Title: Multi-omics analysis of mouse fecal microbiome reveals supplier-dependent functional differences and novel metagenome-assembled genomes

Authors: Zachary L McAdams[1]#, Susheel Bhanu Busi[2]#, Kevin L Gustafson[3], Nathan Bivens[4], Craig L Franklin[1,2,5,6] Paul Wilmes[2,7], and Aaron C Ericsson[1,2,5,6]*

Affiliations:

[1]Molecular Pathogenesis and Therapeutics Program, University of Missouri, Columbia, MO

[2]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[3]Department of Veterinary Pathobiology, University of Missouri, Columbia, MO

[4]University of Missouri (MU) Genomics Technology Core Facility, University of Missouri, Columbia, MO

[5]MU Metagenomics Center, University of Missouri, Columbia, MO

[6]MU Mutant Mouse Resource and Research Center, University of Missouri, Columbia, MO

[7]Department of Life Sciences and Medicine, Faculty of Science, Technology, and Medicine, University of Luxembourg, Belvaux, Luxembourg

#These authors contributed equally to this project

***Corresponding author**: ACE

Abstract

Host genetics, sex, and other within-source factors have been associated with characteristic effects on the fecal microbiome in mice, however, the commercial source of mice remains the dominant factor. Increasing evidence indicates that supplier-specific microbiomes in particular confer differences in disease susceptibility in models of inflammatory conditions, as well as baseline behavior and body morphology. However, current knowledge regarding the compositional differences between suppliers is based on 16S rRNA amplicon sequencing data, and functional differences between these communities remain poorly defined. Here, we applied a meta-omic (metagenomic and metatranscriptomic) approach to biomolecules (DNA/RNA) extracted from murine fecal samples representative of two large U.S. suppliers of research mice, which differ in composition, and influence baseline physiology and behavior as well as disease severity in mouse models of intestinal disease. We reconstructed high-quality metagenome-assembled genomes (MAGs), frequently containing genomic content unique to each supplier. These differences were observed both within pangenomes of dominant taxa as well as the epibiont *Saccharimonadaceae*. Additionally, transcriptional activity and pathway analyses revealed key functional differences between the metagenomes associated with each supplier, including differences

in carbohydrate enzyme activity and dissimilatory sulfate reduction by sulfate-reducing bacteria (SRB). These data provide a detailed characterization of the baseline differences in the fecal metagenome of laboratory mice from two U.S. commercial suppliers suggesting that these functional differences are influenced by differences in the initial inoculum of colony founders, as well as additional taxa gained during growth of the production colony.

Key words (5-10): gut microbiome • metagenomics • metatranscriptomics • metagenome-assembled genomes • Jackson Laboratory • Envigo Laboratory

Introduction

Host-associated microbiomes, such as the gut microbiome (GM), exert strong effects on host physiology, susceptibility or resistance to various conditions, and response to treatment and dietary challenges. Investigations at the population-level suggest that differences in the human GM are responsible for a large portion of the variability within individual host responses to a given dietary challenge[1,2] or medical treatment,[3–5] implying that the GM is an important consideration in precision health and medicine strategies. Similarly, the GM of laboratory mice within the biomedical research community is highly variable due to numerous covariates,[6,7] and these compositional differences have been associated with differences in host fitness in the context of uniform host genetics and environment.[8–12] One of the dominant factors contributing to the population-level variability in specific pathogen-free (SPF) mouse microbiomes is the commercial source of mice.[13–15] Previous studies have demonstrated reproducible differences in the GM richness and beta-diversity, irrespective of host genotype (i.e., strain) within each supplier.[12,13] Specifically, the GM of mice supplied by the Jackson Laboratory (Jackson) and Envigo are characteristically of low and high richness, relative to each other, and each comprises unique taxa, in addition to an apparent core population of bacteria common to both sources. The latter includes members of the semi-standardized altered Schaedler flora (ASF),[16,17] reflecting the common procedures used to establish mouse production colonies on a commercial scale. Suppliers often surgically transfer embryos of the desired genotype to a pseudopregnant surrogate dam colonized with ASF, which then seeds the initial generation of offspring with that limited microbiome comprising 8 to 10 cultivable bacteria.[18] These mice are then used to establish multiple generations of filial mating to expand the colony, during which time mice are housed in large open-top caging systems and allowed to become colonized with additional bacteria from the environment. It is believed that subtle environmental differences are responsible for the reported supplier-origin differences, as well as the differences between multiple distribution facilities of the same supplier[13] or changes within a supplier over time.[19,20]

However, GMs with different taxonomic compositions may possess qualitatively similar functional capacities.[21,22] It is therefore unclear whether the different GMs colonizing mice from Jackson and Envigo are functionally different. Owing to the reported influence of these GMs in multiple mouse models of disease,[23–25] we hypothesized that the compositional differences result in substantial functional differences, as evaluated by the metatranscriptome. Any detected differences in the functional capacity of the fecal microbiome could therefore be due to differences in the ASF isolates maintained by each institution, the environmental exposures during colony expansion, or both. As researchers continue to leverage the inherent differential effects of these complex GMs as a population-level model of human host/microbe interactions, it is important to understand the differences in the metagenome and transcriptional activity of mice from these different suppliers, and the origin of any detected differences. With those goals in mind, fecal samples from healthy adult CD-1 mice colonized with a Jackson-origin or Envigo-origin GM (GM1 and GM4, respectively) were collected and used as the source of DNA and RNA for metagenomic and metatranscriptomic analyses using a re-iterative co-assembly procedure. We report here the identification of 86 high- and medium-quality novel and previously identified metagenome-assembled genomes (MAGs), analyzed and compared in the context of the two source GMs, and a detailed description of the functional differences between mice from these two commercial sources of SPF mice.

Results

Metagenomic, metatranscriptomic and taxonomic summary

To establish a taxonomic and functional profile, using IMP[26] (v3, commit #9672c874; available at https://git-r3lab.uni.lu/IMP/imp3) 2.09 x 10^9 metagenomic and metatranscriptomic reads were co-assembled and binned into MAGs. Subsequently, the completeness and contamination were assessed using CheckM. Per established criteria in the field,[27] **Table 1** lists the 29 high-quality (>90% completion and < 5% contamination) MAGs from the entire dataset (**Figure 1A**). An additional 35 medium-quality (> 80% completion and <10% contamination) MAGs, 22 medium-quality MAGs with completeness >50%, 17 low-quality (partial) MAGs with between 31% and 49% completeness, and 25 MAGs with >50% completeness but >10% contamination were identified (**Figure 1A)**. A complete list of the 128 identified MAGs is provided as **Supplementary Table 1**. Over 75% of MAGs contained greater than 20 tRNA-encoding genes, with over half encoding 30 or more tRNA genes (**Figure 1B**). Over 75% of the 128 assembled MAGs had an average coverage of 10× or greater (median 24.4×, range 2.1× to 1540×; **Supplementary Table 1**) and roughly 65% of MAGs (including the majority of high-quality MAGS) were assembled from less than 200 scaffolds (**Figure 1C**). Comparison of metagenomic composition and the metatranscriptome revealed a strong correlation, suggesting transcriptional activity of the majority of

detected genes (**Figure 1D**). As expected, there was also a strong correlation between the number of genes detected and total size of the assembled MAGs (**Supplementary Figure 1**).

Of the 64 high- and medium-quality MAGs with > 80% completion and <10% contamination listed in **Supplementary Table 1**, over one third (23/64) were identified as members of the Gram-positive family *Lachnospiraceae* (phylum *Bacillota*). The second most common family was the Gram-negative *Muribaculaceae*, within the phylum *Bacteroidota*. Other MAGs within the phylum *Bacillota* included several members of the *Ruminococcaceae*, *Clostridiacae*, *Bacillaceae*, and *Lactobacillaceae*, among others. Additional MAGs within the *Bacteroidota* included members of the genera *Alistipes*, *Bacteroides*, *Parabacteroides*, *Odoribacter*, and *Prevotella*. Six of the high- and medium-quality 64 MAGs in **Supplementary Table 1** were external to either of those two dominant phyla, including one identified as *Parasutterella excrementihominis* (*Burkholderiaceae* within phylum *Pseudomonadota*), and five identified as members of the family *Saccharimonadaceae* (phylum *Patescibacteria*).

Candidate Phyla Radiation taxa demonstrate strain-level differences between vendors

As newly recognized epibionts within the candidate phylum radiation (CPR), the *Saccharimonadaceae* were of particular interest since their reports in laboratory mouse strains are limited. MAGs identified as *Saccharimonadaceae* have been found in diverse environmental samples including deep sea hydrothermal vents, glacial-fed stream biofilms,[28] and petrochemical plant sludge.[29] Regarding host-associated samples, *Saccharimonadaceae* are most commonly identified in human oral cavity samples,[30,31] although a handful of rumen[32] and fecal[33] samples have also yielded MAGs. A thorough search of the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) found 321 MAGs within this phylum, including four MAGs from mouse feces. Comparison of the phylogenetic relationship of the newly generated five MAGs within *Saccharimonadaceae* revealed similarity to other host-associated isolates, and particularly the mouse-origin MAGs (**Figure 2A**). Construction of a *Saccharimonadaceae* pangenome from the current data revealed portions of highly conserved core genomic content, and regions of genomic material specific to MAGs from either of the two supplier-dependent microbiomes (**Figure 2B**), suggesting the vendors each harbor distinct strains of this taxonomy, with distinct functional capacities. These data also suggested the co-evolution and transfer of genetic material between bacterial strains within each source.

Distinct source-dependent MAGs within multiple taxonomies

To further investigate the genomic heterogeneity within other common taxonomies, separate pangenomes were created for various members of the Gram-negative phylum *Bacteroidota*, including *Alistipes* spp. (10 MAGs, **Figure 3A**), *Prevotella* spp. (9 MAGs, **Figure 3B**), and family *Muribaculaceae*

(17 MAGs, **Figure 3C**). As in the *Saccharimonadaceae* pangenome comparison, each genus or family revealed regions of genomic content conserved between multiple MAGs from each of the supplier-origin microbiomes, along with conserved core genomic content encoding for single copy gene (SCG) clusters, suggesting that the transfer of genetic material is an ongoing process within each of these taxonomies, at each production source. Similarly, pangenomes were constructed from dominant members of the Gram-positive phylum *Bacillota*, including *Lactobacillus* spp. (14 MAGs, **Figure 4A**) and family *Lachnospiraceae* (20 MAGs, **Figure 4B**). These pangenomes revealed conserved genomic content including highly conserved SCG clusters within each taxonomy, as well as source-dependent differences in the genomic content of MAGs, which may be interpreted as evidence of distinct lineages of Gram-positive taxonomies in mice from each supplier. Collectively, these data indicate the presence of substantial differences in the functional capacity of the dominant bacterial families detected in the microbiome of mice from different suppliers.

