated. CC and AW used resources of the high-memory computer nodes and computation support from the Computing Center of the CAU Kiel. AA and RF are supported by EMBL.

JFB acknowledges a grant from Hub Innovergne (“Investissements d’Avenir” 16-IDEX-0001 CAP 20–25).

GB and SG acknowledge support by the Institut Pasteur and the French National Research Agency (ANR) through grants Methevol (ANR-19-CE02-0005-01) and Archevol (ANR-16-CE02-0005-01).

## AUTHOR CONTRIBUTIONS

Conception and design: AA, RS, CME, SG, GB

Methodology and analysis: CC, AM, GB, CME, JFB

Collection, data processing and data generation: AA, RF, CC, AW, AM, GB

Data interpretation: CC, AM, GB, JFB, RS, CME

Manuscript preparation: CC, AM, GB, JFB, RS, CME

## DECLARATION OF INTERESTS

The authors declare no competing interests.



## STAR METHODS

### RESOURCE AVAILABILITY

#### *Lead Contact*

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Christine Moissl-Eichinger [christine.moissl-eichinger@medunigraz.at](mailto:christine.moissl-eichinger@medunigraz.at).

#### *Data and Code Availability*

All the recovered genomes are available at [\[http://ftp.ebi.ac.uk/pub/databases/metagenomics/genome\\_sets/archaea\\_gut-genomes.tar.gz\]](http://ftp.ebi.ac.uk/pub/databases/metagenomics/genome_sets/archaea_gut-genomes.tar.gz). All the other considered genomes and metagenomes are publicly available in NCBI, and referenced. This study did not generate code, mentioned tools used for the data analysis were applied with default parameters unless specified otherwise.

## METHOD DETAILS

- *Dataset description*

We selected archaeal genomes sampled from the human gut publicly available as part of the Unified Human Gastrointestinal Genome (UHGG) collection (Data access June 2020, <https://www.ebi.ac.uk/metagenomics/genomes>) (Almeida et al. 2020) and the MGnify resource (Mitchell et al. 2020). Briefly, this collection of non-redundant metagenomes-assembled genomes (MAGs) and isolates was collected from public repositories and metadata for each genome was gathered as well (see Almeida et al. 2020 for more details). Genomes were compared using Mash v2.1 (Ondov et al. 2016) and for genomes that were estimated to be identical, had a Mash distance of 0, only one was selected. In addition, we included genomes of ‘*Ca. Methanomethylophilus alvus*’ (Borrel et al. 2012), ‘*Ca. Methanomassiliicoccus intestinalis*’ (Borrel et al. 2013), as well as human gut-derived MAGs of Methanomassiliicoccales Mx02, Mx03, Mx06 and additional *Ca. M. intestinalis*’ (Borrel et al. 2017), and the human isolate *Methanobrevibacter arboriphilus* ANOR1 (Khelaifia et al. 2014b) to complete the dataset. Those genomes were assigned a genome accession number (GUT\_GENOME286998 to GUT\_GENOME287004), as given in Supplementary Table 1. This brought the total number of genomes used for the analysis in this paper up to 1,167 genomes.

- *Genome quality and taxonomic classification*

The completeness of the non-redundant 1,167 genomes was evaluated by CheckM v1.0.11 (Parks et al. 2015) and only genomes which were more than 50% complete and had less than 5% contamination were selected (following the protocol from Almeida et al. 2020).

DRep v2.0.0 (Olm et al. 2017) was used to dereplicate the complete dataset at 95% and 99% average nucleotide identity (ANI). 95% ANI values were selected to separate between species boundaries (n=27) and a cutoff of 99% was chosen to distinguish archaeal MAGs on strain level (n=98) (Jain et al. 2018). The resulting strain- and species representatives are given in Supplementary Table 1.

All genomes were taxonomically annotated following the procedure given in (Almeida et al. 2020). The taxonomic assignment was performed using the Genome Taxonomy Database Toolkit (GTDB-Tk) version 0.3.1 (database release 04-RS89) (Chaumeil et al. 2020) and default parameters. Methodology is detailed in Supplementary Fig. 1.

- *Genome annotation and protein catalogue*

Protein-coding sequences (CDS) were predicted and annotated with Prokka v1.14.5 (Seemann 2014) using the parameters ‘--kingdom Archaea’ to include non fragmented archaea curated proteins from the UniProtKB database and ‘--rfam’ to scan for ncRNAs. CDS were further characterized using eggNOG mapper v2.0.0 (Huerta-Cepas et al. 2017) and the eggNOG database v5.0 (Huerta-Cepas et al. 2019).

The protein catalogue was generated by combining all predicted CDS (total number 1,790,493) derived from the 1,167 non-redundant archaeal genomes. MMseqs2 linclust (Steinegger and Söding 2017) was used to cluster the concatenated proteins dataset using the options '--cov-mode 1 -c 0.8' (minimum coverage threshold of 80% the length of the shortest sequence) and '--kmer-per-seq 80'. To reduce the risk of contaminants, the proteins were filtered to remove all non clustered proteins. This gave a total of 28,581 proteins clustering at 50% identity (Supplementary Table 2) visualized using the library heatmap (Kolde 2019) in R.

In addition to the protein catalogue, the various species- and strain-subsets of the total 1,167 archaeal genomes (Supplementary Table 1) were submitted to MaGe MicroScope (Microbial Genome Annotation & Analysis Platform, Vallenet et al. 2009), for detailed analyses of genomic synteny.

- *Correlation with geographic and demographic parameters*

Beside genomic information (genome length, number of contigs, N50, GC content, genome completeness, genome contamination, number of rRNAs and tRNAs), eleven metadata categories (numerical 2, categorical 9) could be considered for the data set. Information about the geographic origin was available for 1,063 genomes (91% of the data set, covered countries from maximum to minimum: United States, Israel, Spain, Sweden, Fiji, United Kingdom, Austria, Denmark, Netherlands, France, China, Peru, Germany, Madagascar, United Republic of Tanzania, Australia, Canada, Ireland, Italy, Russia, El Salvador, Iceland, Mongolia, Norway). Information about lifestyle was available for 1,054 genomes (90%, max-min: urban, rural, semi-urban), health state (healthy, diseased) for 894 genomes (77%), age group (adult, elderly, child, teenager, infant) for 825 genomes (71%), gender (female, male) for 620 genomes (53%), body mass index group (normal weight, overweight, obesity class 1, underweight, obesity class 2, extreme obesity class 3) for 505 genomes (43%), name of disease (colorectal cancer CRC, infection, type 2 diabetes T2D, Adenoma, obesity, Ulcerative colitis UC, nonalcoholic fatty liver disease NAFLD, Parkinson, Ankylosing spondylitis - arthritis AS, fecal microbiota transplant FMT, cirrhosis) for 303 genomes (26%), treatment (antibiotics) for 241 genomes (21%).

Supervised classification and regressions with RandomForest were applied to predict respective metadata categories from the unified archaeal protein catalogue with the q2-sample-classifier plugin (Bokulich et al. 2018). First subsets of each metadata category were created from the entire protein matrix and randomly split into a training and test set with the proportions 80 to 20%. By using K-fold cross validation, the training set served as a learning model to predict class probabilities with settings for optimized feature-selection and parameter tuning. In the end, model accuracy was determined by comparing the predicted values between the training and test data sets.

- *Pan-genome analysis*

Pan-genome analysis was performed using Roary version 3.11.2 (Page et al. 2015) using a minimum amino acid identity of 80% and no paralog splitting. Since Roary was not intended for comparing extremely diverse sets of genomes, it was used for pan-genome analysis for archaeal genomes of the same families or the same genus. We used Heaps' Law ( $\eta = \kappa * N - \alpha$ ) to estimate whether we have an open or a closed pan-genome (Tettelin et al. 2008), a first estimate for all predicted CDS and a second estimate excluding singletons from gene presence absence matrices. This analysis was carried out in the R package 'micropan' (Snipen and Liland 2015) using a default permutation value of 100, where  $\eta$  is the predicted number of genes for a particular number of Genomes (N), and  $\kappa$  (intercept parameter) and  $\alpha$  (decay parameter) are the constants used to fit the curve after the genomes are ordered in a random way. An open pan-genome is indicated by an  $\alpha < 1$  while a closed pan-genome is indicated by a  $\alpha > 1$ .

- *Estimation of replication rates*

Replication rates were estimated with iRep (Brown et al. 2016) after mapping reads with Bowtie 2 (Langmead and Salzberg 2012) and post-processed with samtools (Li et al. 2009). Since the original raw reads were not available for each representative genome, as the remaining read sets were not made publically available, read coverage between the origin of replication and its terminus (main input for iRep) could only be mapped for 69 out of 98 genomes.

- *In-depth taxonomic and phylogenetic analyses of the various genera*

ANI distances and tree matrices were calculated using the online resources of the enveomics platform (Rodriguez-R and Konstantinidis 2016), MaGe (Vallenet et al. 2009), as well as Microbial Genomes Atlas (MiGA) (Rodriguez-R et al. 2018). Phylogenetic trees, based on the ANI tree matrix, were annotated using the iTOL tool (Interactive Tree Of Life) (Letunic and Bork 2019), and processed using InkScape. For specific considerations involving additional genomes from animals, a subselection of the archaeal genomes was reanalyzed together with the additional genomes following the same settings as described for the protein catalogue procedures above (respective datasets are given in the Supplementary Table 3 and Supplementary Table 9 as described in the main text).

McrA genes were extracted via MaGe, hosting all strain level genomes. McrA genes were aligned using MegaX (Kumar et al. 2018), and a maximum likelihood tree was calculated (default settings).

- *Horizontal gene transfer analysis*

The information from protein catalogue incl. eggNOG annotation was used for the assessment of the horizontal-gene transfer candidates. For this, the taxonomic “best hit” information for all annotated genes was used; only genomes revealing 0% contamination were used for this assessment. All data are provided in Supplementary Table 6. Data were visualized using Krona (Ondov, Bergman, and Phillippy 2011).

- *Detection of virulence and resistance genes*

To predict potential virulence genes in all 1,167 archaeal genomes, ABRicate version 0.5 (<https://github.com/tseemann/abricate>) was used to profile the following databases: CARD (Jia et al. 2017), Resfinder (Zankari et al. 2012), PlasmidFinder (Carattoli et al. 2014), ARG-ANNOT (Gupta et al. 2014), EcoOH (Ingle et al. 2016), MEGARes 2.0 (Doster et al. 2020) as well as NCBI AMRFinderPlus (Feldgarden et al. 2019). Since ABRicate is solely based on DNA sequences, blastX searches using DIAMOND (Buchfink, Xie, and Huson 2015) was used to complement results from ABRicate on the level of protein sequences in the virulence factor database (VFDB version 20191122) (Chen et al. 2005; Liu et al. 2019) and CARD together with the Resistance Gene Identifier (RGI) (Alcock et al. 2020).

Specific groups of proteins and genes involved in human interaction were investigated according to available annotations from MaGe (Vallenet et al. 2009) and eggNOG-mapper (Huerta-Cepas et al. 2017).

- *Viral identification, quality estimation and comparisons to viral databases*

To assess the presence of prophages, and estimate completeness and quality of potential proviruses, CheckV (Nayfach, Camargo, et al. 2020) was used to scan all MAGs. To ensure we overcome possible contamination issues that can potentially result from the binning process, we selected proviruses flanked within archaeal contigs for this analysis. CheckV looks for host-virus boundaries based on difference in GC content and gene annotation in a sliding window approach. Proviruses that had a CheckV quality assignment of medium-quality (50-90% completeness), high-quality (>90% completeness) or complete were considered for further analysis. Quality assignments by CheckV are based on Minimum Information about an Uncultivated Virus (MIUViG) standards (Roux et al. 2019). The selected phages were subsequently clustered with MMseqs2 using the ‘linclust’ function with the same parameters previously specified and Mmseqs function ‘result2repseq’ to select a viral cluster representative.

ORFs of viral populations with the previously specified MIUViG quality were used as input for vConTACT2 (Bin Jang et al. 2019) including Viral RefSeq genome (version 97). VconTACT2 is used to affiliate a family or a genus rank group to viral populations and thus to determine taxonomic diversity.

A recent study was published by Gregory et al. 2020 (Gregory et al. 2020) where a human Gut Virome Database (GVD) harbors 33,242 viral populations including 0.1 % archaeal viruses resulting from 2,697 gut metagenomes from 32 studies. This dataset was used as a reference database to scan the identified viral scaffolds using MMseqs2 “easy-linclud” function at 50, 80, 90 and 95% identity.

In addition, we BLASTED (50% identity and 80% coverage cut off) identified proviruses to characterized proviruses and lytic phages infecting members of the same genera we find in our dataset. The proviruses Msmi-Pro1 (coordinates 1693232–173205) infecting *Methanobrevibacter smithii* ATCC 35061 (Accession Number CP000678) (Krupovič, Forterre, and Bamford 2010), and the provirus  $\phi$ mru infecting *Methanobrevibacter ruminantium* were used (Ouwkerk, Gilbert, and Klieve 2011). Additional lytic phage Drs3 infecting *Methanobacterium formicicum* (Wolf et al. 2019),  $\psi$ M1 infecting *Methanothermobacter thermoautotrophicum* (Meile et al. 1989), and six viruses infecting various *Halorubrum* species (Pietilä et al. 2012) were used for comparison.

- *Comparison to environmental archaea*

For considerations based on 16S rRNA genes, 16S rRNA genes of representative genomes were extracted using Metaxa2 (Bengtsson-Palme et al. 2015) (n=314; not all 16S rRNA genes could be recovered). This dataset was supplemented with data from amplicon sequencing studies and clone sequences from archaeal signatures from human gastrointestinal samples (dataset described in Borrel et al, 2020; n=381 in total). These sequences were aligned and classified using the SILVA rRNA database (Quast et al. 2013). More specifically, the retrieved 16S rRNA genes were subjected to the ACT tool (alignment, classification and tree service) (Carver et al. 2005), using the following parameters: Basic alignment parameters: “removed”; Search and classify, min. identity with query sequence: “0.95”, Number of neighbours per query sequence: “10”; Compute tree, Workflow: “Denovo including neighbours” and default parameters; Advanced tree computation parameters, Positional variability filter: “None”, Domain: “Archaea”. Unclassified sequences were removed from the dataset. Via SILVA SINA, 10 next neighbours were selected, and information on their isolation source was gathered through NCBI (Supplementary Material 2; Supplementary Table 7; the final dataset contained 566 sequences). Grouping was performed on genus/species level, and information on the percentage of host-associated archaea in all groups was displayed as a circle packing plot (RawGraphs online tool, <https://app.rawgraphs.io/>).

For genome-based analyses, a set of 623 archaeal MAGs identified from environmental and gastrointestinal samples (for example, rumen, guinea pigs and baboon faeces) was used as a reference dataset for comparison to the set of archaea isolated from the human gut microbiome (Parks et al. 2017; Almeida et al. 2020). All environmental genomes used were as well over 50% complete up to 90% complete with less than 5% contamination as well. To estimate the pairwise Average Nucleotide Identity (ANI) distance between environmental archaeal genomes dataset (Supplementary Table 7) and the archaeal genomes from the human gut microbiome, we used fastANI (Jain et al. 2018), a tool that effectively discriminated intra- and inter-species boundaries for over 90K prokaryotic genomes.

- *Metabolic interaction of the archaeome with the gastro-intestinal environment*

Proteins involved in methanogenesis were searched in all genomes using custom Hidden Markov Models (HMM) profiles (threshold e-value  $10^{-5}$ ) implemented in Maccyfinder (Abby et al. 2014). This allowed to determine the presence of enzymatic complexes on the basis of the presence of all or most subunits. The presence in the 26 methanogenic species was first evaluated based on the representative genome (that are the most complete/less contaminated). If a majority of the MAGs in a species have an enzyme, then this enzyme was considered as present in the species, even if absent from the best representative genome.

- *Functional interaction of the archaeome with the gastro-intestinal environment (exploring oxygen resistance, adhesins and bile salt hydrolases)*

Specific functions were searched (“search by keywords”-function) in MaGe (Vallenet et al. 2009) (all strain genomes have been made publicly accessible). Presence and absence information was used for tree annotation through iTOL (Letunic and Bork 2019). The backbone tree was based on ANI similarity as described above.

#### Tools used for data visualization

PCoAs and other graphical displays based on the unified archaeal protein matrix were calculated and visualized in Qiime2 (Bolyen et al. 2018) and Calypso (Zakrzewski et al. 2017). Venn diagrams were created with creately (<https://creately.com/>). Alluvial plots and circle packing plots were generated with RAWGraphs (<https://app.rawgraphs.io/>). Strip charts were created with Calypso. Phylogenetic trees, based on the ANI tree matrix, were annotated using the iTOL tool (Interactive Tree Of Life). All figure panels were created using InkScape.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were conducted by using R, Qiime2 and Calypso. Statistical significance was determined by non-parametric tests including Spearman rank correlations, PERMANOVA, Wilcoxon rank sum tests for pairwise analysis, Mann-Whitney U tests for unpaired data and Kruskal-Wallis tests if significance had to be determined for all groups. Significance was considered at an alpha <0.05 after 999 permutations. P-values were corrected for multi hypothesis testing using the false discovery rate (FDR).

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Unified Human Gastrointestinal Genome (UHGG)	Almeida et al. 2020	See Supplementary Table 1 for details
<i>Ca. Methanomethylophilus alvus</i>	Borrel et al. 2012	See Supplementary Table 1 for details
<i>Ca. Methanomassiliicoccus intestinalis</i>	Borrel et al. 2013	See Supplementary Table 1 for details
Methanomassiliicoccales Mx02, Mx03, Mx06 and additional <i>Ca. M. intestinalis</i>	Borrel et al. 2017	See Supplementary Table 1 for details
<i>Methanobrevibacter arboriphilus</i> ANOR1	Khelaifia et al. 2014b	See Supplementary Table 1 for details
623 archaea MAGs identified from environmental and gastrointestinal samples	Parks et al. 2017	See Supplementary Table 2 Parks et al. 2017 or Supplementary Table 7 this study.
Raw sequencing data	Qin et al. 2010; Vrieze et al. 2012; Qin et al. 2012; Karlsson et al. 2012; Le Chatelier et al. 2013; Li et al. 2014;	NCBI Sequence Read Archive (SRA) - see Supplementary Table 1 for details

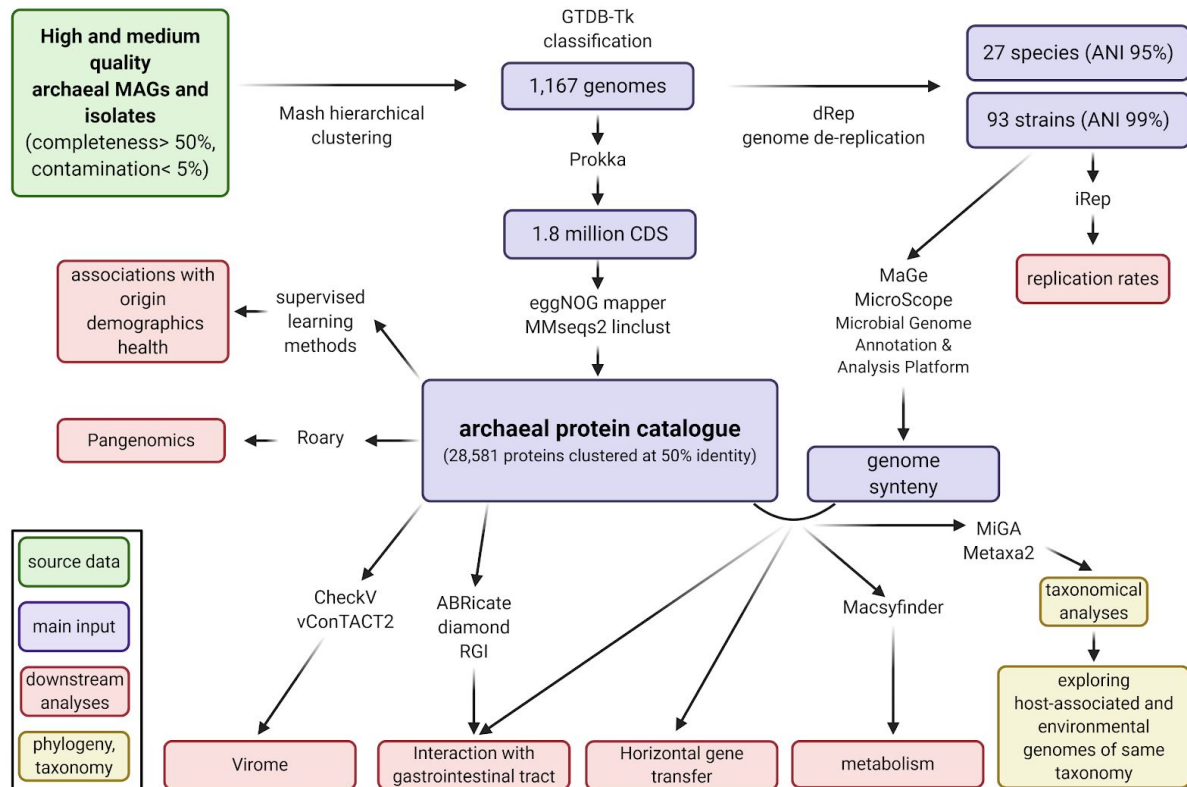
	Zeller et al. 2014; Nielsen et al. 2014; Franzosa et al. 2014; Gevers et al. 2014; Youngster et al. 2014; Khelaifia et al. 2014b; Bäckhed et al. 2015; Bengtsson-Palme et al. 2015; Cleary et al. 2015; Zeevi et al. 2015; Zhang et al. 2015; Feng et al. 2015; Obregon-Tito et al. 2015; Sankaranarayanan et al. 2015; Rampelli et al. 2015; Vogtmann et al. 2016; Liu et al. 2016; Pehrsson et al. 2016; Palleja et al. 2016; Vincent et al. 2016; Schirmer et al. 2016; Vatanen et al. 2016; Xie et al. 2016; Borrel et al. 2017; Asnicar et al. 2017; Smits et al. 2017; Wen et al. 2017; Yu et al. 2017; Gu et al. 2017; He et al. 2017; Bedarf et al. 2017; Costea et al. 2017; Lloyd-Price et al. 2017; Loomba et al. 2017; Petersen et al. 2017; Wu et al. 2017; Ye et al. 2018; Mehta et al. 2018; Pasolli et al. 2019	
<b>Software and Algorithms</b>		
ABRicate		<a href="https://github.com/tseemann/abricate">https://github.com/tseemann/abricate</a>
biorender		<a href="https://biorender.com/">https://biorender.com/</a>
bowtie2	Langmead and Salzberg 2012	<a href="https://github.com/BenLangmead/bowtie2">https://github.com/BenLangmead/bowtie2</a>
Calypso	Zakrzewski et al. 2017	<a href="http://cgenome.net:8080/calypso-8.84/faces/uploadFiles.xhtml">http://cgenome.net:8080/calypso-8.84/faces/uploadFiles.xhtml</a>
CheckM	Parks et al. 2015	<a href="https://github.com/Ecogenomics/CheckM">https://github.com/Ecogenomics/CheckM</a>
CheckV	Nayfach, Camargo et al. 2020	<a href="https://bitbucket.org/berkeleylab/checkv">https://bitbucket.org/berkeleylab/checkv</a>
DIAMOND	Buchfink, Xie, and Huson 2015	<a href="http://www.diamondsearch.org/index.php">http://www.diamondsearch.org/index.php</a>
dRep	Olm et al. 2017	<a href="https://drep.readthedocs.io/en/latest/">https://drep.readthedocs.io/en/latest/</a>
eggNOG database	Huerta-Cepas et al. 2019	<a href="http://eggnog5.embl.de/#/app/home">http://eggnog5.embl.de/#/app/home</a>
eggNOG mapper	Huerta-Cepas et al. 2017	<a href="https://github.com/eggnogdb/eggnog-mapper">https://github.com/eggnogdb/eggnog-mapper</a>



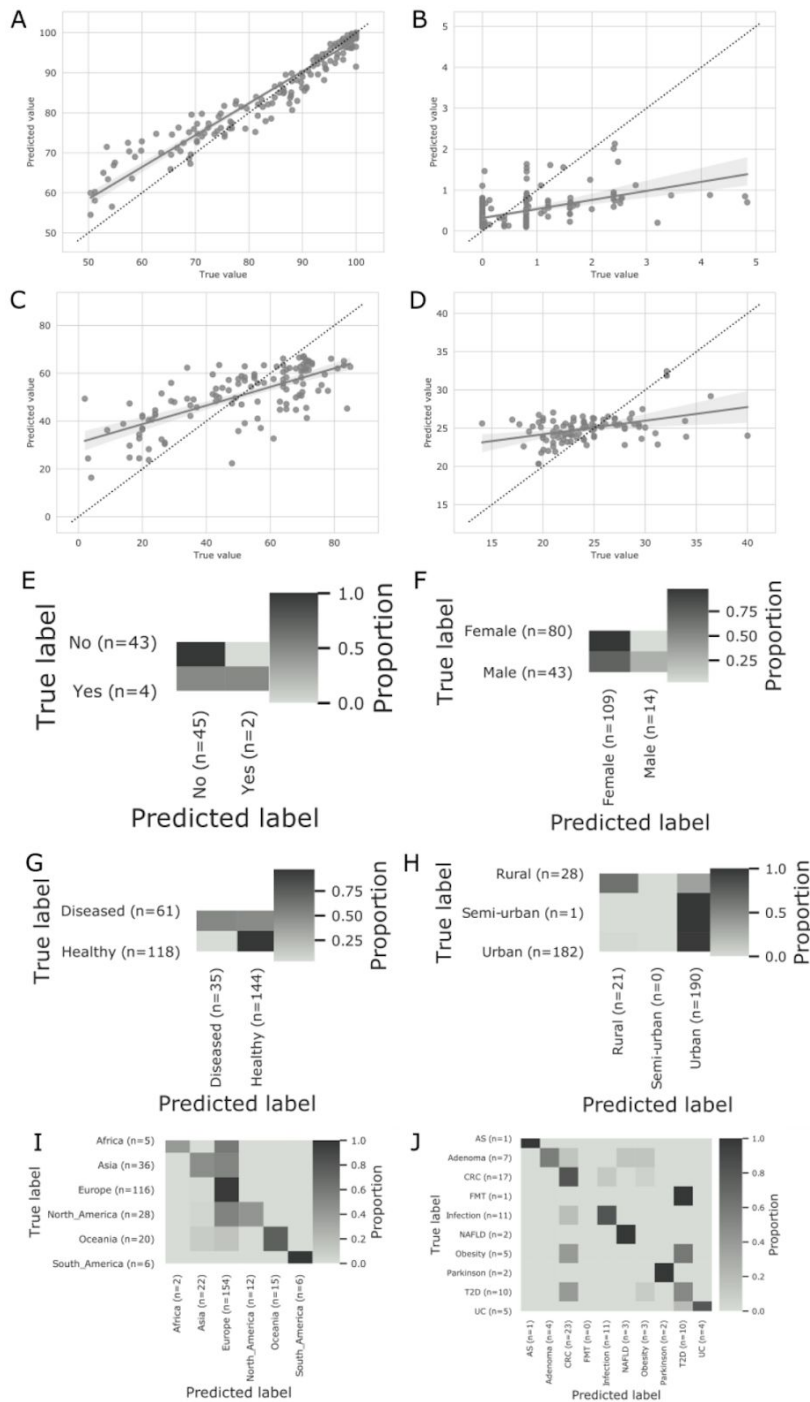
fastANI	Jain et al. 2018	<a href="https://github.com/ParBLISS/FastANI">https://github.com/ParBLISS/FastANI</a>
Genome Taxonomy Database Toolkit (GTDB-Tk)	Chaumeil et al. 2020	<a href="https://github.com/GenomeTaxonomy/GTDBTk">https://github.com/GenomeTaxonomy/GTDBTk</a>
inkscape		
iRep	Brown et al. 2016	<a href="https://github.com/christophertbrown/iRep">https://github.com/christophertbrown/iRep</a>
iTOL tool (Interactive Tree Of Life)	Letunic and Bork 2019	<a href="https://itol.embl.de/">https://itol.embl.de/</a>
Mash	Ondov et al. 2016	<a href="https://mash.readthedocs.io/en/latest/index.html">https://mash.readthedocs.io/en/latest/index.html</a>
MaGe MicroScope (Microbial Genome Annotation & Analysis Platform)	Vallenet et al. 2009	<a href="https://mage.genoscope.cns.fr/microscope/home/index.php">https://mage.genoscope.cns.fr/microscope/home/index.php</a>
Macsyfinder	Abby et al. 2014	<a href="https://github.com/gem-pasteur/macsyfinder">https://github.com/gem-pasteur/macsyfinder</a>
MEGA X: Molecular Evolutionary Genetics Analysis, v. 10.1.8	Kumar et al. 2018	<a href="https://www.megasoftware.net/">https://www.megasoftware.net/</a>
Metaxa2	Bengtsson-Palme et al. 2015	<a href="https://microbiology.se/software/metaxa2/">https://microbiology.se/software/metaxa2/</a>
Microbial Genomes Atlas (MiGA)	Rodriguez-R et al. 2018	<a href="http://microbial-genomes.org/">http://microbial-genomes.org/</a>
MMseqs2	Steinegger and Söding 2017	<a href="https://github.com/soedinglab/MMseqs2">https://github.com/soedinglab/MMseqs2</a>
Prokka	Seemann 2014	<a href="https://github.com/tseemann/prokka">https://github.com/tseemann/prokka</a>
Qiime2	Bolyen et al. 2018	<a href="https://github.com/qiime2/qiime2">https://github.com/qiime2/qiime2</a>
RAWGraphs		<a href="https://rawgraphs.io/">https://rawgraphs.io/</a>
Resistance Gene Identifier (RGI)	Jia et al. 2017; Alcock et al. 2020	<a href="https://github.com/arpcard/rgi">https://github.com/arpcard/rgi</a>
Roary	Page et al. 2015	<a href="https://sanger-pathogens.github.io/Roary/">https://sanger-pathogens.github.io/Roary/</a>
R-Studio		