Functional differences between source-dependent GM

Based on the above observations and our original hypothesis, the metatranscriptome was compared between GM profiles to determine if the detected differences in metagenomic content were also associated with differences in transcriptional activity. Transcripts were compared to (and cross-referenced against) multiple databases including the Kyoto Encyclopedia of Genes and Genomes (KEGG),[34,35] the Protein family (Pfam) database,[36] and the CAZy database[37] of carbohydrate active enzymes and accessory molecules. **Figure 5A** shows KEGG-identified microbial-associated pathways comprising a multitude of differentially expressed KEGG orthologs (**Figure 5B**). A list of differentially expressed KEGG-identified host-associated pathways is shown in **Supplementary Figure 2**. Similarly, comparison of the bulk metatranscriptome annotations against the Pfam (**Figure 5C**) and CAZy (**Figure 5D**) databases resulted in many differences, with greater transcriptional activity of different genes in each GM. **Supplementary File 1** lists all significantly differing KEGG, Pfam, and CAZyme annotations as determined by DESeq2[38] ($p < 0.05$).

Source-dependent differences within the KEGG annotations included several GM1- and GM4-specific genes involved in a wide range of functions. To increase our ability to discern biologically meaningful differences in the function of these GMs, the top 25% most significant KEGG IDs (lowest p values identified by DESeq2) found to significantly differ between GM1 and GM4 were manually reviewed and curated to identify multiple KEGG IDs within a pathway, and thus likely representing true differences in the functional activity of that pathway (**Figure 5A-B**). Several GM1-specific genes involved in diverse metabolic functions were identified including starch and sucrose (CBH1, K01225; SI, K01203), and

fructose and mannose (algG, K01795; mtlK, K00045), arachidonic acid (EPHX2, K08726), and phenylalanine (mhpF, K04073; DDC, K1593) metabolism.

Source-dependent differences within the KEGG annotations also included several GM4-specific genes involved in numerous functions including flagellar assembly (flgH, K02393; flgI, K02394, flgA, K02386), quorum sensing (srfATE, K15657), lipopolysaccharide biosynthesis (lpxC, K02535; lpxI, K09949), and sulfur metabolism (dsrA, K11180). Pfam annotations also identified increased expression of genes within the dissimilatory sulfite reduction pathway (DsrC, DsrD, and FdhE) and chemotactic responses (CheZ, TarH) by bacteria within GM4. Among the many genes found to be differentially expressed, patterns emerged suggesting increased activity of certain pathways in GM4, including the tricarboxylic acid (TCA) cycle and cytochrome c oxidase activity. Increased TCA cycle activity is suggested by increased expression of enzymes within the TCA cycle (succinate dehydrogenase D; sdhD); enzymes involved in acetyl-CoA production (malonyl-CoA/succinyl-CoA reductase; mcr); three different ccb-type cytochrome c oxidase subunits (I, II, and III) and the fixS cytochrome c oxidase maturation protein; and the cytochrome c-type biogenic protein ccmE. Additionally, GM4 had increased expression of enzymes associated with acetate (acetoacetate decarboxylase, adc), propanoate (methylmalonyl-CoA mutase, mcmA1), and butanoate production (mcr) using TCA cycle compounds, suggesting that the increased release of stored energy by the GM may be associated with increased production of compounds beneficial to the host such as short chain fatty acids (SCFAs).

Lastly, numerous source-dependent differences in carbohydrate active enzymes (CAZymes) and accessory molecules were identified (**Figure 5D)**. The glycoside hydrolase (GH) family 48 (GH48.hmm) including chitinase and cellulobiohydrolases enzymes was differentially expressed in GM1 using both CAZyme and Pfam (Glyco_hydro_48) annotations. Other GM1-associated CAZyme molecules included the auxiliary activities (AA) of multicopper oxidases (AA1.hmm) and glycosyltransferase (GT) families that bind the LPS inner core polysaccharide[39] (GT99.hmm) and the host-produced extracellular polysaccharide heparan[40] (GT64.hmm). GM4-associated CAZymes included multiple non-catalytic carbohydrate binding motifs (CBMs) with diverse targets including β-1,3-glucan and LPS (CBM39.hmm), cyclodextrins (CBM20.hmm), lactose (CBM71.hmm), and fucose (CBM47.hmm). CBMs specific to cellulose and chitin were identified in both GM1 (CBM2.hmm, CBM72.hmm) and GM4 (CBM28.hmm). Collectively, these data demonstrate extensive differences in the baseline transcriptional activity at the enzyme and pathway levels of supplier-origin gut microbiomes.

Supplier-origin GMs indicate variable levels of enzymatic activity associated with eukaryotes

While most studies focus on bacterial abundances and differences, we observed eukaryotic organisms present within each GM (*Methods).* The largest portion of eukaryotes identified belonged to the phylum *Ochrophyta* followed by *Dinoflagellata* and *Chlorophyta*, all within the kingdom Protista **(Supplemental Figure 3).** Eukaryotes identified within the kingdom Fungi were constrained to the phylum *Ascomycota* with limited taxonomic resolution. No significant differences in the relative abundance of eukaryotic phyla were observed between GMs. Interestingly, while no differences in the relative abundance of eukaryotic phyla were observed between GM1 and GM4 (**Supplemental Figure 3**), glycoside hydrolase CAZyme expression was negatively correlated with GM1 eukaryotes while positively correlated to GM4 eukaryotes (**Supplemental Figure 4**). Lastly, we identified several genes with increased expression in GM1 previously associated with a wide range of host metabolism and disease pathways (**Supplementary Figure 2**), however, the biological significance of the differential expression of these pathways remains unclear.

Discussion

The Jackson (GM1)- and Envigo (GM4)-origin GMs influence many host phenotypes including intestinal inflammation,[23] colonization resistance,[25] and behavior and body morphology.[24] The robust taxonomic differences between these supplier-origin GMs influencing phenotypic differences have previously been identified using targeted amplicon (e.g., 16S rRNA) sequencing, however, this approach yields limited taxonomic resolution of detected amplicons, and a complete lack of information regarding functional capacity or transcriptional activity. Using an iterative co-assembly procedure, we combined metagenomic and metatranscriptomic sequencing of the fecal microbiome of laboratory mice to provide a valuable resource describing the baseline metagenomic and transcriptional differences of Jackson- and Envigo-origin GMs (GM1 and GM4, respectively). The current data build upon earlier reports of differences in the composition of the GM in mice from different suppliers[25,41,42] by providing a more detailed assessment of those differences as well as functional differences.

Many of those functional differences were attributable to differences in bacteria associated with the ancestral ASF used in the colony founders, including *Lactobacillus murinus* [ASF361] and *L. intestinalis* [ASF360]. These differences could therefore ostensibly be controlled or changed during the initiation of new production colonies. An additional notable aspect of the source-dependent differences in *Lactobacillus* function is the growing body of evidence supporting *Lactobacillus* spp. as psychobiotics,[43] or live organisms capable of conferring benefits to mental health when ingested. Recent studies have demonstrated differences in anxiety-related behavior and spontaneous locomotor and exploratory behavior between isogenic mice harboring GM1 or GM4[24] and *L. intestinalis* and related species have

both been shown to confer vagus nerve-dependent effects on behavior.[44–46] Differences in the genomic content of these MAGs provides one possible explanation for the host phenotypic differences.

Numerous differentially expressed KEGG orthologs representing several microbial- and host-associated pathways were identified between the supplier-origin GMs **(Figure 5A-B, Supplemental Figure 2)**. Consistent with the previously reported differences in *Pseudomonadota*[13] of GM1 relative to GM4, the Jackson-origin GM was associated with decreased expression of lipopolysaccharide biosynthesis and flagellar assembly. Low richness microbiomes have been associated with increased body weight, growth,[47] and intestinal inflammation.[23] Here we have identified that, in addition to fatty acid degradation, multiple carbohydrate metabolic pathways including starch and sucrose, galactose, and fructose metabolism were increased in the low-richness GM. The differential expression of these metabolic pathways may increase energy availability to the host likely contributing to the GM1-associated increase in body weight and growth[24] and increased intestinal inflammation in models of intestinal disease.[10,48,49]

A differentially expressed KEGG pathway in GM4 that can be linked to previously recognized compositional differences is the dissimilatory sulfite reduction (DSR) pathway, expressed by sulfate-reducing bacteria (SRB) such as *Desulfovibrio* and *Bilophila* spp. These taxa, identified as unique GM4-associated features[50] are responsible for production of H_2S, a compound with biphasic effects on inflammation, hypertension, and tumorigenesis depending on its intra- and extracellular concentrations.[51–57] Thus, augmentation of intracellular H_2S production by luminal SRB may result in the low levels adequate to confer protective effects in certain scenarios, or sufficiently high to adversely influence disease susceptibility in others.

These data are also of interest from an evolutionary perspective, as they provide a glimpse of the short-term evolutionary landscape within the GM at each supplier. Pathogenic bacteria frequently undergo rapid evolution within a host organism through recombination and mutation,[58,59] and the same events occur between and within commensal members of the microbiota.[60,61] Moreover, pathobionts can arise from commensal organisms through the same mechanisms.[62] In the data presented here, the consistent finding of source-specific genomic content within genera suggests separate evolutionary trajectories at each supplier, occurring in all dominant taxonomies with multiple closely related members. Notably, this feature was particularly evident in the relatively small pangenome of *Saccharimonadaceae*. These findings are in agreement with the recent study from Yilmaz et al. demonstrating the long-term evolution of microbiota and development of multiple co-existing substrains of bacteria within individual taxonomies.[63]

Lastly, we were surprised to recover a large number of high-quality MAGs associated with the family *Saccharimonadaceae* (formerly known as TM7), epibionts[64,65] which were unrecognized until their identification using molecular methods. Successful culture requires co-culture with the cognate host bacteria, including *Actinomyces odontolyticus* and other members of the human oral cavity.[65,66] That being said, these highly auxotrophic epibionts with extremely limited genomes are found in virtually all environmental conditions while being surprisingly scarce in metagenomic data from fecal microbiomes. Our analysis agrees with that by Dinis *et al*.,[67] which demonstrated that the vast majority of host-associated MAGs from this phylum were from human oral cavity samples or rumen contents, with much fewer fecal samples represented. It is unclear which bacteria serve as the host for fecal members of *Saccharibacteria*.

Detecting differences in microbial diversity and composition between Jackson- and Envigo-origin GMs has previously relied upon targeted amplicon sequencing of the 16S rRNA gene.[13,24] While informative, this approach is limited by taxonomic resolution and does not provide the baseline functional capacity or transcriptional activity of these GMs. Our metagenomic and metatranscriptomic sequencing of Jackson- and Envigo origin GMs has established that distinct differences in both the functional capacity and baseline transcriptional activity at the gene and metabolic pathway levels exist amongst the dominant taxa within supplier-origin GMs. Collectively, these data will serve as a valuable resource to leverage the host-microbiome relationship in mouse models of disease and behavior in future.