samtools	Li et al. 2009	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a> <a href="http://www.htslib.org/">http://www.htslib.org/</a>
vConTACT2	Bin Jang et al. 2019	<a href="https://bitbucket.org/MAVERICLab/vcontact2/src">https://bitbucket.org/MAVERICLab/vcontact2/src</a>
pheatmap (R package)	<a href="https://cran.r-project.org/web/packages/pheatmap/index.html">https://cran.r-project.org/web/packages/pheatmap/index.html</a>	<a href="https://cran.r-project.org/web/packages/pheatmap/index.html">https://cran.r-project.org/web/packages/pheatmap/index.html</a>
ggplot2 (R package)	<a href="https://cran.r-project.org/web/packages/ggplot2/index.html">https://cran.r-project.org/web/packages/ggplot2/index.html</a>	<a href="https://cran.r-project.org/web/packages/ggplot2/index.html">https://cran.r-project.org/web/packages/ggplot2/index.html</a>
ggpubr (R package)	<a href="https://cran.r-project.org/web/packages/ggpubr/index.html">https://cran.r-project.org/web/packages/ggpubr/index.html</a>	<a href="https://cran.r-project.org/web/packages/ggpubr/index.html">https://cran.r-project.org/web/packages/ggpubr/index.html</a>
<b>Other</b>		
ARG-ANNOT	Gupta et al. 2014	
CARD	Jia et al. 2017	<a href="https://card.mcmaster.ca/">https://card.mcmaster.ca/</a>
EcOH	Ingle et al. 2016	<a href="https://microreact.org/project/Ny5Gg4Wg-">https://microreact.org/project/Ny5Gg4Wg-</a>
Human Gut Virome Database (GVD)	Gregory et al. 2020	
MEGARES 2.00	Doster et al. 2020	<a href="https://megares.meglab.org/">https://megares.meglab.org/</a>
NCBI AMRFinderPlus	Feldgarden et al. 2019	<a href="https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/">https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/</a>
PlasmidFinder	Carattoli et al. 2014	<a href="https://cge.cbs.dtu.dk/services/PlasmidFinder/">https://cge.cbs.dtu.dk/services/PlasmidFinder/</a>
Resfinder	Zankari et al. 2012	<a href="https://cge.cbs.dtu.dk/services/ResFinder/">https://cge.cbs.dtu.dk/services/ResFinder/</a>
SILVA rRNA database	Quast et al. 2013	
VFDB	Chen et al. 2005; Liu et al. 2019	<a href="http://www.mgc.ac.cn/VFs/main.htm">http://www.mgc.ac.cn/VFs/main.htm</a>
Viral RefSeq database (version 97)	-	<a href="https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi">https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi</a>

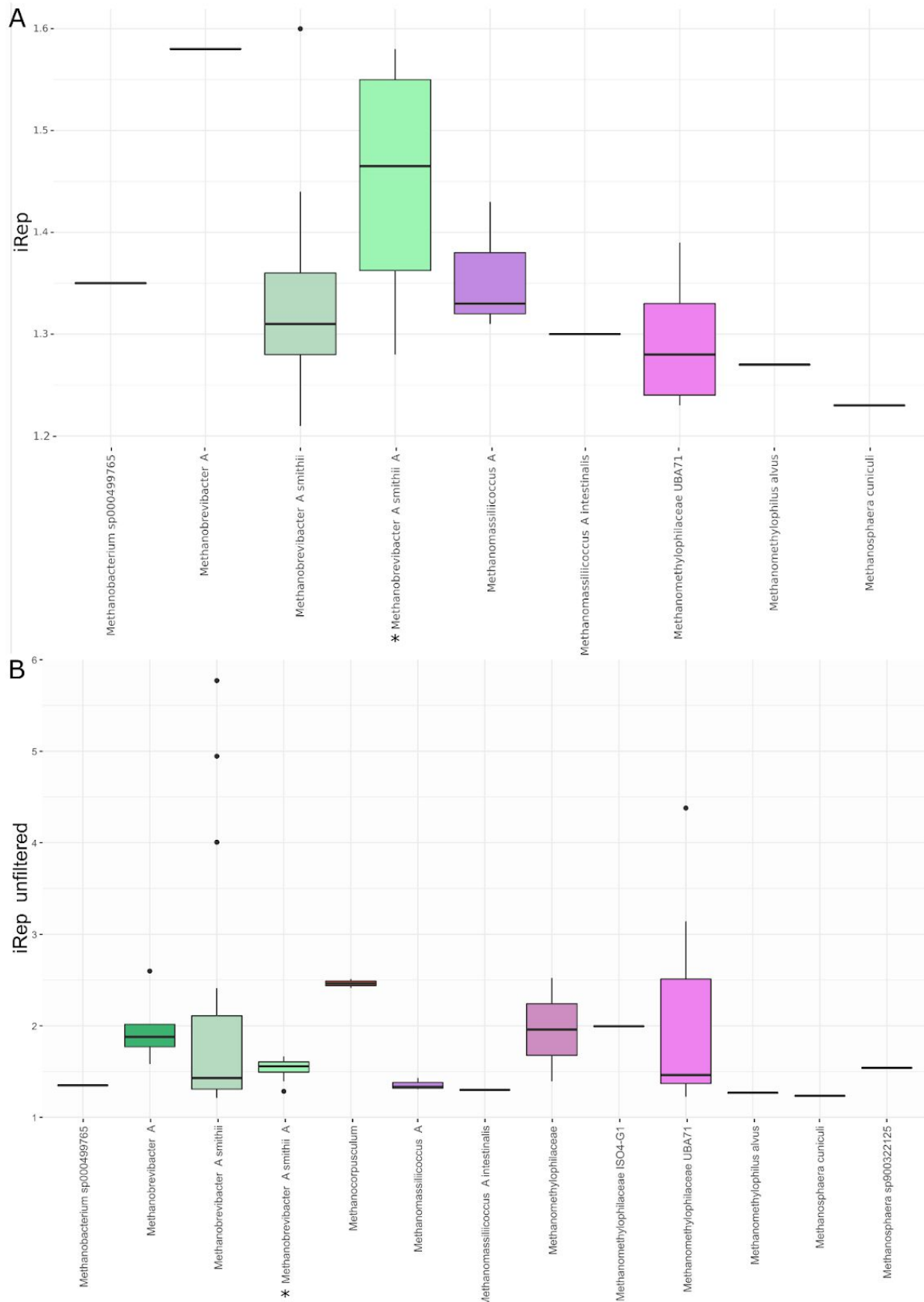
## SUPPLEMENTAL INFORMATION



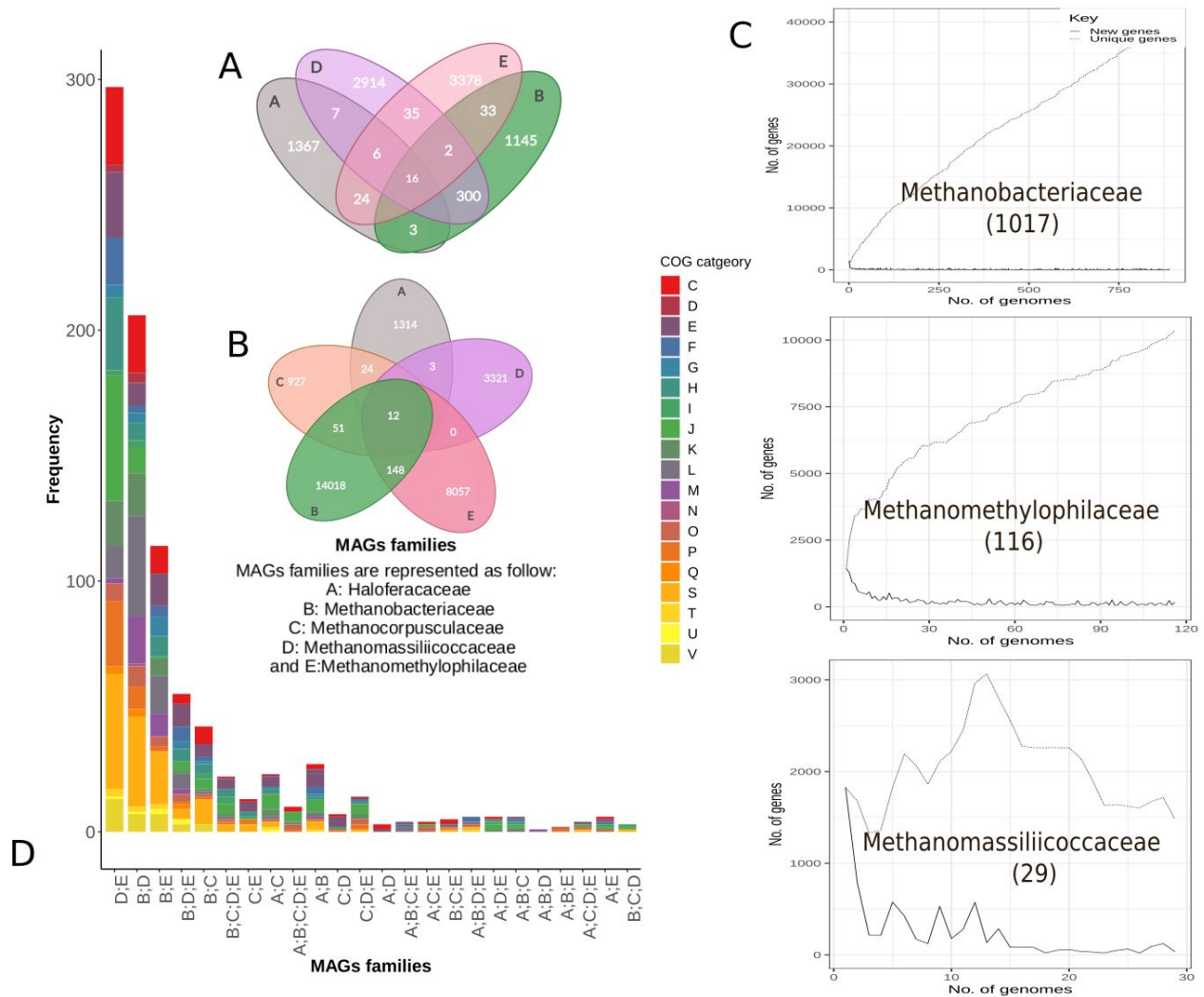
**Supplementary Fig. 1: Methodology.** Flow chart covering the major analysis steps of the study. Colored boxes show the source data (green), main input for the analysis (magenta), downstream analysis (red) and the phylogenetic and taxonomic analysis of the presented data set (yellow). Different steps are connected by arrows highlighting a selection of used bioinformatic tools for each step.



**Supplementary Fig. 2: Predicting metadata values as a function of protein composition by supervised learning methods.** Scatterplots and heatmaps of metadata predictions based on the unified archaeal MAG protein catalogue. Completeness (A), contamination (B), age (C), BMI (D), antibiotics (E), gender (F), health status (G), lifestyle (H), continent (I), and name of disease (J).

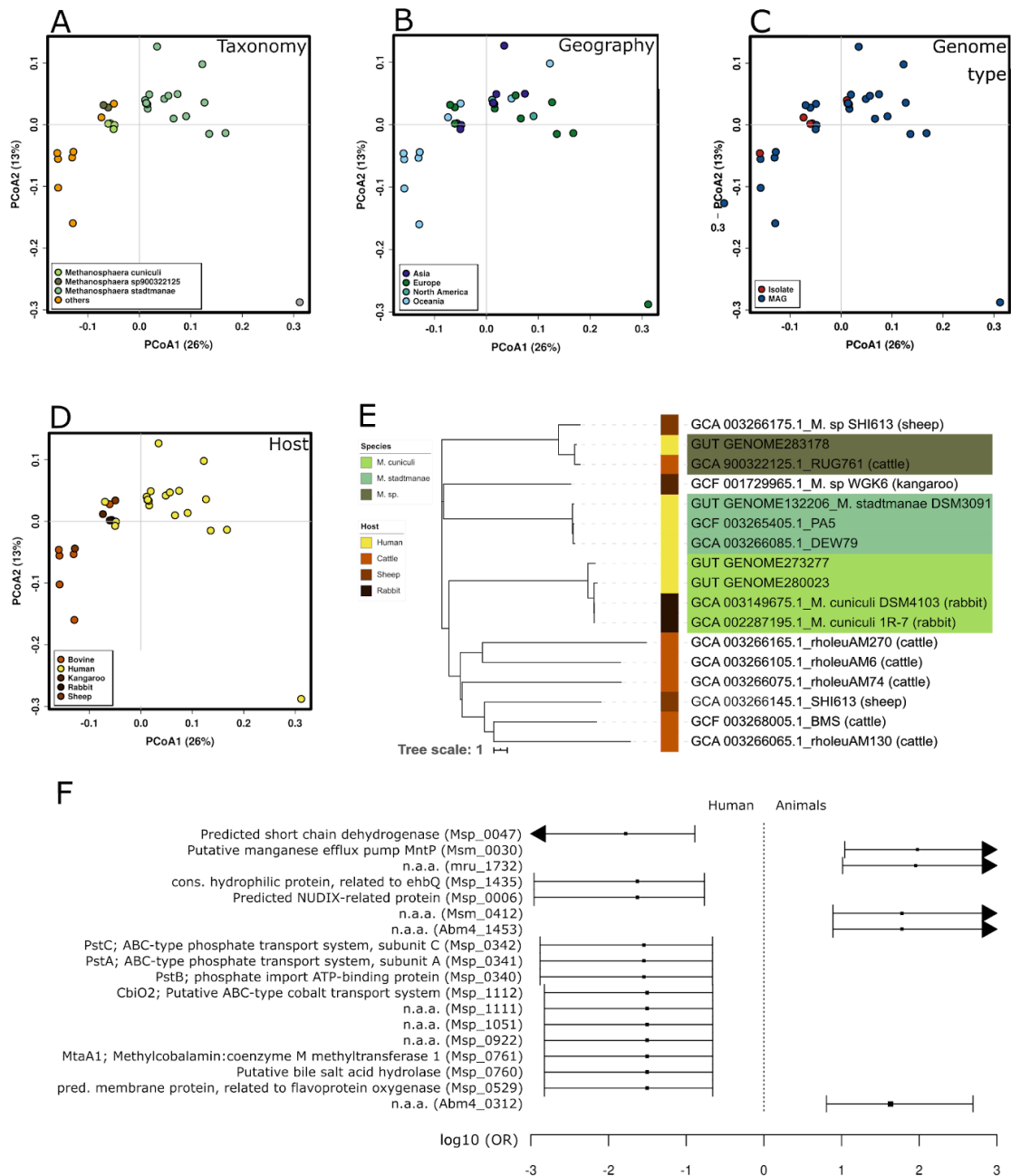


**Supplementary Fig. 3. Replication rates of archaeal genomes from the human gut. A:** iRep analysis of 26 genomes, showing the active replication of different GTDB-tk classified genomes. **B:** unfiltered iRep analysis on 69 genomes, showing the active replication of different GTDB-tk classified genomes. \**Candidatus* Methanobrevibacter intestinalis

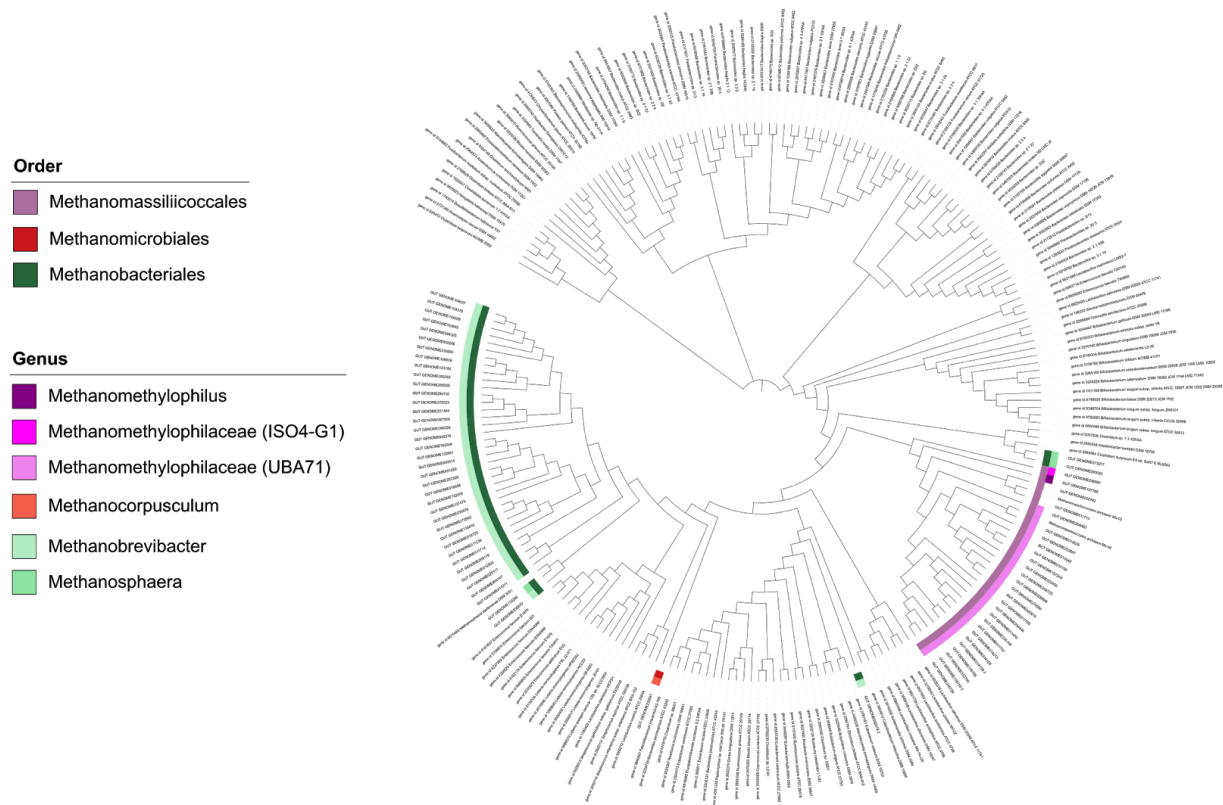


**Supplementary Fig. 4: Venn diagrams summarizing the number of shared proteins between the different archaeal families A) for complete genomes only B) for all 1167 genomes.** Venn diagrams were done by creately (<https://app.creately.com>). **C)** Pan-genome analysis per family where over 10 members per family exist. The three archaeal families Methanomassiliicoccaceae ( $\alpha=0.69$ ), Methanobacteriaceae ( $\alpha=0.71$ ) and Methanomethylophilaceae ( $\alpha=0.86$ ) have an open pangenome. **D)** Bar plot representing the frequency of COG categories of the proteins shared between the 5 archaeal MAGs taxonomic families (CELLULAR PROCESSES AND SIGNALING: [D] Cell cycle control, cell division, chromosome partitioning, [M] Cell wall/membrane/envelope biogenesis, [N] Cell motility, [O] Post-translational modification, protein turnover, and chaperones, [T] Signal transduction mechanisms, [U] Intracellular trafficking, secretion, and vesicular transport, [V] Defense mechanisms. INFORMATION STORAGE AND PROCESSING: [J] Translation, ribosomal structure and biogenesis, [K] Transcription, [L] Replication, recombination, and repair. METABOLISM: [C] Energy production and conversion, [E] Amino acid transport and metabolism, [F] Nucleotide transport and metabolism, [G] Carbohydrate transport and metabolism, [H] Coenzyme transport and metabolism, [I] Lipid transport and metabolism, [P] Inorganic ion transport and metabolism, [Q] Secondary metabolites biosynthesis, transport, and catabolism. POORLY CHARACTERIZED: [S] Function unknown) Supplementary Table 2.

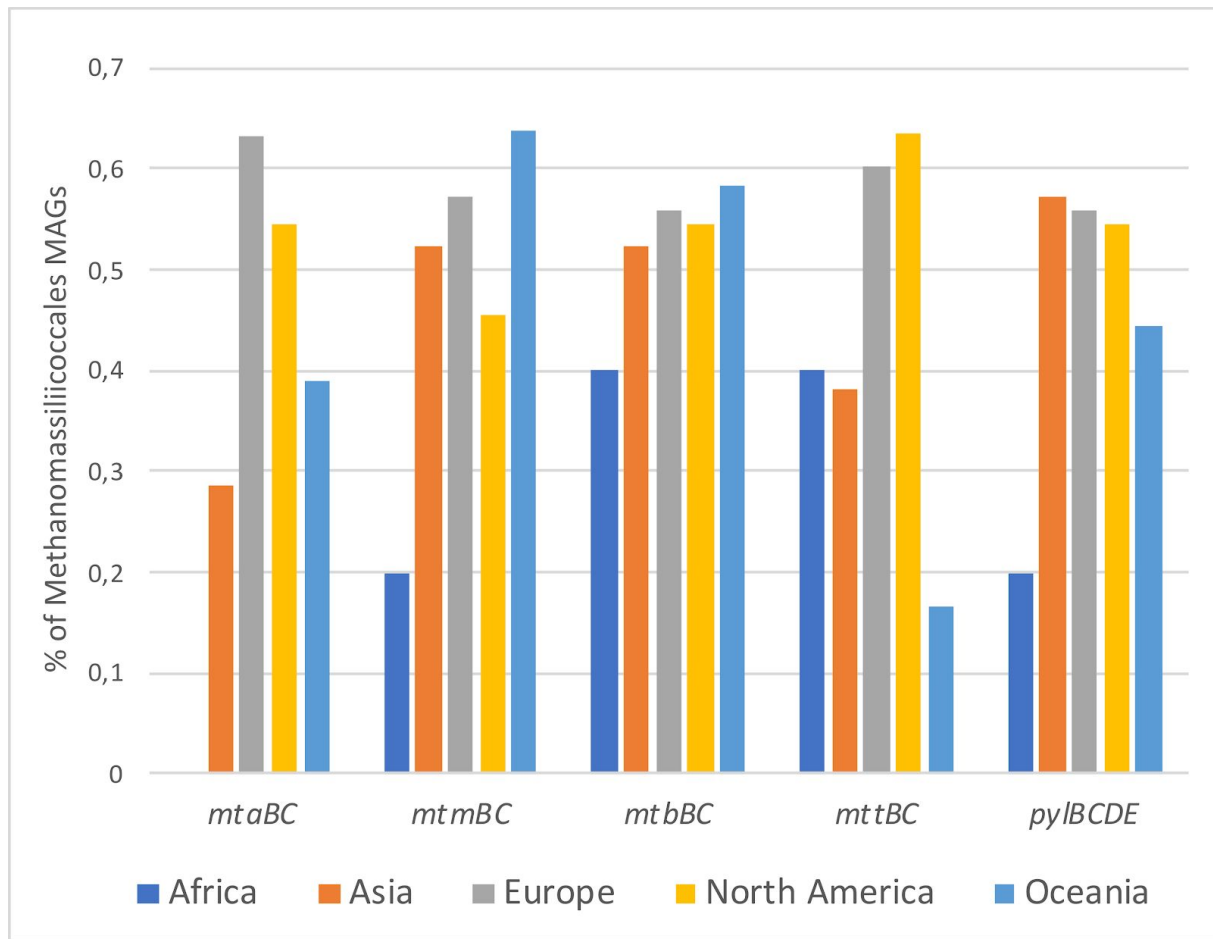




**Supplementary Fig. 5: Profiles of human-associated *Methanosphaera* genomes** (Supplementary Table 9). For comparison, eleven genomes from animal-associated *Methanosphaera* were included. PCoA plots of the genomic profiles according to taxonomy (A), geography (B), genome type (C), and host (D) and phylogenetic tree of the genus *Methanosphaera* with human- and animal-associated representatives (E). Human-associated species are highlighted in green colors. Colored bar displays the origin: human (yellow) and animals (shades of brown). (F): Forest plot showing the outcome of the Wilcoxon rank test comparison of genomes from humans vs. animals (only proteins with FDR<0.05 are shown), bar displays the odds ratio (OR) (Supplementary Table 9).



**Supplementary Fig. 6: Archaeal bile salt hydrolase genes (this study) integrated in the bacterial tree of BSHs (Song et al. 2019). Archaeal genes are highlighted by the colored ring, indicating the respective taxonomic affiliation.**



**Supplementary Fig. 7: Geographic distribution of methyl-compound utilization capacity by Methanomassiliicoccales representatives.** The presence of *mtaBC*, *mtmBC*, *mtbBC* and *mttBC* genes needed for methanol, monomethylamine, dimethylamine and, trimethylamine utilization, respectively, as well as *pylBCDE* genes responsible for the biosynthesis of pyrrolysine (an amino-acid specifically present in methylamine methyltransferases) was searched in all Methanomassiliicoccales MAGs. Methanomassiliicoccales were separated according to the geographic location (continents) of their host, and the percentage of them having the above mentioned genes is displayed.

**Supplementary Table 1.** Detailed description of all MAGs and isolates, which were analyzed in this study. This table includes the list of all 1167 archaeal genomes, a strain list (99% ANI distance), a species list (95% ANI distance) and all available metadata. These data sets were the basis for Fig. 1, Fig.3 and Supplementary Fig. 2,

**Supplementary Table 2.** Unified human archaeal protein catalogue based on clustering at 50% identity of all genome CDS, with associated lineage and genome information and a summary of the number of genes shared per archaeal family. Basis for Fig 3D and Supplementary Fig. 4 A,B and D.