Methods

Mice and sample collection

Mice contributing fecal samples were eight-week-old, female, CD-1 mice produced by breeding colonies maintained at the MU Mutant Mouse Resource and Research Center in accordance with the Guide for the Care and Use of Laboratory Animals approved by the University of Missouri Institutional Animal Care and Use Committee (IACUC, protocol 9587). Mice colonized with a Jackson-origin GM (GM1) or Envigo-origin GM (GM4)[42] were housed in microisolator polycarbonate cages on individually ventilated racks, under positive pressure. A sample size of three mice per GM was selected to attain a power of 80% and a 5% alpha error rate reflecting changes in microbial composition, based on previously observed robust differences in beta-diversity and the presence of unique taxa within each supplier-origin GM.[24,25,42] All husbandry was performed in accordance with barrier conditions including use of autoclaved, irradiated chow, autoclaved, acidified water, and autoclaved bedding. Biweekly cage changes occurred in a laminar flow hood using bead-sterilized forceps to transfer mice between cages, by personnel wearing bleach-disinfected latex gloves. Mice were on a 14:10 light/dark cycle and

were determined to be free of all known pathogens based on comprehensive quarterly sentinel testing through IDEXX BioAnalytics.

Fecal samples were collected by placing each mouse into an empty, autoclaved, microisolator cage, allowing the mouse to defecate normally, and collecting the pellet into a sterile 1 mL cryovial using an autoclaved wooden toothpick. Cryovials were then sealed and flash-frozen in liquid nitrogen. Separate fecal pellets were collected from each mouse for DNA and RNA extraction.

DNA extraction

Fecal DNA was extracted using PowerFecal kits (Qiagen) per the manufacturer's protocol, with the exception that the initial sample disaggregation was performed with a TissueLyser II (Qiagen), operated at 30 Hz. DNA yields were eluted in 50 µL sterile water, quantified using Qubit 2.0 fluorometer and Quant-iT dsDNA Broad Range (BR) Assay kits, and diluted to a uniform volume and concentration.

RNA extraction

Fecal RNA was extracted using MagMAX mirVana Total RNA Isolation kits (Thermo Fisher) per the manufacturer's protocol. RNA yields were eluted in 50 µL sterile water, quantified using Qubit 2.0 fluorometer and Quant-iT RNA Broad Range (BR) Assay kits, and diluted to a uniform volume and concentration.

Metagenomic library preparation

Metagenomic libraries were generated from genomic DNA (250 ng) per manufacturer's protocol with reagents supplied in the Illumina DNA Prep, Tagmentation Kit. The sample concentration was determined using the Qubit dsDNA high-sensitivity (HS) assay kit. Genomic DNA was fragmented, and short adapter sequences ligated to the ends by bead-link transposomes. Tagmented DNA was amplified using a minimum number of PCR cycles (5) to complete adapter sequences required for cluster generation and the addition of unique dual indexes. Final libraries were purified by addition of Axyprep Mag PCR Clean-up beads. The final construct of each purified library was evaluated using the Fragment Analyzer, quantified using the Qubit HS dsDNA assay kit, and diluted according to Illumina's standard sequencing protocol.

Metatranscriptomic library preparation

Metatranscriptomic libraries were generated from total RNA (800 ng) per manufacturer's protocol with reagents supplied in NEBNext® rRNA Depletion Kit (Bacteria) followed by fragmentation and synthesis of cDNA using the Illumina Stranded mRNA Prep, Ligation Kit. The sample concentration was determined using the Qubit RNA high-sensitivity (HS) assay kit, and the RNA integrity checked using

the Agilent Fragment Analyzer automated electrophoresis system. The rRNA was first removed from total RNA by hybridization probes using the NEBNext kit instead of poly-A RNA enrichment. The rRNA-depleted samples were then precipitated and fragmented, and double-stranded cDNA was generated from fragmented RNA, and short adapter sequences ligated to the ends. The cDNA constructs were amplified using a minimum number of PCR cycles (10) to complete adapter sequences required for cluster generation and the addition of unique dual indexes. Final libraries were purified by addition of Axyprep Mag PCR Clean-up beads. The final construct of each purified library was evaluated using the Fragment Analyzer, quantified using the Qubit HS dsDNA assay kit, and diluted according to Illumina's standard sequencing protocol. Paired-end 150 base pair length reads were sequences using an Illumina NovaSeq 6000 instrument. All six metagenomic and six metatranscriptomic libraries were pooled to yield approximately 40 Gb per metagenomic library and 190 million paired end reads per metatranscriptome library.

Meta-omic preprocessing, assembly, binning, and analyses

For processing metagenomic sequence data, we used the Integrated Meta-omic Pipeline (IMP) workflow[68] to process paired forward and reverse reads using version 3.0 (commit# 9672c874; available at https://git-r3lab.uni.lu/IMP/imp3). IMP includes pre-processing, assembly, genome reconstructions and additional functional analysis of genes based on custom databases in a reproducible manner. Briefly, adapter trimming is followed by filtering the reads against the mouse reference genome (GRCm38, https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20/) to remove any reads mapping to the host, i.e. mice. Thereafter, an iterative co-assembly of both the metagenomic and metatranscriptomic reads using MEGAHIT v1.2.9[69] is performed. Concurrently, MetaBAT2 v2.12.1,[70] MaxBin2 v2.2.7,[71] and binny[72] were used for binning the assembly, for reconstructing metagenome-assembled genomes (MAGs). Upon completion of binning, we used DASTool[73] to select a non-redundant set of MAGs using a recommended threshold score of 0.7. Furthermore, CheckM v1.1.3[74] was used to assess the quality of the MAGs, and the GTDB-tooklit[75] was used to assign the taxonomy per MAG. To estimate the overall abundances of eukaryotes, EUKulele v1.0.5[76] was used on the assemblies, with both the MMETSP and the PhyloDB databases. Each of the databases were run separately to confirm the detected eukaryotic profiles, whereby conflicts in assigned taxonomy were resolved by selecting the best hit score. To understand the overall metabolic and functional potential of the metagenome and reconstructed MAGs we used MANTIS[77] which annotates both assemblies and MAGs alike using several databases such as KEGG,[34,35] PFAM,[36] and CAZyme.[37] All the parameters, databases, and relevant code for the analyses described above are openly available at https://github.com/susheelbhanu/mice_multiomics_mmrrc and included in the Code availability section.

Phylogenomics, pangenome construction and differential analyses

To perform the pangenome analyses, bins with the same level of taxonomic resolution, i.e., genus or family level, were collected. They were subsequently subjected to the pangenome workflow as described here http://merenlab.org/2016/11/08/pangenomics-v2, by Meren *et al. w*ithin the anvi'o[78] ecosystem. For the *Saccharibacteria* pangenome analysis, two existing genomes (accession IDs: CP040003 and CP040004.1) from Genbank were downloaded. The pangenome was run using the --min-bit 0.5, --mcl-inflation 10 and --min-occurence 2 parameters, excluding the partial gene calls. A phylogenomic tree was built using MUSCLE v3.8.1551[79] and FastTree2 v2.1.10[80] on all single-copy gene clusters in the pangenome that were present in at least 30 genomes and had a functional homogeneity index below 0.9, and geometric homogeneity index above 0.9. The phylogenomic tree was used to order the genomes, the frequency of gene clusters (GC) to order the GC dendrogram. For the Saccharibacteria phylogenetic tree, we used the *Entrez Direct* tools available at https://www.ncbi.nlm.nih.gov/books/NBK179288/, to fetch all genomes labelled as 'Saccharibacteria', within NCBI. Following this, the genomes were input to GToTree v1.5.51130[81] pipeline with the -D parameter, allowing to retrieve taxonomic information for the NCBI accessions, where the tree was built using 'Bacteria and Archaea' marker genes. Briefly, HMMER3 v3.3.2[82] was used to retrieve the single-copy genes after gene-calling with Prodigal v2.6.3[83] and aligned using TrimAl v1.4.rev15.[84] The entire workflow is based on GNU Parallel v20210222134.

Data analyses and figures

The heatmaps were generated using the ggplot2 package while the volcano plots were built using the EnhancedVolcano package found at *https://github.com/kevinblighe/EnhancedVolcano*. The correlation matrices were generated using the corrplot package. Furthermore, for the differential analyses, we used DESeq2[38] with FDR and multiple-testing adjustments to assess enriched KOs, pathways, and expression levels. For the Saccharibacteria tree visualization the following packages from the R environment were used: ape, ggree, ggtreeExtra and treeio.

Funding

This work was supported by the National Institutes of Health (NIH) under Grants R03 OD028259 and U42 OD010918. ZM was supported by the NIH under T32 GM008396. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 863664).

Disclosure Statement

The authors have no competing interests to declare.

Data Availability

Raw sequencing data samples and the MAGs are available at NCBI's sequence read archive under BioProject accession PRJNA876568. The BioSample accession IDs and the metadata associated with each sample are listed in **Supplementary Table 2**.

Code Availability

The detailed code for the downstream analyses including the assemblies using IMP is available at *https://github.com/susheelbhanu/mice_multiomics_mmrrc*. The code used to generate the Figure 5 and Supplemental Figures 1-3 is available at *https://github.com/ericsson-lab/metaG_metaT*.

Figure and Table Legends

Figure 1. Dot plot showing the completeness (%) and contamination (%) among the 128 metagenome-assembled genomes (MAGs) recovered from all six samples, legend in inset, dot size correlates to the number (1 to 6) MAGs represented (**A**); bar charts showing the number of tRNAs found in low-, medium-, and high-quality MAGs (**B**), and the number of contigs used to construct MAGs (**C**); and dot plot showing the number of expressed genes in relation to total detected genes in each MAG (**D**).

Figure 2. Phylogenetic tree (ignoring branch lengths) showing the relationship between the newly identified *Saccharimonadaceae* MAGs and 321 MAGs within the NCBI Sequence Read Archive (SRA) annotated to the *Saccharimonadaceae* family, asterisk represents gut-associated samples (**A**); and pangenome of novel *Saccharimonadaceae* MAGs showing genomic content specific to MAGs from each source (**B**).

Figure 3. Pangenomes of *Alistipes* (**A**), *Prevotella* (**B**), and family *Muribaculaceae* (**C**) constructed from the present data, each showing the conserved core genomic content, and additional genomic content, common to multiple MAGs from each supplier

Figure 4. Pangenome of *Lactobacillus* (**A**) and family *Lachnospiraceae* (**B**) constructed from the present data, each showing the conserved core genomic content, and additional genomic content, common to multiple MAGs from each supplier.

Figure 5. Heat map of differentially expressed select KEGG pathways (A) and volcano plots of individual KEGG (B), Pfam (C), and CAZyme (D) IDs between Jackson (GM1)- and Envigo (GM4)-origin microbiomes. Differentially abundance testing was performed using DESeq2 with $p < 0.05$ considered significant.

Supplemental Figure 1. Dot plot representing the significant correlation between the number of detected genes and assembled MAG size (Mb). Spearman correlation. R = 0.92, $p < 0.001$.