**Supplementary Table 3.** Metadata information on the human and animal *Methanobrevibacter* analysis dataset, and protein catalogue information for those genomes (cut-off: presence in at least 10 genomes). Basis for Fig. 4.

**Supplementary Table 4.** Comparison of *M. smithii* vs. *M. smithii\_A*, based on genome size and discriminative proteins (Wilcoxon top 25). Basis for information provided in chapter "The *Methanobrevibacter smithii* clade splits into two separate species".

**Supplementary Table 5.** Identified proviruses and clustering into viral populations including quality summaries and host where they were identified from. Basis for Fig. 5.

**Supplementary Table 6.** Proportions of non-archaeal proteins for Methanosphaera, Methanobrevibacter and Methanomassiliicoccales, organized for human and animal-derived genomes, and isolates. Basis for Fig. 6.

**Supplementary Table 7. Comparison to environmental archaea.** i) Origin of 16S rRNA genes. Basis for Fig 7A. ii) Subset of the environmental archaeal genomes identified by Parks et al. 2017. Basis for Fig. 7 B,C,D and E.

**Supplementary Table 8.** Presence/absence of enzymatic complexes and pathways in the 27 species reported in this study. The presence/absence of the complexes and pathways is inferred from the detection of all enzymes that compose them. Three putative species represented by only 1 MAG having a poor quality were not included in this analysis. "+" indicates the presence of the complex or the pathways and "-" indicates its absence. MCR, Methyl-coenzyme M reductase; MTR, N5-Methyltetrahydromethanopterin: coenzyme M methyltransferase; H4MPT mWL, H4MPT version of the methyl branch of the Wood-Ljungdhal pathway; MtaBC, methanol:corrinoide protein methyltransferase and associated corrinoide-binding protein; MtmBC, monomethylamine:corrinoide protein methyltransferase and associated corrinoide-binding protein; MtbBC, dimethylamine:corrinoide protein methyltransferase and associated corrinoide-binding protein; MtsAB, trimethylamine:corrinoide protein methyltransferase and associated corrinoide-binding protein; MthsAB, Methylthiol:coenzyme M methyltransferase and associated corrinoide-binding protein; PylBCDE, pyl-system involved in pyrrolysine biosynthesis; FdhAB, formate dehydrogenase; "Alcohol utilisation" is based on the presence of Adh (alcohol dehydrogenase), Aldh (Aldehyde dehydrogenase) and Fno (NADPH-dependent F420 reductase); Fpo-like, Ferredoxin:heterodisulfide oxidoreductase; Eha, Energy-conserving hydrogenase A; Ehb, Energy-conserving hydrogenase B; Ech, Energy-conserving hydrogenase C; HdrBC, Heterodisulfide reductase catalytic subunit and C subunit; MvhADG, F420-nonreducing [Ni-Fe]-hydrogenase; FrhABG, F420H2-reducing hydrogenase; \* in "Genome name" column indicates names of candidatus species and/or genera proposed in this study. Basis for Fig. 8B.

**Supplementary Table 9.** Metadata information on the human and animal *Methanosphaera* analysis dataset, protein catalogue information for those genomes, and Wilcoxon rank test comparison. Basis for Supplementary Fig. 5.

**Supplementary Material 1.** Alignment of the mcrA genes of *M. smithii* and *M. smithii\_A*, fasta file.

**Supplementary Material 2.** Environmental and human archaea 16S rRNA genes, fasta file, basis for Fig. 7A.

**Supplementary Material 3.** Interactive Krona output, basis for Fig. 6.

## REFERENCES

- Abby, Sophie S., Bertrand Néron, Hervé Ménager, Marie Touchon, and Eduardo P. C. Rocha. 2014. "MacSyFinder: A Program to Mine Genomes for Molecular Systems with an Application to CRISPR-Cas Systems." *PLOS ONE* 9 (10): e110726. <https://doi.org/10.1371/journal.pone.0110726>.
- Alcock, Brian P., Amogelang R. Raphenya, Tammy T. Y. Lau, Kara K. Tsang, Mégane Bouchard, Arman Edalatmand, William Huynh, et al. 2020. "CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database." *Nucleic Acids Research* 48 (D1): D517–25. <https://doi.org/10.1093/nar/gkz935>.
- Almeida, Alexandre, Alex L. Mitchell, Miguel Boland, Samuel C. Forster, Gregory B. Gloor, Aleksandra Tarkowska, Trevor D. Lawley, and Robert D. Finn. 2019. "A New Genomic Blueprint of the Human Gut Microbiota." *Nature* 568 (7753): 499–504. <https://doi.org/10.1038/s41586-019-0965-1>.
- Almeida, Alexandre, Stephen Nayfach, Miguel Boland, Francesco Strozzi, Martin Beracochea, Zhou Jason Shi, Katherine S. Pollard, et al. 2020. "A Unified Catalog of 204,938 Reference Genomes from the Human Gut Microbiome." *Nature Biotechnology*, July, 1–10. <https://doi.org/10.1038/s41587-020-0603-3>.
- Asnicar, Francesco, Serena Manara, Moreno Zolfo, Duy Tin Truong, Matthias Scholz, Federica Armanini, Pamela Ferretti, et al. 2017. "Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling." *MSystems* 2 (1). <https://doi.org/10.1128/mSystems.00164-16>.
- Bäckhed, Fredrik, Josefine Roswall, Yangqing Peng, Qiang Feng, Huijue Jia, Petia Kovatcheva-Datchary, Yin Li, et al. 2015. "Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life." *Cell Host & Microbe* 17 (5): 690–703. <https://doi.org/10.1016/j.chom.2015.04.004>.
- Bang, Corinna, Claudia Ehlers, Alvaro Orell, Daniela Prasse, Marlene Spinner, Stanislav N. Gorb, Sonja-Verena Albers, and Ruth A. Schmitz. 2014. "Biofilm Formation of Mucosa-Associated Methanoarchaeal Strains." *Frontiers in Microbiology* 5. <https://doi.org/10.3389/fmicb.2014.00353>.
- Bang, Corinna, and Ruth A. Schmitz. 2018. "Archaea: Forgotten Players in the Microbiome." *Emerging Topics in Life Sciences* 2 (4): 459–68. <https://doi.org/10.1042/ETLS20180035>.
- Bedarf, J. R., F. Hildebrand, L. P. Coelho, S. Sunagawa, M. Bahram, F. Goeser, P. Bork, and U. Wüllner. 2017. "Functional Implications of Microbial and Viral Gut Metagenome Changes in Early Stage L-DOPA-Naïve Parkinson's Disease Patients." *Genome Medicine* 9 (1): 39. <https://doi.org/10.1186/s13073-017-0428-y>.
- Bengtsson-Palme, Johan, Martin Angelin, Mikael Huss, Sanela Kjellqvist, Erik Kristiansson, Helena Palmgren, D. G. Joakim Larsson, and Anders Johansson. 2015. "The Human Gut Microbiome as a Transporter of Antibiotic Resistance Genes between Continents." *Antimicrobial Agents and Chemotherapy* 59 (10): 6551–60. <https://doi.org/10.1128/AAC.00933-15>.
- Bengtsson-Palme, Johan, Martin Hartmann, Karl Martin Eriksson, Chandan Pal, Kaisa Thorell, Dan Göran Joakim Larsson, and Rolf Henrik Nilsson. 2015. "Metaxa2: Improved Identification and Taxonomic Classification of Small and Large Subunit rRNA in Metagenomic Data." *Molecular Ecology Resources* 15 (6): 1403–14. <https://doi.org/10.1111/1755-0998.12399>.
- Biavati, B., M. Vasta, and J. G. Ferry. 1988. "Isolation and Characterization of 'Methanosphaera Cuniculi' Sp. Nov." *Applied and Environmental Microbiology* 54 (3): 768–71.
- Bin Jang, Ho, Benjamin Bolduc, Olivier Zabolocki, Jens H. Kuhn, Simon Roux, Evelien M. Adriaenssens, J. Rodney Brister, et al. 2019. "Taxonomic Assignment of Uncultivated Prokaryotic Virus Genomes Is Enabled by Gene-Sharing Networks." *Nature Biotechnology* 37 (6): 632–39. <https://doi.org/10.1038/s41587-019-0100-8>.
- Bobay, Louis-Marie, and Howard Ochman. 2018. "Biological Species in the Viral World." *Proceedings of the National Academy of Sciences* 115 (23): 6040–45. <https://doi.org/10.1073/pnas.1717593115>.

- Bokulich, Nicholas A., Matthew R. Dillon, Evan Bolyen, Benjamin D. Kaehler, Gavin A. Huttley, and J. Gregory Caporaso. 2018. "Q2-Sample-Classifer: Machine-Learning Tools for Microbiome Classification and Regression." *Journal of Open Source Software* 3 (30): 934. <https://doi.org/10.21105/joss.00934>.
- Bolyen, Evan, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, et al. 2018. "QIIME 2: Reproducible, Interactive, Scalable, and Extensible Microbiome Data Science." e27295v2. PeerJ Inc. <https://doi.org/10.7287/peerj.preprints.27295v2>.
- Borrel, Guillaume, Jean-François Brugère, Simonetta Gribaldo, Ruth A. Schmitz, and Christine Moissl-Eichinger. 2020. "The Host-Associated Archaeome." *Nature Reviews Microbiology* 18 (11): 622–36. <https://doi.org/10.1038/s41579-020-0407-y>.
- Borrel, Guillaume, Hugh M. B. Harris, Nicolas Parisot, Nadia Gaci, William Tottey, Agnès Mihajlovski, Jennifer Deane, et al. 2013. "Genome Sequence of 'Candidatus Methanomassiliicoccus Intestinalis' Issoire-Mx1, a Third Thermoplasmatales-Related Methanogenic Archaeon from Human Feces." *Genome Announcements* 1 (4). <https://doi.org/10.1128/genomeA.00453-13>.
- Borrel, Guillaume, Hugh M. B. Harris, William Tottey, Agnès Mihajlovski, Nicolas Parisot, Eric Peyretailade, Pierre Peyret, Simonetta Gribaldo, Paul W. O'Toole, and Jean-François Brugère. 2012. "Genome Sequence of 'Candidatus Methanomethylophilus Alvus' Mx1201, a Methanogenic Archaeon from the Human Gut Belonging to a Seventh Order of Methanogens." *Journal of Bacteriology* 194 (24): 6944–45. <https://doi.org/10.1128/JB.01867-12>.
- Borrel, Guillaume, Angela McCann, Jennifer Deane, Marta C. Neto, Denise B. Lynch, Jean-François Brugère, and Paul W. O'Toole. 2017. "Genomics and Metagenomics of Trimethylamine-Utilizing Archaea in the Human Gut Microbiome." *The ISME Journal* 11 (9): 2059–74. <https://doi.org/10.1038/ismej.2017.72>.
- Brown, Christopher T., Matthew R. Olm, Brian C. Thomas, and Jillian F. Banfield. 2016. "Measurement of Bacterial Replication Rates in Microbial Communities." *Nature Biotechnology* 34 (12): 1256–63. <https://doi.org/10.1038/nbt.3704>.
- Brugère, Jean-François, John F. Atkins, Paul W. O'Toole, and Guillaume Borrel. 2018. "Pyrrolysine in Archaea: A 22nd Amino Acid Encoded through a Genetic Code Expansion." *Emerging Topics in Life Sciences* 2 (4): 607–18. <https://doi.org/10.1042/ETLS20180094>.
- Brugère, Jean-François, Guillaume Borrel, Nadia Gaci, William Tottey, Paul W. O'Toole, and Corinne Malpuech-Brugère. 2014. "Archaeobiotics." *Gut Microbes* 5 (1): 5–10. <https://doi.org/10.4161/gmic.26749>.
- Brum, Jennifer R., J. Cesar Ignacio-Espinoza, Simon Roux, Guilhem Doulier, Silvia G. Acinas, Adriana Alberti, Samuel Chaffron, et al. 2015. "Patterns and Ecological Drivers of Ocean Viral Communities." *Science* 348 (6237). <https://doi.org/10.1126/science.1261498>.
- Buchfink, Benjamin, Chao Xie, and Daniel H. Huson. 2015. "Fast and Sensitive Protein Alignment Using DIAMOND." *Nature Methods* 12 (1): 59–60. <https://doi.org/10.1038/nmeth.3176>.
- Carattoli, Alessandra, Ea Zankari, Aurora García-Fernández, Mette Voldby Larsen, Ole Lund, Laura Villa, Frank Møller Aarestrup, and Henrik Hasman. 2014. "In Silico Detection and Typing of Plasmids Using PlasmidFinder and Plasmid Multilocus Sequence Typing." *Antimicrobial Agents and Chemotherapy* 58 (7): 3895–3903. <https://doi.org/10.1128/AAC.02412-14>.
- Carver, Tim J., Kim M. Rutherford, Matthew Berriman, Marie-Adele Rajandream, Barclay G. Barrell, and Julian Parkhill. 2005. "ACT: The Artemis Comparison Tool." *Bioinformatics* 21 (16): 3422–23. <https://doi.org/10.1093/bioinformatics/bti553>.
- Chaumeil, Pierre-Alain, Aaron J. Mussig, Philip Hugenholtz, and Donovan H. Parks. 2020. "GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database." *Bioinformatics* 36 (6): 1925–27. <https://doi.org/10.1093/bioinformatics/btz848>.
- Chellapandi, P, M Bharathi, C Sangavai, and R Prathiviraj. 2018. "Methanobacterium Formicicum as a Target Rumen Methanogen for the Development of New Methane Mitigation Interventions: A Review." *Veterinary and Animal Science* 6 (December): 86–94. <https://doi.org/10.1016/j.vas.2018.09.001>.



- Chen, I.-Min A., Ken Chu, Krishna Palaniappan, Manoj Pillay, Anna Ratner, Jinghua Huang, Marcel Huntemann, et al. 2019. "IMG/M v.5.0: An Integrated Data Management and Comparative Analysis System for Microbial Genomes and Microbiomes." *Nucleic Acids Research* 47 (D1): D666–77. <https://doi.org/10.1093/nar/gky901>.
- Chen, Lihong, Jian Yang, Jun Yu, Zhijian Yao, Lilian Sun, Yan Shen, and Qi Jin. 2005. "VFDB: A Reference Database for Bacterial Virulence Factors." *Nucleic Acids Research* 33 (suppl\_1): D325–28. <https://doi.org/10.1093/nar/gki008>.
- Cleary, Brian, Ilana Lauren Brito, Katherine Huang, Dirk Gevers, Terrance Shea, Sarah Young, and Eric J. Alm. 2015. "Detection of Low-Abundance Bacterial Strains in Metagenomic Datasets by Eigengene Partitioning." *Nature Biotechnology* 33 (10): 1053–60. <https://doi.org/10.1038/nbt.3329>.
- Clemente, Jose C., Erica C. Pehrsson, Martin J. Blaser, Kuldeep Sandhu, Zhan Gao, Bin Wang, Magda Magris, et al. 2015. "The Microbiome of Uncontacted Amerindians." *Science Advances* 1 (3): e1500183. <https://doi.org/10.1126/sciadv.1500183>.
- Costea, Paul I, Luis Pedro Coelho, Shinichi Sunagawa, Robin Munch, Jaime Huerta-Cepas, Kristoffer Forslund, Falk Hildebrand, Almagul Kushugulova, Georg Zeller, and Peer Bork. 2017. "Subspecies in the Global Human Gut Microbiome." *Molecular Systems Biology* 13 (12): 960. <https://doi.org/10.15252/msb.20177589>.
- Cui, Heng-Lin, Dilbr Tohty, Pei-Jin Zhou, and Shuang-Jiang Liu. 2006. "Halorubrum Lipolyticum Sp. Nov. and Halorubrum Aidingense Sp. Nov., Isolated from Two Salt Lakes in Xin-Jiang, China." *International Journal of Systematic and Evolutionary Microbiology*, 56 (7): 1631–34. <https://doi.org/10.1099/ijs.0.64305-0>.
- Doster, Enrique, Steven M. Lakin, Christopher J. Dean, Cory Wolfe, Jared G. Young, Christina Boucher, Keith E. Belk, Noelle R. Noyes, and Paul S. Morley. 2020. "MEGARes 2.0: A Database for Classification of Antimicrobial Drug, Biocide and Metal Resistance Determinants in Metagenomic Sequence Data." *Nucleic Acids Research* 48 (D1): D561–69. <https://doi.org/10.1093/nar/gkz1010>.
- Dridi, Bédis, Marie-Laure Fardeau, Bernard Ollivier, Didier Raoult, and Michel Drancourt. 2012. "Methanomassiliicoccus Luminyensis Gen. Nov., Sp. Nov., a Methanogenic Archaeon Isolated from Human Faeces." *International Journal of Systematic and Evolutionary Microbiology*, 62 (Pt\_8): 1902–7. <https://doi.org/10.1099/ijs.0.033712-0>.
- Duhaime, Melissa B., Natalie Solonenko, Simon Roux, Nathan C. Verberkmoes, Antje Wichels, and Matthew B. Sullivan. 2017. "Comparative Omics and Trait Analyses of Marine Pseudoalteromonas Phages Advance the Phage OTU Concept." *Frontiers in Microbiology* 8. <https://doi.org/10.3389/fmicb.2017.01241>.
- Duhaime, Melissa B., and Matthew B. Sullivan. 2012. "Ocean Viruses: Rigorously Evaluating the Metagenomic Sample-to-Sequence Pipeline." *Virology*, Special issue: viruses of microbes, 434 (2): 181–86. <https://doi.org/10.1016/j.virol.2012.09.036>.
- Duvallet, Claire, Sean M. Gibbons, Thomas Gurry, Rafael A. Irizarry, and Eric J. Alm. 2017. "Meta-Analysis of Gut Microbiome Studies Identifies Disease-Specific and Shared Responses." *Nature Communications* 8 (1): 1784. <https://doi.org/10.1038/s41467-017-01973-8>.
- Faveri, M., L. F. H. Gonçalves, M. Feres, L. C. Figueiredo, L. A. Gouveia, J. A. Shibli, and M. P. A. Mayer. 2011. "Prevalence and Microbiological Diversity of Archaea in Peri-Implantitis Subjects by 16S Ribosomal RNA Clonal Analysis." *Journal of Periodontal Research* 46 (3): 338–44. <https://doi.org/10.1111/j.1600-0765.2011.01347.x>.
- Feldgarden, Michael, Vyacheslav Brover, Daniel H. Haft, Arjun B. Prasad, Douglas J. Slotta, Igor Tolstoy, Gregory H. Tyson, et al. 2019. "Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates." *Antimicrobial Agents and Chemotherapy* 63 (11). <https://doi.org/10.1128/AAC.00483-19>.
- Feng, Qiang, Suisha Liang, Huijue Jia, Andreas Stadlmayr, Longqing Tang, Zhou Lan, Dongya Zhang, et al. 2015. "Gut Microbiome Development along the Colorectal Adenoma–Carcinoma

- Sequence." *Nature Communications* 6 (1): 6528. <https://doi.org/10.1038/ncomms7528>.
- Franzosa, Eric A., Xochitl C. Morgan, Nicola Segata, Levi Waldron, Joshua Reyes, Ashlee M. Earl, Georgia Giannoukos, et al. 2014. "Relating the Metatranscriptome and Metagenome of the Human Gut." *Proceedings of the National Academy of Sciences* 111 (22): E2329–38. <https://doi.org/10.1073/pnas.1319284111>.
- Gaci, Nadia, Guillaume Borrel, William Tottey, Paul William O'Toole, and Jean-François Brugère. 2014. "Archaea and the Human Gut: New Beginning of an Old Story." *World Journal of Gastroenterology : WJG* 20 (43): 16062–78. <https://doi.org/10.3748/wjg.v20.i43.16062>.
- Gevers, Dirk, Subra Kugathasan, Lee A. Denson, Yoshiki Vázquez-Baeza, Will Van Treuren, Boyu Ren, Emma Schwager, et al. 2014. "The Treatment-Naive Microbiome in New-Onset Crohn's Disease." *Cell Host & Microbe* 15 (3): 382–92. <https://doi.org/10.1016/j.chom.2014.02.005>.
- Gregory, Ann C., Sergei A. Solonenko, J. Cesar Ignacio-Espinoza, Kurt LaButti, Alex Copeland, Sebastian Sudek, Ashley Maitland, et al. 2016. "Genomic Differentiation among Wild Cyanophages despite Widespread Horizontal Gene Transfer." *BMC Genomics* 17 (1): 930. <https://doi.org/10.1186/s12864-016-3286-x>.
- Gregory, Ann C., Olivier Zablocki, Ahmed A. Zayed, Allison Howell, Benjamin Bolduc, and Matthew B. Sullivan. 2020. "The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut." *Cell Host & Microbe*, August. <https://doi.org/10.1016/j.chom.2020.08.003>.
- Gu, Yanyun, Xiaokai Wang, Junhua Li, Yifei Zhang, Huanzi Zhong, Ruixin Liu, Dongya Zhang, et al. 2017. "Analyses of Gut Microbiota and Plasma Bile Acids Enable Stratification of Patients for Antidiabetic Treatment." *Nature Communications* 8 (1): 1785. <https://doi.org/10.1038/s41467-017-01682-2>.
- Gupta, Sushim Kumar, Babu Roshan Padmanabhan, Seydina M. Diene, Rafael Lopez-Rojas, Marie Kempf, Luce Landraud, and Jean-Marc Rolain. 2014. "ARG-ANNOT, a New Bioinformatic Tool to Discover Antibiotic Resistance Genes in Bacterial Genomes." *Antimicrobial Agents and Chemotherapy* 58 (1): 212–20. <https://doi.org/10.1128/AAC.01310-13>.
- Hansen, Elizabeth E., Catherine A. Lozupone, Federico E. Rey, Meng Wu, Janaki L. Guruge, Aneesha Narra, Jonathan Goodfellow, et al. 2011. "Pan-Genome of the Dominant Human Gut-Associated Archaeon, *Methanobrevibacter Smithii*, Studied in Twins." *Proceedings of the National Academy of Sciences* 108 (Supplement 1): 4599–4606. <https://doi.org/10.1073/pnas.1000071108>.
- He, Qing, Yuan Gao, Zhuye Jie, Xinlei Yu, Janne Marie Laursen, Liang Xiao, Ying Li, et al. 2017. "Two Distinct Metacommunities Characterize the Gut Microbiota in Crohn's Disease Patients." *GigaScience* 6 (7). <https://doi.org/10.1093/gigascience/gix050>.
- Hoedt, Emily C., Páiraic Ó Cuív, Paul N. Evans, Wendy J. M. Smith, Chris S. McSweeney, Stuart E. Denman, and Mark Morrison. 2016. "Differences Down-under: Alcohol-Fueled Methanogenesis by Archaea Present in Australian Macropodids." *The ISME Journal* 10 (10): 2376–88. <https://doi.org/10.1038/ismej.2016.41>.
- Huerta-Cepas, Jaime, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. 2017. "Fast Genome-Wide Functional Annotation through Orthology Assignment by EggNOG-Mapper." *Molecular Biology and Evolution* 34 (8): 2115–22. <https://doi.org/10.1093/molbev/msx148>.
- Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, et al. 2019. "EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses." *Nucleic Acids Research* 47 (D1): D309–14. <https://doi.org/10.1093/nar/gky1085>.
- Ingle, Danielle J., Mary Valcanis, Alex Kuzevski, Marija Tauschek, Michael Inouye, Tim Stinear, Myron M. Levine, Roy M. Robins-Browne, and Kathryn E. Holt. 2016. "In Silico Serotyping of E. Coli from Short Read Data Identifies Limited Novel O-Loci but Extensive Diversity of O:H Serotype Combinations within and between Pathogenic Lineages." *Microbial Genomics* 2 (7). <https://doi.org/10.1099/mgen.0.000064>.
- Jain, Chirag, Luis M. Rodriguez-R, Adam M. Phillippy, Konstantinos T. Konstantinidis, and Srinivas

- Aluru. 2018. "High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries." *Nature Communications* 9 (1): 5114. <https://doi.org/10.1038/s41467-018-07641-9>.
- Jennings, Matthew E., Nicholas Chia, Lisa A. Boardman, and William W. Metcalf. 2017. "Draft Genome Sequence of Methanobrevibacter Smithii Isolate WWM1085, Obtained from a Human Stool Sample." *Genome Announcements* 5 (39). <https://doi.org/10.1128/genomeA.01055-17>.
- Jia, Baofeng, Amogelang R. Raphenya, Brian Alcock, Nicholas Waglechner, Peiyao Guo, Kara K. Tsang, Briony A. Lago, et al. 2017. "CARD 2017: Expansion and Model-Centric Curation of the Comprehensive Antibiotic Resistance Database." *Nucleic Acids Research* 45 (D1): D566–73. <https://doi.org/10.1093/nar/gkw1004>.
- Karlsson, Fredrik H., Frida Fåk, Intawat Nookaew, Valentina Tremaroli, Björn Fagerberg, Dina Petranovic, Fredrik Bäckhed, and Jens Nielsen. 2012. "Symptomatic Atherosclerosis Is Associated with an Altered Gut Metagenome." *Nature Communications* 3 (1): 1245. <https://doi.org/10.1038/ncomms2266>.
- Khelaifia, Saber, Aurelia Caputo, Claudia Andrieu, Frederique Cadoret, Nicholas Armstrong, Caroline Michelle, Jean-Christophe Lagier, Felix Djossou, Pierre-Edouard Fournier, and Didier Raoult. 2018. "Genome Sequence and Description of Haloferax Massiliense Sp. Nov., a New Halophilic Archaeon Isolated from the Human Gut." *Extremophiles* 22 (3): 485–98. <https://doi.org/10.1007/s00792-018-1011-1>.
- Khelaifia, Saber, Marc Garibal, Catherine Robert, Didier Raoult, and Michel Drancourt. 2014a. "Draft Genome Sequencing of Methanobrevibacter Oralis Strain JMR01, Isolated from the Human Intestinal Microbiota." *Genome Announcements* 2 (1). <https://doi.org/10.1128/genomeA.00073-14>.
- Khelaifia, Saber, Marc Garibal, Catherine Robert, Didier Raoult, and Michel Drancourt. 2014b. "Draft Genome Sequence of a Human-Associated Isolate of Methanobrevibacter Arboriphilicus, the Lowest-G+C-Content Archaeon." *Genome Announcements* 2 (1). <https://doi.org/10.1128/genomeA.01181-13>.
- Kim, Joon Yong, Tae Woong Whon, Mi Young Lim, Yeon Bee Kim, Namhee Kim, Min-Sung Kwon, Juseok Kim, et al. 2020. "The Human Gut Archaeome: Identification of Diverse Haloarchaea in Korean Subjects." *Microbiome* 8 (1): 114. <https://doi.org/10.1186/s40168-020-00894-x>.
- Kitts, Paul A., Deanna M. Church, Françoise Thibaud-Nissen, Jinna Choi, Vichet Hem, Victor Sapojnikov, Robert G. Smith, et al. 2016. "Assembly: A Resource for Assembled Genomes at NCBI." *Nucleic Acids Research* 44 (D1): D73–80. <https://doi.org/10.1093/nar/gkv1226>.
- Kolde, M. R. 2019. *Package 'Pheatmap'*. 2015.
- Koskinen, Kaisa, Manuela R. Pausan, Alexandra K. Perras, Michael Beck, Corinna Bang, Maximilian Mora, Anke Schilhabel, Ruth Schmitz, and Christine Moissl-Eichinger. 2017. "First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin." *MBio* 8 (6). <https://doi.org/10.1128/mBio.00824-17>.
- Krupovič, Mart, Patrick Forterre, and Dennis H. Bamford. 2010. "Comparative Analysis of the Mosaic Genomes of Tailed Archaeal Viruses and Proviruses Suggests Common Themes for Virion Architecture and Assembly with Tailed Viruses of Bacteria." *Journal of Molecular Biology* 397 (1): 144–60. <https://doi.org/10.1016/j.jmb.2010.01.037>.
- Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." *Molecular Biology and Evolution* 35 (6): 1547–49. <https://doi.org/10.1093/molbev/msy096>.
- Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. <https://doi.org/10.1038/nmeth.1923>.
- Le Chatelier, Emmanuelle, Trine Nielsen, Junjie Qin, Edi Prifti, Falk Hildebrand, Gwen Falony, Mathieu Almeida, et al. 2013. "Richness of Human Gut Microbiome Correlates with Metabolic Markers." *Nature* 500 (7464): 541–46. <https://doi.org/10.1038/nature12506>.
- Letunic, Ivica, and Peer Bork. 2019. "Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments." *Nucleic Acids Research* 47 (W1): W256–59.