Supplementary Figure 2. Heatmaps of host-associated pathways differentially expressed in fecal metatranscriptomic data in Jackson (GM1)- and Envigo (GM4)-origin microbiomes.

Supplementary Figure 3. Relative abundance heatmaps of phyla representing greater than 1% of eukaryotes in Jackson (GM1)- and Envigo (GM4)-origin microbiomes.

Supplementary Figure 4. Heatmaps demonstrating correlations between classes of CAZyme molecules and detected eukaryotes in GM1 **(A)** and GM4 **(B).** *: $p < 0.05$.

Table 1. High quality MAGs (>90% completion and < 5% contamination) identified in GM1- and GM4-origin gut microbiomes.

References

1. Korem T, Zeevi D, Zmora N, Weissbrod O, Bar N, Lotan-Pompan M, Avnit-Sagi T, Kosower N, Malka G, Rein M, et al. Bread Affects Clinical Parameters and Induces Gut Microbiome-Associated Personal Glycemic Responses. Cell Metab 2017; 25:1243-1253 e5.

2. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, et al. Personalized Nutrition by Prediction of Glycemic Responses. Cell 2015; 163:1079–94.

3. Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre ML, Luke JJ, Gajewski TF. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. Science 2018; 359:104–8.

4. Routy B, Chatelier EL, Derosa L, Duong CPM, Alou MT, Daillere R, Fluckiger A, Messaoudene M, Rauber C, Roberti MP, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science 2018; 359:91–7.

5. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, Prieto PA, Vicente D, Hoffman K, Wei SC, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science 2018; 359:97–103.

6. Franklin CL, Ericsson AC. Microbiota and reproducibility of rodent models. Lab Animal 2017; 46:114–22.

7. Ericsson AC, Gagliardi J, Bouhan D, Spollen WG, Givan SA, Franklin CL. The influence of caging, bedding, and diet on the composition of the microbiota in different regions of the mouse gut. Scientific reports 2018; 8:4065.

8. Sofi MH, Gudi R, Karumuthil-Melethil S, Perez N, Johnson BM, Vasu C. pH of drinking water influences the composition of gut microbiome and type 1 diabetes incidence. Diabetes 2014; 63:632–44.

9. Wolf KJ, Daft JG, Tanner SM, Hartmann R, Khafipour E, Lorenz RG. Consumption of acidic water alters the gut microbiome and decreases the risk of diabetes in NOD mice. J Histochem Cytochem 2014; 62:237–50.

10. Hart ML, Ericsson AC, Franklin CL. Differing complex microbiota alter disease severity of the IL-10-/- mouse model of inflammatory bowel disease. Frontiers in microbiology 2017; 8.

11. Zhao Y, Tarbell KV. Comment on Sofi et al. pH of Drinking Water Influences the Composition of Gut Microbiome and Type 1 Diabetes Incidence. Diabetes 2014;63:632-644. Diabetes 2015; 64:e19.

12. Guo Y, Wang Q, Li D, Onyema OO, Mei Z, Manafi A, Banerjee A, Mahgoub B, Stoler MH, Barker TH, et al. Vendor-specific microbiome controls both acute and chronic murine lung allograft rejection by altering CD4+Foxp3+ regulatory T cell levels. Am J Transplant 2019; 19:2705–18.

461 13. Ericsson AC, Davis JW, Spollen W, Bivens N, Givan S, Hagan CE, McIntosh M, Franklin CL.
462 Effects of Vendor and Genetic Background on the Composition of the Fecal Microbiota of Inbred Mice.
463 Plos One 2015; 10:e0116704.

464 14. Rasmussen TS, Vries L de, Kot W, Hansen LH, Castro-Mejia JL, Vogensen FK, Hansen AK,
465 Nielsen DS. Mouse Vendor Influence on the Bacterial and Viral Gut Composition Exceeds the Effect of
466 Diet. Viruses 2019; 11:435.

467 15. Hufeldt MR, Nielsen DS, Vogensen FK, Midtvedt T, Hansen AK. Variation in the gut microbiota of
468 laboratory mice is related to both genetic and environmental factors. Comp Med 2010; 60:336–47.

469 16. Orcutt RP, Gianni FJ, Judge RJ. Development of an “altered” Schaedler flora for NCI gnotobiotic
470 rodents. Microecol Ther 1987; 17:59.

471 17. Schaedler RW, Dubos R, Costello R. The Development of the Bacterial Flora in the Gastrointestinal
472 Tract of Mice. J Exp Med 1965; 122:59–66.

473 18. Dewhirst FE, Chien CC, Paster BJ, Ericson RL, Orcutt RP, Schauer DB, Fox JG. Phylogeny of the
474 defined murine microbiota: altered Schaedler flora. Appl Environ Microb 1999; 65:3287–92.

475 19. Mandal RK, Denny JE, Waide ML, Li Q, Bhutiani N, Anderson CD, Baby BV, Jala VR, Egilmez NK,
476 Schmidt NW. Temporospatial shifts within commercial laboratory mouse gut microbiota impact
477 experimental reproducibility. Bmc Biol 2020; 18:83.

478 20. Hoy YE, Bik EM, Lawley TD, Holmes SP, Monack DM, Theriot JA, Relman DA. Variation in
479 Taxonomic Composition of the Fecal Microbiota in an Inbred Mouse Strain across Individuals and Time.
480 PloS one 2015; 10:e0142825.

481 21. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM,
482 FitzGerald MG, Fulton RS, et al. Structure, Function and Diversity of the Healthy Human Microbiome.
483 Nature 2012; 486:207–14.

484 22. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ,
485 Roe BA, Affourtit JP, et al. A core gut microbiome in obese and lean twins. Nature 2009; 457:480–4.

486 23. Hart ML, Ericsson AC, Franklin CL. Differing Complex Microbiota Alter Disease Severity of the IL-
487 10−/− Mouse Model of Inflammatory Bowel Disease. Front Microbiol 2017; 8:792.

488 24. Ericsson AC, Hart ML, Kwan J, Lanoue L, Bower LR, Araiza R, Lloyd KCK, Franklin CL. Supplier-
489 origin mouse microbiomes significantly influence locomotor and anxiety-related behavior, body
490 morphology, and metabolism. Commun Biology 2021; 4:716.

491 25. Velazquez EM, Nguyen H, Heasley KT, Saechao CH, Gil LM, Rogers AWL, Miller BM, Rolston MR,
492 Lopez CA, Litvak Y, et al. Endogenous Enterobacteriaceae underlie variation in susceptibility to
493 Salmonella infection. Nat Microbiol 2019; 4:1057–64.

494 26. Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, Wampach L, Schneider
495 JG, Hogan A, Beaufort C de, et al. Integrated multi-omics of the human gut microbiome in a case study
496 of familial type 1 diabetes. Nat Microbiol 2016; 2:16180.

27. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 2017; 35:725–31.

28. Busi SB, Bourquin M, Fodelianakis S, Michoud G, Kohler TJ, Peter H, Pramateftaki P, Styllas M, Tolosano M, Staercke VD, et al. Genomic and metabolic adaptations of biofilms to ecological windows of opportunity in glacier-fed streams. Nat Commun 2022; 13:2168.

29. Antunes TC, Marconatto L, Borges LG dos A, Giongo A, Sand STVD. Analysis of microbial community biodiversity in activated sludge from a petrochemical plant. Ambiente E Agua - Interdiscip J Appl Sci 2021; 16:1–22.

30. Lima CPV, Grisi DC, Guimarães MDCM, Salles LP, Kruly P de C, Do T, Borges LGDA, Dame-Teixeira N. Enrichment of sulphate-reducers and depletion of butyrate-producers may be hyperglycaemia signatures in the diabetic oral microbiome. J Oral Microbiol 2022; 14:2082727.

31. Saito D, Lemos LN, Ferreira ATRN, Saito CPB, Oliveira RF de, Cannavan F de S, Tsai SM. Draft Genome Sequences of Five Putatively Novel Saccharibacteria Species Assembled from the Human Oral Metagenome. Microbiol Resour Announc 2022; 11:e00246-22.

32. Mousavi SH, Motahar SFS, Salami M, Kavousi K, Mamaghani ASA, Ariaeenejad S, Salekdeh GH. In vitro bioprocessing of corn as poultry feed additive by the influence of carbohydrate hydrolyzing metagenome derived enzyme cocktail. Sci Rep-uk 2022; 12:405.

33. Chen Y-F, Hsieh A-H, Wang L-C, Huang Y-J, Tsai Y-C, Tseng W-Y, Kuo Y-L, Luo S-F, Yu K-H, Kuo C-F. Fecal microbiota changes in NZB/W F1 mice after induction of lupus disease. Sci Rep-uk 2021; 11:22953.

34. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017; 45:D353–61.

35. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28:27–30.

36. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 2016; 44:D279-85.

37. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 2009; 37:D233-8.

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

39. Lodowska J, Wolny D, Węglarz L. The sugar 3-deoxy-d-manno-oct-2-ulosonic acid (Kdo) as a characteristic component of bacterial endotoxin — a review of its biosynthesis, function, and placement in the lipopolysaccharide core. Can J Microbiol 2013; 59:645–55.

534 40. Cartmell A, Lowe EC, Baslé A, Firbank SJ, Ndeh DA, Murray H, Terrapon N, Lombard V, Henrissat
535 B, Turnbull JE, et al. How members of the human gut microbiota overcome the sulfation problem posed
536 by glycosaminoglycans. Proc National Acad Sci 2017; 114:7037–42.

537 41. Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee CA,
538 Lynch SV, et al. Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria. Cell 2009;
539 139:485–98.

540 42. Hart ML, Ericsson AC, Lloyd KCK, Grimsrud KN, Rogala AR, Godfrey VL, Nielsen JN, Franklin CL.
541 Development of outbred CD1 mouse colonies with distinct standardized gut microbiota profiles for use
542 in complex microbiota targeted studies. Sci Rep-uk 2018; 8:10107.

543 43. Sarkar A, Lehto SM, Harty S, Dinan TG, Cryan JF, Burnet PWJ. Psychobiotics and the
544 Manipulation of Bacteria-Gut-Brain Signals. Trends Neurosci 2016; 39:763–81.

545 44. Wang S, Ishima T, Zhang J, Qu Y, Chang L, Pu Y, Fujita Y, Tan Y, Wang X, Hashimoto K.
546 Ingestion of Lactobacillus intestinalis and Lactobacillus reuteri causes depression- and anhedonia-like
547 phenotypes in antibiotic-treated mice via the vagus nerve. J Neuroinflamm 2020; 17:241.

548 45. Bravo JA, Forsythe P, Chew MV, Escaravage E, Savignac HM, Dinan TG, Bienenstock J, Cryan JF.
549 Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in
550 a mouse via the vagus nerve. Proc Natl Acad Sci USA 2011; 108:16050–5.