- <https://doi.org/10.1093/nar/gkz239>.
- Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79.
- Li, Junhua, Huijue Jia, Xianghang Cai, Huanzi Zhong, Qiang Feng, Shinichi Sunagawa, Manimozhiyan Arumugam, et al. 2014. "An Integrated Catalog of Reference Genes in the Human Gut Microbiome." *Nature Biotechnology* 32 (8): 834–41. <https://doi.org/10.1038/nbt.2942>.
- Liu, Bo, Dandan Zheng, Qi Jin, Lihong Chen, and Jian Yang. 2019. "VFDB 2019: A Comparative Pathogenomic Platform with an Interactive Web Interface." *Nucleic Acids Research* 47 (D1): D687–92. <https://doi.org/10.1093/nar/gky1080>.
- Liu, Wenjun, Jiachao Zhang, Chunyan Wu, Shunfeng Cai, Weiqiang Huang, Jing Chen, Xiaoxia Xi, et al. 2016. "Unique Features of Ethnic Mongolian Gut Microbiome Revealed by Metagenomic Analysis." *Scientific Reports* 6 (1): 34826. <https://doi.org/10.1038/srep34826>.
- Lloyd-Price, Jason, Anup Mahurkar, Gholamali Rahnava, Jonathan Crabtree, Joshua Orvis, A. Brantley Hall, Arthur Brady, et al. 2017. "Strains, Functions and Dynamics in the Expanded Human Microbiome Project." *Nature* 550 (7674): 61–66. <https://doi.org/10.1038/nature23889>.
- Loomba, Rohit, Victor Seguritan, Weizhong Li, Tao Long, Niels Klitgord, Archana Bhatt, Parambir Singh Dulai, et al. 2017. "Gut Microbiome-Based Metagenomic Signature for Non-Invasive Detection of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease." *Cell Metabolism* 25 (5): 1054-1062.e5. <https://doi.org/10.1016/j.cmet.2017.04.001>.
- Lurie-Weinberger, Mor Nadia, Michael Peeri, Tamir Tuller, and Uri Gophna. 2012. "Extensive Inter-Domain Lateral Gene Transfer in the Evolution of the Human Commensal *Methanospaera Stadtmanae*." *Frontiers in Genetics* 3. <https://doi.org/10.3389/fgene.2012.00182>.
- Lyu, Zhe, and Yahai Lu. 2018. "Metabolic Shift at the Class Level Sheds Light on Adaptation of Methanogens to Oxidative Environments." *The ISME Journal* 12 (2): 411–23. <https://doi.org/10.1038/ismej.2017.173>.
- Mahnert, Alexander, Nicholas P. Robinson, Marcus Blohs, Manuela-Raluca Pausan, and Christine Moissl-Eichinger. 2018. "The Human Archaeome: Methodological Pitfalls and Knowledge Gaps." *Emerging Topics in Life Sciences* 2 (4): 469–82. <https://doi.org/10.1042/ETLS20180037>.
- Matarazzo, Flavia, Adriana Costa Ribeiro, Magda Feres, Marcelo Favari, and Marcia Pinto Alves Mayer. 2011. "Diversity and Quantitative Analysis of Archaea in Aggressive Periodontitis and Periodontally Healthy Subjects." *Journal of Clinical Periodontology* 38 (7): 621–27. <https://doi.org/10.1111/j.1600-051X.2011.01734.x>.
- Mehta, Raaj S., Galeb S. Abu-Ali, David A. Drew, Jason Lloyd-Price, Ayshwarya Subramanian, Paul Lochhead, Amit D. Joshi, et al. 2018. "Stability of the Human Faecal Microbiome in a Cohort of Adult Men." *Nature Microbiology* 3 (3): 347–55. <https://doi.org/10.1038/s41564-017-0096-0>.
- Meile, Leo, Urs Jenal, Daniel Studer, Martin Jordan, and Thomas Leisinger. "Characterization of  $\psi$  M1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg." *Archives of microbiology* 152, no. 2 (1989): 105-110.
- Miller, Terry L., and Meyer J. Wolin. 1985. "Methanospaera Stadtmaniae Gen. Nov., Sp. Nov.: A Species That Forms Methane by Reducing Methanol with Hydrogen." *Archives of Microbiology* 141 (2): 116–22. <https://doi.org/10.1007/BF00423270>.
- Mitchell, Alex L., Alexandre Almeida, Martin Beracochea, Miguel Boland, Josephine Burgin, Guy Cochrane, Michael R. Crusoe, et al. 2020. "MGnify: The Microbiome Analysis Resource in 2020." *Nucleic Acids Research* 48 (D1): D570–78. <https://doi.org/10.1093/nar/gkz1035>.
- Moissl-Eichinger, Christine, Manuela Pausan, Julian Taffner, Gabriele Berg, Corinna Bang, and Ruth A. Schmitz. 2018. "Archaea Are Interactive Components of Complex Microbiomes." *Trends in Microbiology* 26 (1): 70–85. <https://doi.org/10.1016/j.tim.2017.07.004>.
- Nayfach, Stephen, Antonio Pedro Camargo, Emiley Eloie-Fadrosch, Simon Roux, and Nikos Kyrpides.



2020. "CheckV: Assessing the Quality of Metagenome-Assembled Viral Genomes." *BioRxiv*, May, 2020.05.06.081778. <https://doi.org/10.1101/2020.05.06.081778>.
- Nayfach, Stephen, Simon Roux, Rekha Seshadri, Daniel Udvary, Neha Varghese, Frederik Schulz, Dongying Wu, et al. 2020. "A Genomic Catalog of Earth's Microbiomes." *Nature Biotechnology*, November, 1–11. <https://doi.org/10.1038/s41587-020-0718-6>.
- Nayfach, Stephen, Zhou Jason Shi, Rekha Seshadri, Katherine S. Pollard, and Nikos C. Kyrpides. 2019. "New Insights from Uncultivated Genomes of the Global Human Gut Microbiome." *Nature* 568 (7753): 505–10. <https://doi.org/10.1038/s41586-019-1058-x>.
- Nielsen, H. Bjørn, Mathieu Almeida, Agnieszka Sierakowska Juncker, Simon Rasmussen, Junhua Li, Shinichi Sunagawa, Damian R. Plichta, et al. 2014. "Identification and Assembly of Genomes and Genetic Elements in Complex Metagenomic Samples without Using Reference Genomes." *Nature Biotechnology* 32 (8): 822–28. <https://doi.org/10.1038/nbt.2939>.
- Obregon-Tito, Alexandra J., Raul Y. Tito, Jessica Metcalf, Krithivasan Sankaranarayanan, Jose C. Clemente, Luke K. Ursell, Zhenjiang Zech Xu, et al. 2015. "Subsistence Strategies in Traditional Societies Distinguish Gut Microbiomes." *Nature Communications* 6 (1): 6505. <https://doi.org/10.1038/ncomms7505>.
- Olm, Matthew R., Christopher T. Brown, Brandon Brooks, and Jillian F. Banfield. 2017. "DRep: A Tool for Fast and Accurate Genomic Comparisons That Enables Improved Genome Recovery from Metagenomes through de-Replication." *The ISME Journal* 11 (12): 2864–68. <https://doi.org/10.1038/ismej.2017.126>.
- Ondov, Brian D., Nicholas H. Bergman, and Adam M. Phillippy. 2011. "Interactive Metagenomic Visualization in a Web Browser." *BMC Bioinformatics* 12 (1): 385. <https://doi.org/10.1186/1471-2105-12-385>.
- Ondov, Brian D., Todd J. Treangen, Páll Melsted, Adam B. Mallonee, Nicholas H. Bergman, Sergey Koren, and Adam M. Phillippy. 2016. "Mash: Fast Genome and Metagenome Distance Estimation Using MinHash." *Genome Biology* 17 (1): 132. <https://doi.org/10.1186/s13059-016-0997-x>.
- Ouwerkerk, D., Gilbert, R.A. and Klieve, A., 2011. Archaeophage therapy to control rumen methanogens.
- Page, Andrew J., Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T. G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, and Julian Parkhill. 2015. "Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis." *Bioinformatics* 31 (22): 3691–93. <https://doi.org/10.1093/bioinformatics/btv421>.
- Palleja, Albert, Alireza Kashani, Kristine H. Allin, Trine Nielsen, Chenchen Zhang, Yin Li, Thorsten Brach, et al. 2016. "Roux-En-Y Gastric Bypass Surgery of Morbidly Obese Patients Induces Swift and Persistent Changes of the Individual Gut Microbiota." *Genome Medicine* 8 (1): 67. <https://doi.org/10.1186/s13073-016-0312-1>.
- Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. "CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes." *Genome Research* 25 (7): 1043–55. <https://doi.org/10.1101/gr.186072.114>.
- Parks, Donovan H., Christian Rinke, Maria Chuvochina, Pierre-Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz, and Gene W. Tyson. 2017. "Recovery of Nearly 8,000 Metagenome-Assembled Genomes Substantially Expands the Tree of Life." *Nature Microbiology* 2 (11): 1533–42. <https://doi.org/10.1038/s41564-017-0012-7>.
- Pasolli, Edoardo, Francesco Asnicar, Serena Manara, Moreno Zolfo, Nicolai Karcher, Federica Armanini, Francesco Beghini, et al. 2019. "Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle." *Cell* 176 (3): 649-662.e20. <https://doi.org/10.1016/j.cell.2019.01.001>.
- Pausan, Manuela R., Cintia Csorba, Georg Singer, Holger Till, Veronika Schöpf, Elisabeth Santigli, Barbara Klug, Christoph Högenauer, Marcus Blohs, and Christine Moissl-Eichinger. 2019. "Exploring the Archaeome: Detection of Archaeal Signatures in the Human Body." *Frontiers in Microbiology* 10. <https://doi.org/10.3389/fmicb.2019.02796>.

- Pehrsson, Erica C., Pablo Tsukayama, Sanket Patel, Melissa Mejía-Bautista, Giordano Sosa-Soto, Karla M. Navarrete, Maritza Calderon, et al. 2016. "Interconnected Microbiomes and Resistomes in Low-Income Human Habitats." *Nature* 533 (7602): 212–16. <https://doi.org/10.1038/nature17672>.
- Petersen, Lauren M., Eddy J. Bautista, Hoan Nguyen, Blake M. Hanson, Lei Chen, Sai H. Lek, Erica Sodergren, and George M. Weinstock. 2017. "Community Characteristics of the Gut Microbiomes of Competitive Cyclists." *Microbiome* 5 (1): 98. <https://doi.org/10.1186/s40168-017-0320-4>.
- Pietilä, M.K., Atanasova, N.S., Manole, V., Liljeroos, L., Butcher, S.J., Oksanen, H.M. and Bamford, D.H., 2012. Virion architecture unifies globally distributed pleolipoviruses infecting halophilic archaea. *Journal of virology*, 86(9), pp.5067-5079.
- Prangishvili, David, Dennis H. Bamford, Patrick Forterre, Jaime Iranzo, Eugene V. Koonin, and Mart Krupovic. 2017. "The Enigmatic Archaeal Virosphere." *Nature Reviews Microbiology* 15 (12): 724–39. <https://doi.org/10.1038/nrmicro.2017.125>.
- Prangishvili, David, Patrick Forterre, and Roger A. Garrett. 2006. "Viruses of the Archaea: A Unifying View." *Nature Reviews Microbiology* 4 (11): 837–48. <https://doi.org/10.1038/nrmicro1527>.
- Qin, Junjie, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, et al. 2010. "A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing." *Nature* 464 (7285): 59–65. <https://doi.org/10.1038/nature08821>.
- Qin, Junjie, Yingrui Li, Zhiming Cai, Shenghui Li, Jianfeng Zhu, Fan Zhang, Suisha Liang, et al. 2012. "A Metagenome-Wide Association Study of Gut Microbiota in Type 2 Diabetes." *Nature* 490 (7418): 55–60. <https://doi.org/10.1038/nature11450>.
- Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." *Nucleic Acids Research* 41 (D1): D590–96. <https://doi.org/10.1093/nar/gks1219>.
- Rampelli, Simone, Stephanie L. Schnorr, Clarissa Consolandi, Silvia Turroni, Marco Severgnini, Clelia Peano, Patrizia Brigidi, Alyssa N. Crittenden, Amanda G. Henry, and Marco Candela. 2015. "Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota." *Current Biology* 25 (13): 1682–93. <https://doi.org/10.1016/j.cub.2015.04.055>.
- Rodriguez-R, Luis M., Santosh Gunturu, William T. Harvey, Ramon Rosselló-Mora, James M. Tiedje, James R. Cole, and Konstantinos T. Konstantinidis. 2018. "The Microbial Genomes Atlas (MiGA) Webserver: Taxonomic and Gene Diversity Analysis of Archaea and Bacteria at the Whole Genome Level." *Nucleic Acids Research* 46 (W1): W282–88. <https://doi.org/10.1093/nar/gky467>.
- Rodriguez-R, Luis M., and Konstantinos T. Konstantinidis. 2016. "The Enveomics Collection: A Toolbox for Specialized Analyses of Microbial Genomes and Metagenomes." e1900v1. PeerJ Inc. <https://doi.org/10.7287/peerj.preprints.1900v1>.
- Roux, Simon, Evelien M. Adriaenssens, Bas E. Dutilh, Eugene V. Koonin, Andrew M. Kropinski, Mart Krupovic, Jens H. Kuhn, et al. 2019. "Minimum Information about an Uncultivated Virus Genome (MIUViG)." *Nature Biotechnology* 37 (1): 29–37. <https://doi.org/10.1038/nbt.4306>.
- Sankaranarayanan, Krithivasan, Andrew T. Ozga, Christina Warinner, Raul Y. Tito, Alexandra J. Obregon-Tito, Jiawu Xu, Patrick M. Gaffney, et al. 2015. "Gut Microbiome Diversity among Cheyenne and Arapaho Individuals from Western Oklahoma." *Current Biology* 25 (24): 3161–69. <https://doi.org/10.1016/j.cub.2015.10.060>.
- Schirmer, Melanie, Sanne P. Smeekens, Hera Vlamakis, Martin Jaeger, Marije Oosting, Eric A. Franzosa, Rob ter Horst, et al. 2016. "Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity." *Cell* 167 (4): 1125-1136.e8. <https://doi.org/10.1016/j.cell.2016.10.020>.
- Seemann, Torsten. 2014. "Prokka: Rapid Prokaryotic Genome Annotation." *Bioinformatics* 30 (14): 2068–69. <https://doi.org/10.1093/bioinformatics/btu153>.
- Self, William T., Amy M. Grunden, Adnan Hasona, and Keelnatham T. Shanmugam. 2001. "Molybdate



- Transport Florida Agricultural Experiment Station Journal Series No. R-08044." *Research in Microbiology* 152 (3): 311–21. [https://doi.org/10.1016/S0923-2508\(01\)01202-5](https://doi.org/10.1016/S0923-2508(01)01202-5).
- Smits, Samuel A., Jeff Leach, Erica D. Sonnenburg, Carlos G. Gonzalez, Joshua S. Lichtman, Gregor Reid, Rob Knight, et al. 2017. "Seasonal Cycling in the Gut Microbiome of the Hadza Hunter-Gatherers of Tanzania." *Science* 357 (6353): 802–6. <https://doi.org/10.1126/science.aan4834>.
- Snipen, Lars, and Kristian Hovde Liland. 2015. "Micropan: An R-Package for Microbial Pan-Genomics." *BMC Bioinformatics* 16 (1): 79. <https://doi.org/10.1186/s12859-015-0517-0>.
- Snyder, Jamie C., Benjamin Bolduc, and Mark J. Young. 2015. "40 Years of Archaeal Virology: Expanding Viral Diversity." *Virology*, 60th Anniversary Issue, 479–480 (May): 369–78. <https://doi.org/10.1016/j.virol.2015.03.031>.
- Söllinger, Andrea, Clarissa Schwab, Thomas Weinmaier, Alexander Loy, Alexander T. Tveit, Christa Schleper, and Tim Urich. 2016. "Phylogenetic and Genomic Analysis of Methanomassiliicoccales in Wetlands and Animal Intestinal Tracts Reveals Clade-Specific Habitat Preferences." *FEMS Microbiology Ecology* 92 (1). <https://doi.org/10.1093/femsec/fiv149>.
- Song, Ziwei, Yuanyuan Cai, Xingzhen Lao, Xue Wang, Xiaoxuan Lin, Yingyun Cui, Praveen Kumar Kalavagunta, et al. 2019. "Taxonomic Profiling and Populational Patterns of Bacterial Bile Salt Hydrolase (BSH) Genes Based on Worldwide Human Gut Microbiome." *Microbiome* 7 (1): 9. <https://doi.org/10.1186/s40168-019-0628-3>.
- Srinivasan, Gayathri, Carey M. James, and Joseph A. Krzycki. 2002. "Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA." *Science* 296 (5572): 1459–62. <https://doi.org/10.1126/science.1069588>.
- Steinegger, Martin, and Johannes Söding. 2017. "MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets." *Nature Biotechnology* 35 (11): 1026–28. <https://doi.org/10.1038/nbt.3988>.
- Stewart, R.D., Auffret, M.D., Warr, A., Wiser, A.H., Press, M.O., Langford, K.W., Liachko, I., Snelling, T.J., Dewhurst, R.J., Walker, A.W. and Roehe, R., 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nature communications*, 9(1), pp.1-11.
- Taffner, Julian, Armin Erlacher, Anastasia Bragina, Christian Berg, Christine Moissl-Eichinger, and Gabriele Berg. 2018. "What Is the Role of Archaea in Plants? New Insights from the Vegetation of Alpine Bogs." *MSphere* 3 (3). <https://doi.org/10.1128/mSphere.00122-18>.
- Tettelin, Hervé, David Riley, Ciro Cattuto, and Duccio Medini. 2008. "Comparative Genomics: The Bacterial Pan-Genome." *Current Opinion in Microbiology, Antimicrobials/Genomics*, 11 (5): 472–77. <https://doi.org/10.1016/j.mib.2008.09.006>.
- Torsvik, Terje, and Ian D. Dundas. 1974. "Bacteriophage of Halobacterium Salinarium." *Nature* 248 (5450): 680–81.
- Torsvik, Terje, and Ian D. Dundas. 1980. "Persisting Phage Infection in Halobacterium Salinarium Str. 1." *Journal of General Virology*, 47 (1): 29–36. <https://doi.org/10.1099/0022-1317-47-1-29>.
- Vallenet, D., S. Engelen, D. Mornico, S. Cruveiller, L. Fleury, A. Lajus, Z. Rouy, et al. 2009. "MicroScope: A Platform for Microbial Genome Annotation and Comparative Genomics." *Database* 2009 (January). <https://doi.org/10.1093/database/bap021>.
- Vatanen, Tommi, Aleksandar D. Kostic, Eva d’Hennezel, Heli Siljander, Eric A. Franzosa, Moran Yassour, Raivo Kolde, et al. 2016. "Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans." *Cell* 165 (4): 842–53. <https://doi.org/10.1016/j.cell.2016.04.007>.
- Vierbuchen, Tim, Corinna Bang, Hanna Rosigkeit, Ruth A. Schmitz, and Holger Heine. 2017. "The Human-Associated Archaeon Methanosphaera Stadtmanae Is Recognized through Its RNA and Induces TLR8-Dependent NLRP3 Inflammasome Activation." *Frontiers in Immunology* 8. <https://doi.org/10.3389/fimmu.2017.01535>.
- Vincent, Caroline, Mark A. Miller, Thaddeus J. Edens, Sudeep Mehrotra, Ken Dewar, and Ameer R. Manges. 2016. "Bloom and Bust: Intestinal Microbiota Dynamics in Response to Hospital Exposures and Clostridium Difficile Colonization or Infection." *Microbiome* 4 (1): 12.

- <https://doi.org/10.1186/s40168-016-0156-3>.
- Vogtmann, Emily, Xing Hua, Georg Zeller, Shinichi Sunagawa, Anita Y. Voigt, Rajna Hercog, James J. Goedert, Jianxin Shi, Peer Bork, and Rashmi Sinha. 2016. "Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing." *PLOS ONE* 11 (5): e0155362. <https://doi.org/10.1371/journal.pone.0155362>.
- Vrieze, Anne, Els Van Nood, Frits Holleman, Jarkko Salojärvi, Ruud S. Kootte, Joep F. W. M. Bartelsman, Geesje M. Dallinga-Thie, et al. 2012. "Transfer of Intestinal Microbiota From Lean Donors Increases Insulin Sensitivity in Individuals With Metabolic Syndrome." *Gastroenterology* 143 (4): 913-916.e7. <https://doi.org/10.1053/j.gastro.2012.06.031>.
- Wattam, Alice R., James J. Davis, Rida Assaf, Sébastien Boisvert, Thomas Brettin, Christopher Bun, Neal Conrad, et al. 2017. "Improvements to PATRIC, the All-Bacterial Bioinformatics Database and Analysis Resource Center." *Nucleic Acids Research* 45 (D1): D535-42. <https://doi.org/10.1093/nar/gkw1017>.
- Wen, Chengping, Zhijun Zheng, Tiejuan Shao, Lin Liu, Zhijun Xie, Emmanuelle Le Chatelier, Zhixing He, et al. 2017. "Quantitative Metagenomics Reveals Unique Gut Microbiome Biomarkers in Ankylosing Spondylitis." *Genome Biology* 18 (1): 142. <https://doi.org/10.1186/s13059-017-1271-6>.
- Wolf, S., Fischer, M.A., Kupczok, A., Reetz, J., Kern, T., Schmitz, R.A. and Rother, M., 2019. Characterization of the lytic archaeal virus Drs3 infecting *Methanobacterium formicicum*. *Archives of virology*, 164(3), pp.667-674.
- Wu, Hao, Eduardo Esteve, Valentina Tremaroli, Muhammad Tanweer Khan, Robert Caesar, Louise Mannerås-Holm, Marcus Ståhlman, et al. 2017. "Metformin Alters the Gut Microbiome of Individuals with Treatment-Naive Type 2 Diabetes, Contributing to the Therapeutic Effects of the Drug." *Nature Medicine* 23 (7): 850-858. <https://doi.org/10.1038/nm.4345>.
- Xie, Hailiang, Ruijin Guo, Huanzi Zhong, Qiang Feng, Zhou Lan, Bingcai Qin, Kirsten J. Ward, et al. 2016. "Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome." *Cell Systems* 3 (6): 572-584.e3. <https://doi.org/10.1016/j.cels.2016.10.004>.
- Ye, Zi, Ni Zhang, Chunyan Wu, Xinyuan Zhang, Qingfeng Wang, Xinyue Huang, Liping Du, et al. 2018. "A Metagenomic Study of the Gut Microbiome in Behcet's Disease." *Microbiome* 6 (1): 135. <https://doi.org/10.1186/s40168-018-0520-6>.
- Youngblut, Nicholas D., Jacobo de la Cuesta-Zuluaga, Georg H. Reischer, Silke Dauser, Nathalie Schuster, Chris Walzer, Gabrielle Stalder, Andreas H. Farnleitner, and Ruth E. Ley. 2020. "Large Scale Metagenome Assembly Reveals Novel Animal-Associated Microbial Genomes, Biosynthetic Gene Clusters, and Other Genetic Diversity." *BioRxiv*, June, 2020.06.05.135962. <https://doi.org/10.1101/2020.06.05.135962>.
- Youngblut, Nicholas D., Georg H. Reischer, Silke Dauser, Chris Walzer, Gabrielle Stalder, Andreas H. Farnleitner, and Ruth E. Ley. 2020. "Strong Influence of Vertebrate Host Phylogeny on Gut Archaeal Diversity." *BioRxiv*, November, 2020.11.10.376293. <https://doi.org/10.1101/2020.11.10.376293>.
- Youngster, Ilan, Jenny Sauk, Christina Pindar, Robin G. Wilson, Jess L. Kaplan, Mark B. Smith, Eric J. Alm, Dirk Gevers, George H. Russell, and Elizabeth L. Hohmann. 2014. "Fecal Microbiota Transplant for Relapsing *Clostridium Difficile* Infection Using a Frozen Inoculum From Unrelated Donors: A Randomized, Open-Label, Controlled Pilot Study." *Clinical Infectious Diseases* 58 (11): 1515-22. <https://doi.org/10.1093/cid/ciu135>.
- Yu, Jun, Qiang Feng, Sunny Hei Wong, Dongya Zhang, Qiao yi Liang, Youwen Qin, Longqing Tang, et al. 2017. "Metagenomic Analysis of Faecal Microbiome as a Tool towards Targeted Non-Invasive Biomarkers for Colorectal Cancer." *Gut* 66 (1): 70-78. <https://doi.org/10.1136/gutjnl-2015-309800>.
- Zakrzewski, Martha, Carla Proietti, Jonathan J. Ellis, Shihab Hasan, Marie-Jo Brion, Bernard Berger, and Lutz Krause. 2017. "Calypso: A User-Friendly Web-Server for Mining and Visualizing Microbiome-Environment Interactions." *Bioinformatics* 33 (5): 782-83. <https://doi.org/10.1093/bioinformatics/btw725>.

- Zankari, Ea, Henrik Hasman, Salvatore Cosentino, Martin Vestergaard, Simon Rasmussen, Ole Lund, Frank M. Aarestrup, and Mette Voldby Larsen. 2012. "Identification of Acquired Antimicrobial Resistance Genes." *Journal of Antimicrobial Chemotherapy* 67 (11): 2640–44. <https://doi.org/10.1093/jac/dks261>.
- Zeevi, David, Tal Korem, Niv Zmora, David Israeli, Daphna Rothschild, Adina Weinberger, Orly Ben-Yacov, et al. 2015. "Personalized Nutrition by Prediction of Glycemic Responses." *Cell* 163 (5): 1079–94. <https://doi.org/10.1016/j.cell.2015.11.001>.
- Zeikus, J. G., and Diane L. Henning. 1975. "Methanobacterium Arbophilicum Sp. Nov. An Obligate Anaerobe Isolated from Wetwood of Living Trees." *Antonie van Leeuwenhoek* 41 (1): 543–52. <https://doi.org/10.1007/BF02565096>.
- Zeller, Georg, Julien Tap, Anita Y Voigt, Shinichi Sunagawa, Jens Roat Kultima, Paul I Costea, Aurélien Amiot, et al. 2014. "Potential of Fecal Microbiota for Early-Stage Detection of Colorectal Cancer." *Molecular Systems Biology* 10 (11): 766. <https://doi.org/10.15252/msb.20145645>.
- Zhang, Xuan, Dongya Zhang, Huijue Jia, Qiang Feng, Donghui Wang, Di Liang, Xiangni Wu, et al. 2015. "The Oral and Gut Microbiomes Are Perturbed in Rheumatoid Arthritis and Partly Normalized after Treatment." *Nature Medicine* 21 (8): 895–905. <https://doi.org/10.1038/nm.3914>.