551 46. Liang S, Wang T, Hu X, Luo J, Li W, Wu X, Duan Y, Jin F. Administration of Lactobacillus
552 helveticus NS8 improves behavioral, cognitive, and biochemical aberrations caused by chronic restraint
553 stress. Neuroscience 2015; 310:561–77.

554 47. Chatelier EL, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-
555 M, Kennedy S, et al. Richness of human gut microbiome correlates with metabolic markers. Nature
556 2013; 500:541–6.

557 48. Vila AV, Imhann F, Collij V, Jankipersadsing SA, Gurry T, Mujagic Z, Kurilshikov A, Bonder MJ,
558 Jiang X, Tigchelaar EF, et al. Gut microbiota composition and functional changes in inflammatory bowel
559 disease and irritable bowel syndrome. Sci Transl Med 2018; 10.

560 49. Schirmer M, Garner A, Vlamakis H, Xavier RJ. Microbial genes and pathways in inflammatory
561 bowel disease. Nat Rev Microbiol 2019; 17:497–511.

562 50. Moskowitz JE, Andreatta F, Amos-Landgraf J. The gut microbiota modulates differential adenoma
563 suppression by B6/J and B6/N genetic backgrounds in Apc(Min) mice. Mammalian genome : official
564 journal of the International Mammalian Genome Society 2019; 30:237–44.

565 51. Blachier F, Andriamihaja M, Larraufie P, Ahn E, Lan A, Kim E. Production of hydrogen sulfide by
566 the intestinal microbiota and epithelial cells and consequences for the colonic and rectal mucosa. Am J
567 Physiol-gastr L 2021; 320:G125–35.

568 52. Attene-Ramos MS, Wagner ED, Plewa MJ, Gaskins HR. Evidence that hydrogen sulfide is a
569 genotoxic agent. Molecular cancer research : MCR 2006; 4:9–14.

53. Guo FF, Yu TC, Hong J, Fang JY. Emerging Roles of Hydrogen Sulfide in Inflammatory and Neoplastic Colonic Diseases. Front Physiol 2016; 7:156.

54. Flannigan KL, Agbor TA, Motta JP, Ferraz JG, Wang R, Buret AG, Wallace JL. Proresolution effects of hydrogen sulfide during colitis are mediated through hypoxia-inducible factor-1alpha. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 2015; 29:1591–602.

55. Motta JP, Flannigan KL, Agbor TA, Beatty JK, Blackler RW, Workentine ML, Silva GJD, Wang R, Buret AG, Wallace JL. Hydrogen sulfide protects from colitis and restores intestinal microbiota biofilm and mucus production. Inflamm Bowel Dis 2015; 21:1006–17.

56. Wallace JL, Blackler RW, Chan MV, Silva GJD, Elsheikh W, Flannigan KL, Gamaniek I, Manko A, Wang L, Motta JP, et al. Anti-inflammatory and cytoprotective actions of hydrogen sulfide: translation to therapeutics. Antioxidants & redox signaling 2015; 22:398–410.

57. Hsu CN, Hou CY, Chang-Chien GP, Lin S, Tain YL. Maternal N-Acetylcysteine Therapy Prevents Hypertension in Spontaneously Hypertensive Rat Offspring: Implications of Hydrogen Sulfide-Generating Pathway and Gut Microbiota. Antioxidants 2020; 9:856.

58. Feil EJ, Holmes EC, Bessen DE, Chan MS, Day NP, Enright MC, Goldstein R, Hood DW, Kalia A, Moore CE, et al. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. Proc Natl Acad Sci USA 2001; 98:182–7.

59. Duchene S, Holt KE, Weill FX, Hello SL, Hawkey J, Edwards DJ, Fourment M, Holmes EC. Genome-scale rates of evolutionary change in bacteria. Microb Genom 2016; 2:e000094.

60. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nature reviews Microbiology 2016; 14:150–62.

61. Garud NR, Good BH, Hallatschek O, Pollard KS. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. Plos Biol 2019; 17:e3000102.

62. Young BC, Wu CH, Gordon NC, Cole K, Price JR, Liu E, Sheppard AE, Perera S, Charlesworth J, Golubchik T, et al. Severe infections emerge from commensal bacteria by adaptive evolution. eLife 2017; 6:e30637.

63. Yilmaz B, Mooser C, Keller I, Li H, Zimmermann J, Bosshard L, Fuhrer T, Aguero MG de, Trigo NF, Tschanz-Lischer H, et al. Long-term evolution and short-term adaptation of microbiota strains and sub-strains in mice. Cell host & microbe 2021; 29:650-663 e9.

64. Bor B, Poweleit N, Bois JS, Cen L, Bedree JK, Zhou ZH, Gunsalus RP, Lux R, McLean JS, He X, et al. Phenotypic and Physiological Characterization of the Epibiotic Interaction Between TM7x and Its Basibiont Actinomyces. Microbial Ecol 2016; 71:243–55.

65. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu SY, Dorrestein PC, Esquenazi E, Hunter RC, Cheng G, et al. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl Acad Sci USA 2015; 112:244–9.

606 66. Soro V, Dutton LC, Sprague SV, Nobbs AH, Ireland AJ, Sandy JR, Jepson MA, Micaroni M, Splatt
607 PR, Dymock D, et al. Axenic culture of a candidate division TM7 bacterium from the human oral cavity
608 and biofilm interactions with other oral bacteria. Appl Environ Microbiol 2014; 80:6480–9.

609 67. Dinis JM, Barton DE, Ghadiri J, Surendar D, Reddy K, Velasquez F, Chaffee CL, Lee MC,
610 Gavrilova H, Ozuna H, et al. In search of an uncultured human-associated TM7 bacterium in the
611 environment. PloS one 2011; 6:e21280.

612 68. Narayanasamy S, Jarosz Y, Muller EEL, Heintz-Buschart A, Herold M, Kaysen A, Laczny CC, Pinel
613 N, May P, Wilmes P. IMP: a pipeline for reproducible reference-independent integrated metagenomic
614 and metatranscriptomic analyses. Genome Biol 2016; 17:260.

615 69. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large
616 and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 2015; 31:1674–6.

617 70. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. MetaBAT 2: an adaptive binning
618 algorithm for robust and efficient genome reconstruction from metagenome assemblies. Peerj 2019;
619 7:e7359.

620 71. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover
621 genomes from multiple metagenomic datasets. Bioinformatics 2016; 32:605–7.

622 72. Hickl O, Queirós P, Wilmes P, May P, Heintz-Buschart A. binny: an automated binning algorithm to
623 recover high-quality genomes from complex metagenomic datasets. Biorxiv 2022; :2021.12.22.473795.

624 73. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. Recovery of
625 genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 2018;
626 3:836–43.

627 74. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of
628 microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 2015;
629 25:1043–55.

630 75. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with
631 the Genome Taxonomy Database. Bioinformatics 2019; 36:1925–7.

632 76. Krinos A, Hu S, Cohen N, Alexander H. EUKulele: Taxonomic annotation of the unsung eukaryotic
633 microbes. J Open Source Softw 2021; 6:2817.

634 77. Queirós P, Delogu F, Hickl O, May P, Wilmes P. Mantis: flexible and consensus-driven genome
635 annotation. Gigascience 2021; 10:giab042.

636 78. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. Anvi'o: an
637 advanced analysis and visualization platform for 'omics data. Peerj 2015; 3:e1319.

638 79. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space
639 complexity. Bmc Bioinformatics 2004; 5:113.

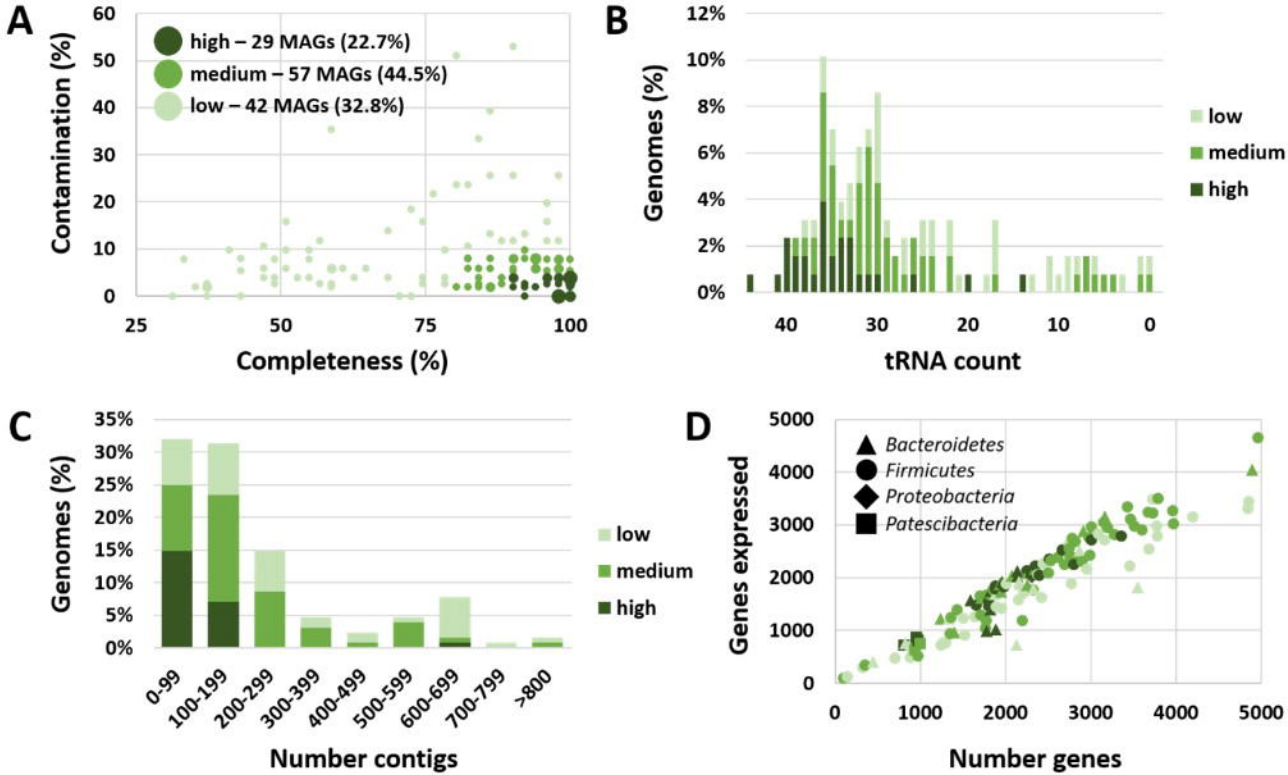
80. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. Plos One 2010; 5:e9490.

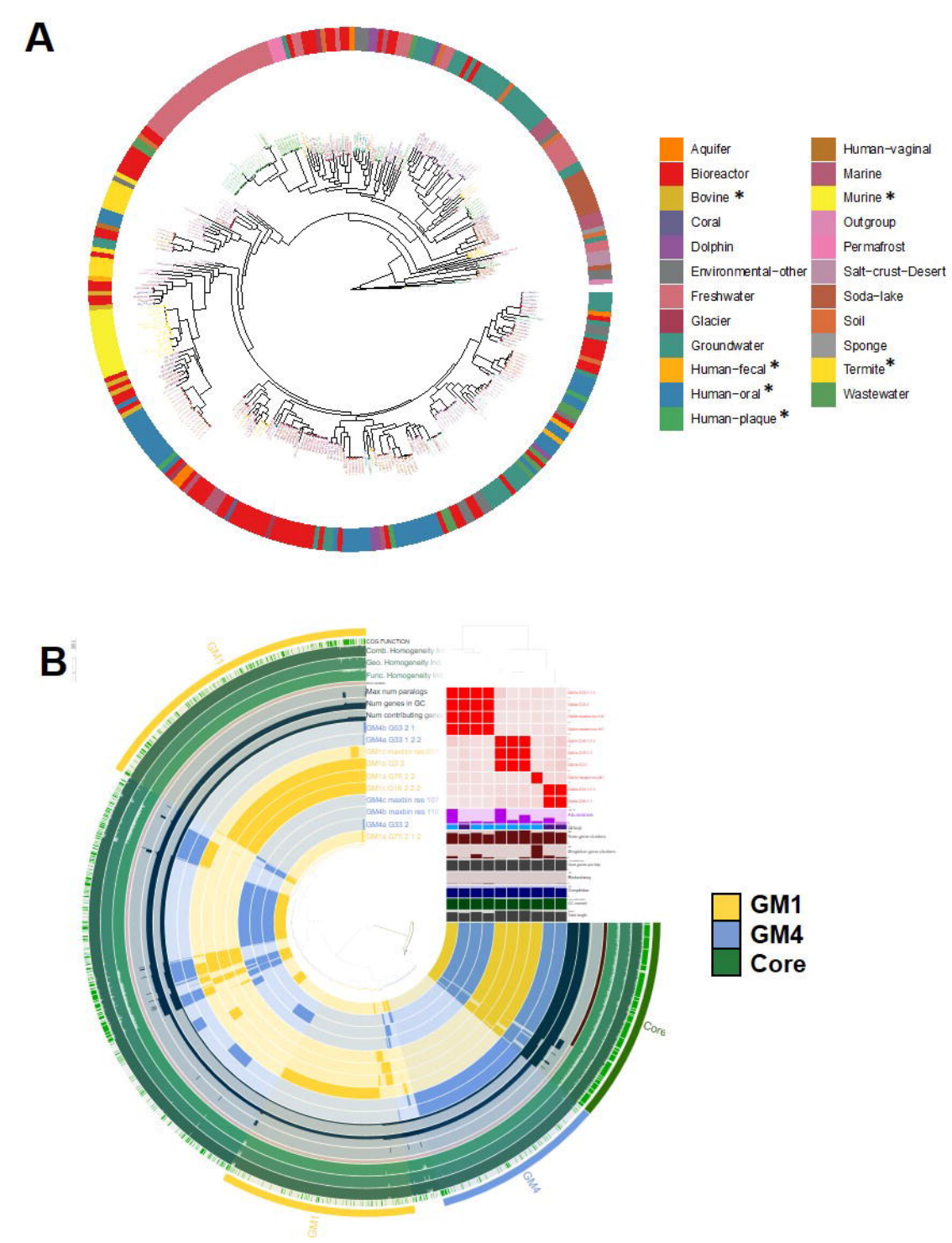
81. Lee MD. GToTree: a user-friendly workflow for phylogenomics. Bioinform Oxf Engl 2019; 35:4162–4.

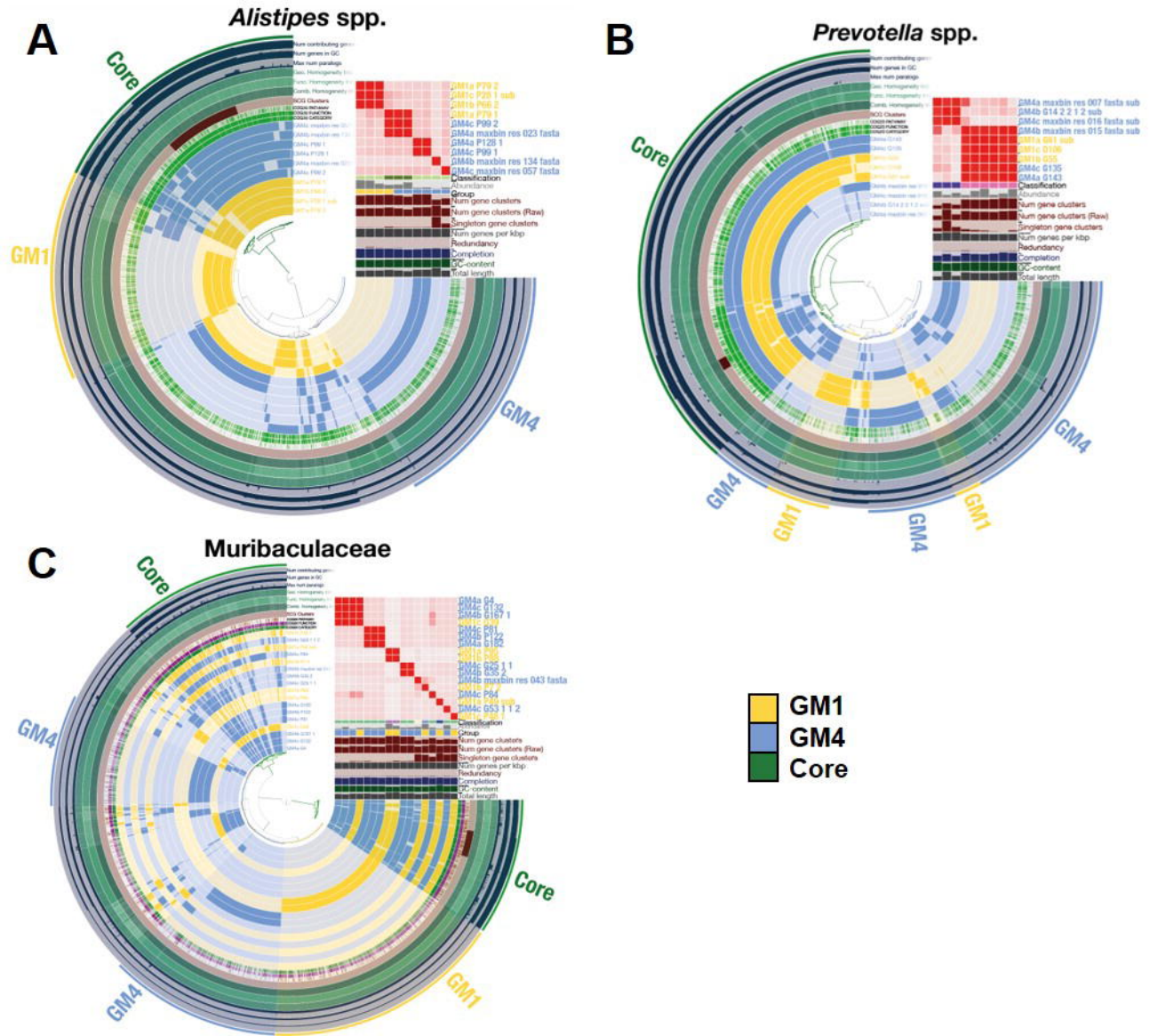
82. Eddy SR. Accelerated Profile HMM Searches. Plos Comput Biol 2011; 7:e1002195.

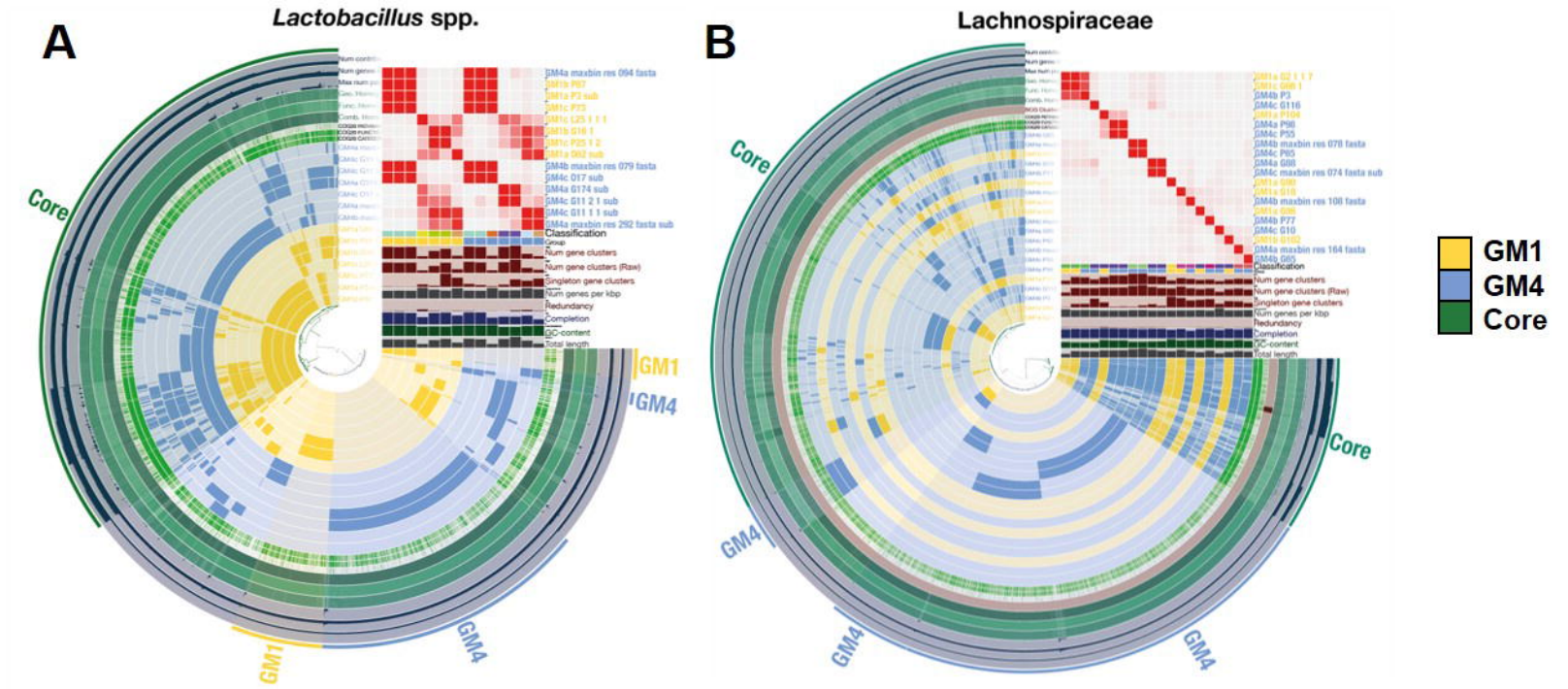
83. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. Bmc Bioinformatics 2010; 11:119.

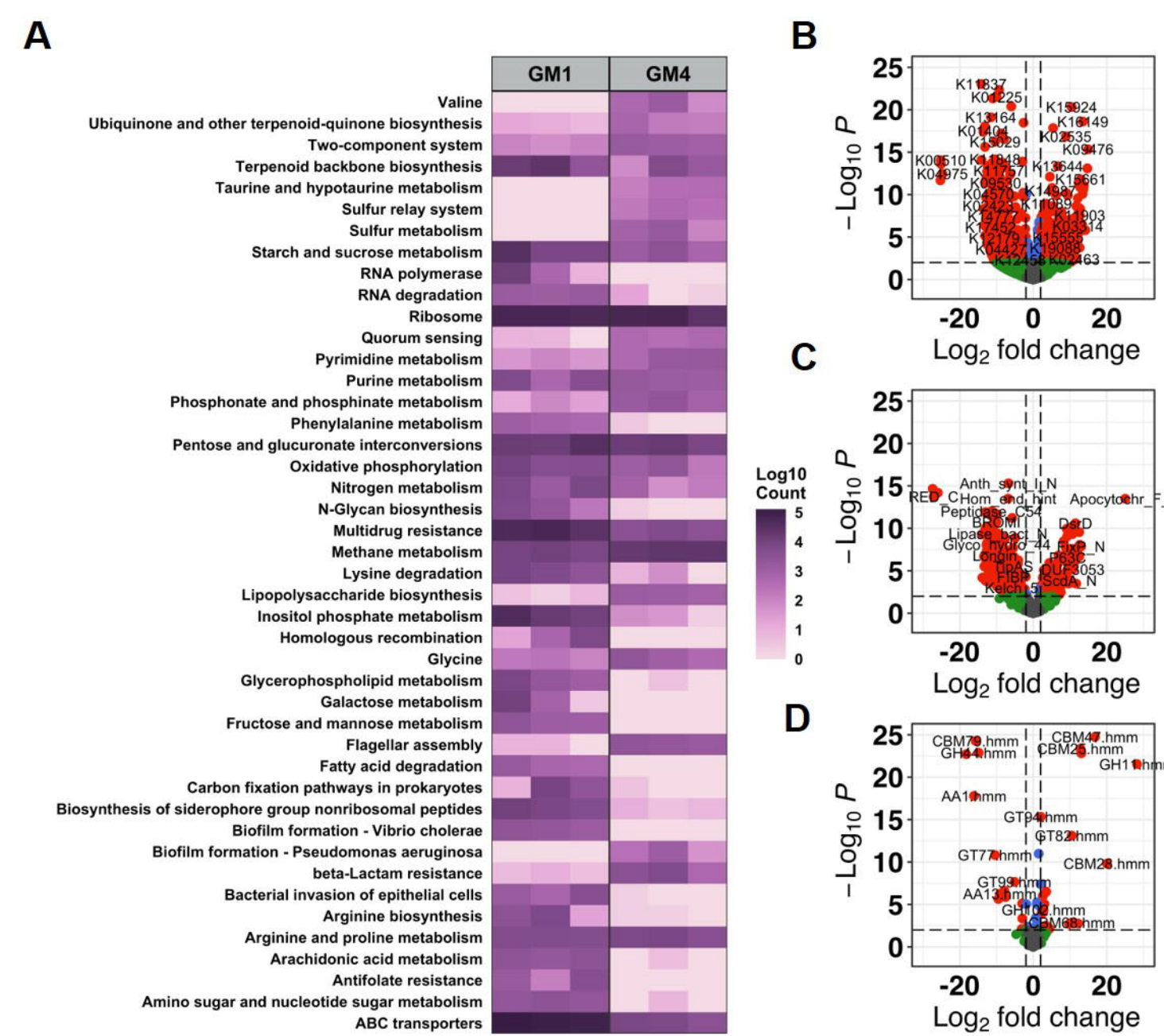
84. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 2009; 25:1972–3.

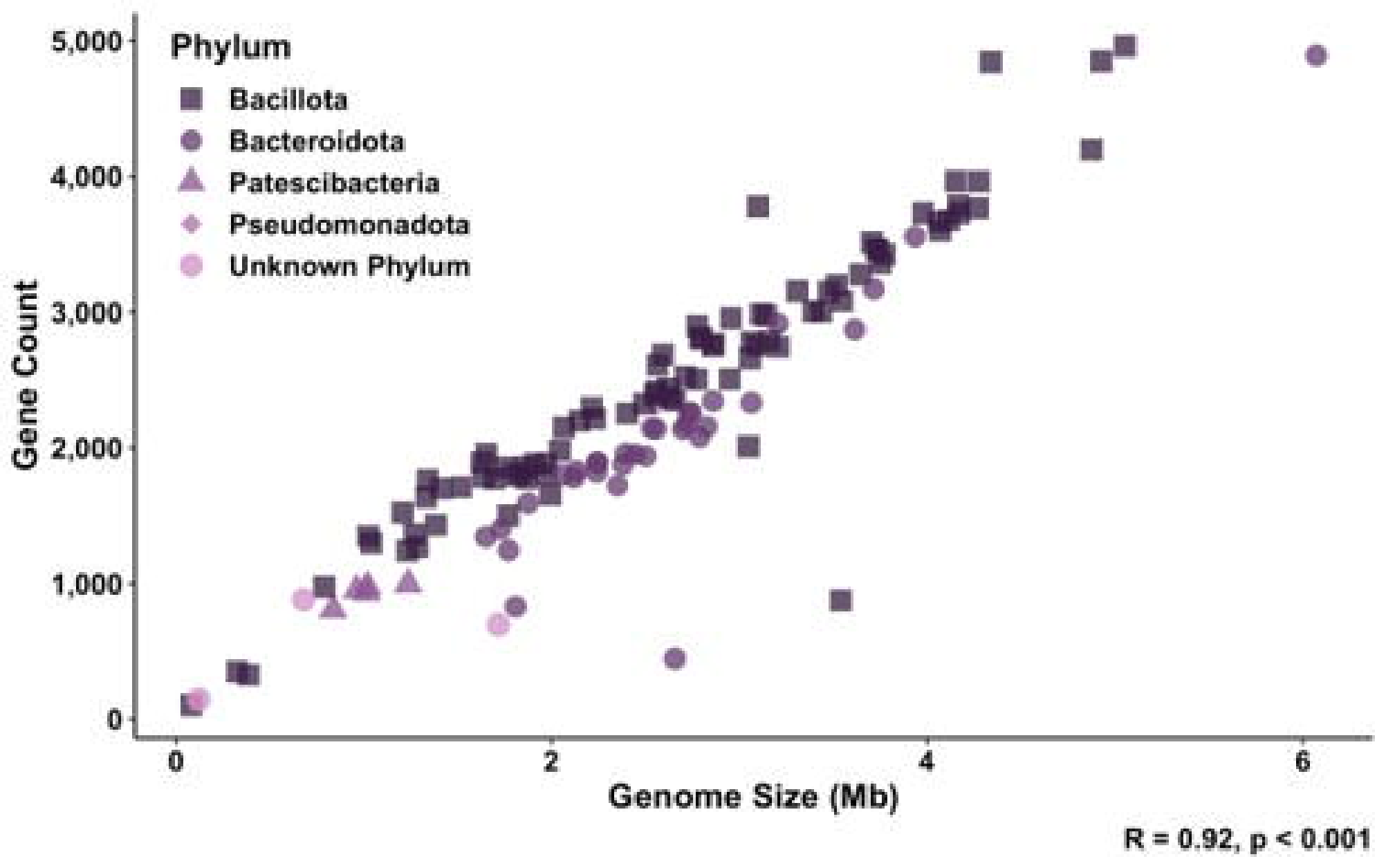

A
high – 29 MAGs (22.7%)
medium – 57 MAGs (44.5%)
low – 42 MAGs (32.8%)
Contamination (%)
Completeness (%)
B
Genomes (%)
tRNA count
low
medium
high
C
Genomes (%)
Number contigs
0-99
100-199
200-299
300-399
400-499
500-599
600-699
700-799
>800
D
Bacteroidetes
Firmicutes
Proteobacteria
Patescibacteria
Genes expressed
Number genes

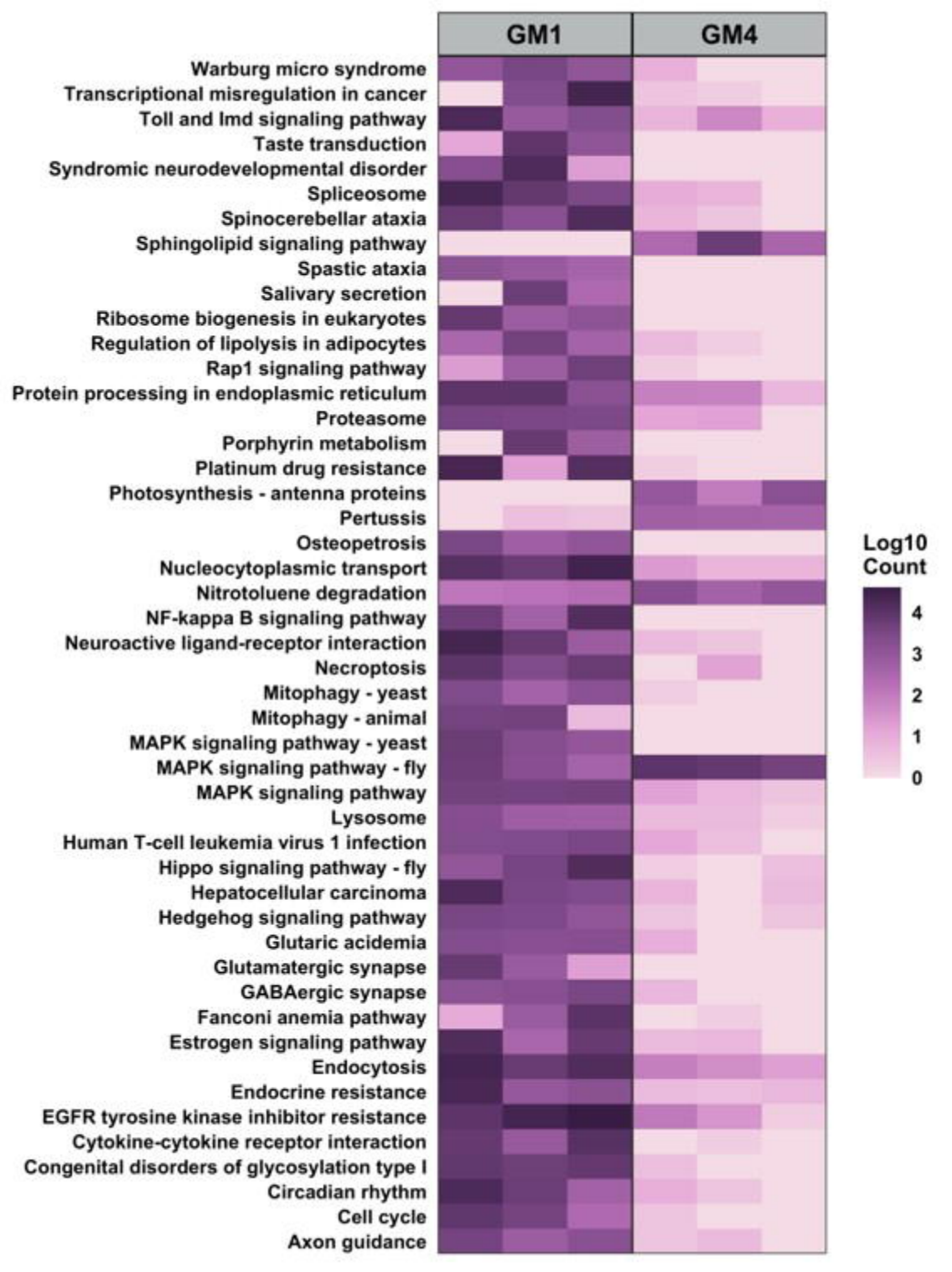

A
Aquifer
Bioreactor
Bovine *
Coral
Dolphin
Environmental−other
Freshwater
Glacier
Groundwater
Human−fecal *
Human−oral *
Human−plaque *
Human−vaginal
Marine
Murine *
Outgroup
Permafrost
Salt−crust−Desert
Soda−lake
Soil
Sponge
Termite *
Wastewater
B
GM1
GM4
Core

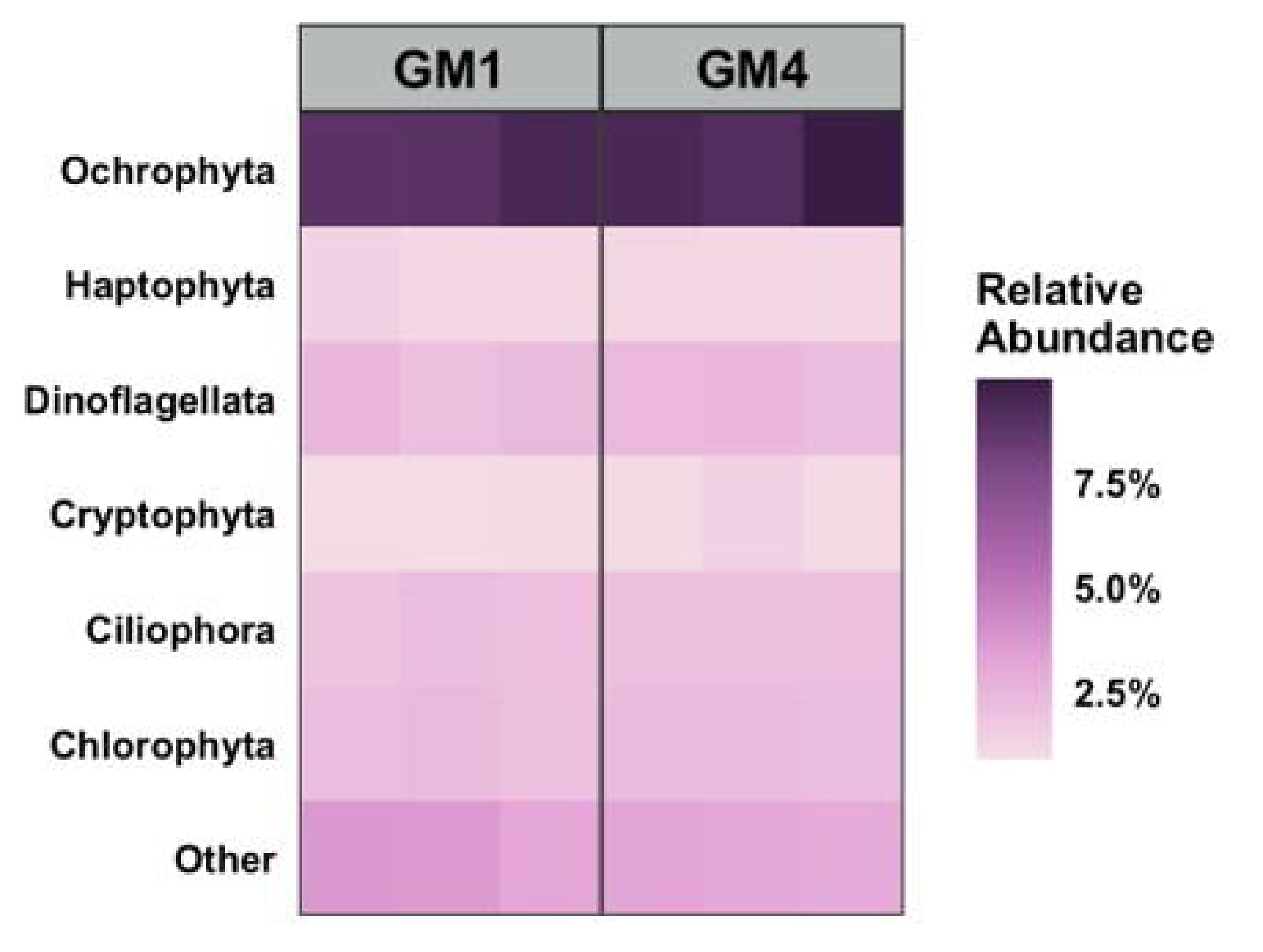

A
Alistipes spp.
Core
GM1
GM4
Classification
Group
Num gene clusters
Num gene clusters (Raw)
Singleton gene clusters
Num genes per kbp
Redundancy
Completion
GC-content
Total length
B
Prevotella spp.
Core
GM4
GM1
GM4
GM1
GM4
Classification
Num gene clusters
Num gene clusters (Raw)
Singleton gene clusters
Num genes per kbp
Redundancy
Completion
GC-content
Total length
C
Muribaculaceae
Core
GM4
Core
GM4
GM1
Classification
Group
Num gene clusters
Num gene clusters (Raw)
Singleton gene clusters
Num genes per kbp
Redundancy
Completion
GC-content
Total length
GM1
GM4
Core


A
Lactobacillus spp.
Core
GM1
GM4
B
Lachnospiraceae
Core
GM4
GM4
GM4
Core
GM1
GM4
Core


A
GM1
GM4
Valine
Ubiquinone and other terpenoid-quinone biosynthesis
Two-component system
Terpenoid backbone biosynthesis
Taurine and hypotaurine metabolism
Sulfur relay system
Sulfur metabolism
Starch and sucrose metabolism
RNA polymerase
RNA degradation
Ribosome
Quorum sensing
Pyrimidine metabolism
Purine metabolism
Phosphonate and phosphinate metabolism
Phenylalanine metabolism
Pentose and glucuronate interconversions
Oxidative phosphorylation
Nitrogen metabolism
N-Glycan biosynthesis
Multidrug resistance
Methane metabolism
Lysine degradation
Lipopolysaccharide biosynthesis
Inositol phosphate metabolism
Homologous recombination
Glycine
Glycerophospholipid metabolism
Galactose metabolism
Fructose and mannose metabolism
Flagellar assembly
Fatty acid degradation
Carbon fixation pathways in prokaryotes
Biosynthesis of siderophore group nonribosomal peptides
Biofilm formation - Vibrio cholerae
Biofilm formation - Pseudomonas aeruginosa
beta-Lactam resistance
Bacterial invasion of epithelial cells
Arginine biosynthesis
Arginine and proline metabolism
Arachidonic acid metabolism
Antifolate resistance
Amino sugar and nucleotide sugar metabolism
ABC transporters
Log10 Count
5
4
3
2
1
0
B
C
D
−Log10 P
Log2 fold change
25
20
15
10
5
0
-20
0
20
K11837
K01225
K13164
K01404
K15029
K00510
K11848
K04975
K11757
K09530
K04570
K02423
K14777
K17452
K12179
K04427
K12458
K15924
K16149
K02535
K09476
K13644
K15661
K14987
K11089
K11903
K03314
K15555
K19088
K02463
Anth_synt_I_N
RED_CHom
end_hint
Apocytochr_F
Peptidase_C54
BROMI
DsrD
Lipase_bact_N
Glyco_hydro_44
FixP_N
Longin
Fe3C
LpxS
DUF3053
FlbB
ScdA_N
Kelch_5
CBM79.hmm
CBM47.hmm
GH44.hmm
CBM25.hmm
GH11.hmm
AA1.hmm
GT94.hmm
GT82.hmm
GT77.hmm
CBM28.hmm
GT99.hmm
AA13.hmm
GH102.hmm
CBM68.hmm


Phylum
Bacillota
Bacteroidota
Patescibacteria
Pseudomonadota
Unknown Phylum
Gene Count
5,000
4,000
3,000
2,000
1,000
0
0
2
4
6
Genome Size (Mb)
R = 0.92, p < 0.001


GM1
GM4
Warburg micro syndrome
Transcriptional misregulation in cancer
Toll and Imd signaling pathway
Taste transduction
Syndromic neurodevelopmental disorder
Spliceosome
Spinocerebellar ataxia
Sphingolipid signaling pathway
Spastic ataxia
Salivary secretion
Ribosome biogenesis in eukaryotes
Regulation of lipolysis in adipocytes
Rap1 signaling pathway
Protein processing in endoplasmic reticulum
Proteasome
Porphyrin metabolism
Platinum drug resistance
Photosynthesis - antenna proteins
Pertussis
Osteopetrosis
Nucleocytoplasmic transport
Nitrotoluene degradation
NF-kappa B signaling pathway
Neuroactive ligand-receptor interaction
Necroptosis
Mitophagy - yeast
Mitophagy - animal
MAPK signaling pathway - yeast
MAPK signaling pathway - fly
MAPK signaling pathway
Lysosome
Human T-cell leukemia virus 1 infection
Hippo signaling pathway - fly
Hepatocellular carcinoma
Hedgehog signaling pathway
Glutaric acidemia
Glutamatergic synapse
GABAergic synapse
Fanconi anemia pathway
Estrogen signaling pathway
Endocytosis
Endocrine resistance
EGFR tyrosine kinase inhibitor resistance
Cytokine-cytokine receptor interaction
Congenital disorders of glycosylation type I
Circadian rhythm
Cell cycle
Axon guidance
Log10 Count
4
3
2
1
0


GM1
GM4
Ochrophyta
Haptophyta
Dinoflagellata
Cryptophyta
Ciliophora
Chlorophyta
Other
Relative Abundance
7.5%
5.0%
2.5%

A
Eukaryota
CBM
CE
GH
GT
PL
CBM
CE
GH
GT
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1
B
Eukaryota
CBM
CE
GH
GT
PL
CBM
CE
GH
GT
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1