RESEARCH



Systems genetics uncovers associations among host amylase locus, gut microbiome, and metabolic traits in mice



Qijun Zhang¹, Evan R. Hutchison¹, Calvin Pan^{2,3}, Matthew F. Warren¹, Mark P. Keller⁴, Alan D. Attie⁴, Aldons J. Lusis^{2,3} and Federico E. Rey^{1,5*}

Abstract

Background Population studies have revealed associations between host genetic and gut microbiome in humans and mice. However, the molecular bases for how host genetic variation impacts the gut microbial community and bacterial metabolic niches remain largely unknown.

Results We leveraged 90 inbred hyperlipidemic mouse strains from the hybrid mouse diversity panel (HMDP), previously studied for a variety of cardio-metabolic traits. Metagenomic analysis of cecal DNA followed by genome-wide association analysis identified genomic loci that were associated with microbial enterotypes in the gut. Among these, we detected a genetic locus surrounding multiple amylase genes that were associated with abundances of Firmicutes (*Lachnospiraceae* family) and Bacteroidetes (*Muribaculaceae* family) taxa encoding distinct starch and sugar degrading capabilities. The genetic variants at the amylase gene locus were associated with distinct gut microbial communities (enterotypes) with different predicted metabolic capacities for carbohydrate degradation. Mendelian randomization analysis revealed host phenotypes, including liver fibrosis and plasma HDL-cholesterol levels, that were associated with gut microbiome enterotypes.

Conclusions This work reveals novel relationships among host genetic variation, gut microbial enterotypes, and host metabolic traits and supports the notion that variation of host amylase may represent a key determinant of gut microbiome in mice.

Keywords Systems genetics, Gut microbiome, Genetic association study, Metagenomics, Hybrid mouse diversity panel

*Correspondence:

Introduction

The microbial communities that inhabit the gut of mammals have profound effects on the biology and health of their hosts. Alterations in the intestinal microbiome have been associated with myriad conditions, including metabolic disorders, cardiovascular disease, [1, 2] and autoimmune and inflammatory disorders [3, 4]. Host genetic variation and environmental factors, including diet, modulate the gut microbiome and its interactions with the host [5, 6]. Distinct gut microbial communities, termed enterotypes, which reflect stratification in a population and defined compositional attributes, have been



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Federico E. Rey

ferey@wisc.edu

¹ Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA

² Division of Cardiology, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

³ Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA, USA

 $^{^{\}rm 4}$ Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

⁵ Department of Medical Microbiology and Immunology, University of Wisconsin-Madison, Madison, WI, USA

identified [7, 8]. However, it remains largely unknown how host genetics modulate microbial enterotypes, and the metabolic functions that contribute to this stratification have not been established.

Carbohydrates represent an important energy source for both human and microbial cells. Consumption of different carbohydrates can influence the gut microbiota and its association with the host [9, 10]. Non-digestible carbohydrates, including many plant polysaccharides, such as resistant starch, are not digested by host enzymes and can therefore serve as substrates for microbial growth in the distal gut. Digestible carbohydrates like starch, on the other hand, are broken down by host amylases [11], releasing their constituent sugars which can be absorbed by the host. Amylase gene copy number varies among individual humans and associations between amylase copy number and diet have been reported for mammals [11-14]. In mice, the genes encoding salivary (*Amy1*) and pancreatic (*Amy2*) amylases are located in a gene cluster on chromosome 3. Previous work showed that the copy number of the Amy2 paralogous gene (Amy2a1, Amy2a2, Amy2a3, Amy2a4, Amy2a5, Amy2b) varies across different mouse strains [13]. Furthermore, a recent study revealed that humans with higher salivary amylase gene (AMY1) copy number harbor gut microbiomes with increased abundance of starch-degrading microbes suggesting that genetic variants of host amylase gene locus may potentially impact gut microbiome and its subsequent effects on the host [15]. However, environmental factors including diet, which is known to influence gut microbiome, are difficult to control in human studies.

Previous studies in mice and humans have investigated the impact of host genetic variation on gut microbiota composition. These efforts mostly applied 16S rRNA gene sequencing [16–18]. Shotgun metagenomics allows comprehensive characterization of the functions and metabolic pathways present in gut communities. However, a limited number of studies have applied metagenomic characterization of gut microbiome in large-scale genetically diverse, phenotypically characterized cohorts [19]. A valuable resource for such an undertaking is the hybrid mouse diversity panel (HMDP), which consists of over 100 common inbred and recombinant inbred strains. In order to study cardio-metabolic traits, each strain from this panel was made hyperlipidemic by transgenic expression of human apolipoprotein E-Leiden (APOE-Leiden) and human cholesteryl ester transfer protein (CETP). This was done by breeding each of the HMDP strains to a C57BL/6 J background strain that possessed these two hyperlipidemia-inducing transgenes. Genetic differences among these animals arise only from sequence variations present in the individual recipient strains. This set of F1 animals was termed Ath-HMDP and represents a rich resource for studies exploring complex interactions underlying atherosclerosis and liver fibrosis [20–22].

Here, we analyzed gut metagenomes from 90 Ath-HMDP strains and used systems genetics to explore links between gut microbiome composition, host genetic variation, and cardiometabolic disease. Our analysis identified three distinct microbial enterotypes and uncovered two host genomic loci significantly associated with their abundance. Additionally, we identified bacterial functions within these host-selected enterotypes that may influence key biomarkers, such as plasma cholesterol and triglycerides, and disease phenotypes including liver fibrosis.

Results

Characterization of the gut metagenomes from 90 Ath-HMDP mouse strains

We generated metagenomic datasets from DNA isolated from cecal contents collected from 356 F1 Ath-HMDP mice encompassing 90 strains (males and females were included for most strains) fed a high-fat diet (33% kcal from fat) with 1% cholesterol for 16 weeks (37.8 million paired-end reads/sample). The generation of this mouse cohort has been previously described [20], including metabolic and disease-associated phenotypes such as plasma lipids, glucose levels, atherosclerotic lesions, and liver fibrosis, all of which varied widely among strains [20, 21] (SupplementaryTable 1).

Phylogenetic and functional analyses of these metagenomes identified 459 bacterial taxa encompassing 7 phyla, 43 classes, 49 orders, 63 families, 131 genera, and 166 species; 2127 KO (KEGG Orthology) functions and 294 metabolic pathways (MetaCyc database) across all mice. Gut microbiota composition was highly variable across the 90 strains; for example, the relative abundance of Firmicutes phylum ranged from 10 to 55%, whereas the relative abundance of Bacteroidetes ranged from 5 to 77% (Supplementary Fig. 1A). The most abundant species (>0.5%) that were present in at least 90% of mice included the Paramuribaculum intestinale, Muribaculum intestinale, Muribaculum gordoncarteri, Duncaniella muris, Duncaniella freteri, Alistipes finegoldii, Parabacteroides goldsteinii, Faecalibaculum rodentium, Lachnospiraceae bacterium, Desulfovibrio SGB41239, Muribaculaceae SGB35953, Bilophila sp., and Parasutterella sp. (Supplementary Fig. 1B, Supplementary Table 2). We defined these taxa as the "core microbiome species" in the Ath-HMDP mice gut.

Enterotypes of the Ath-HMDP gut microbiome

Principal coordinate analysis (PCoA) of the microbial communities described above using Bray–Curtis distance of species abundance resulted in three clusters



Fig. 1 Gut microbial enterotypes and bacterial co-abundance groups. **A** Clustering identified three microbial enterotypes in the Ath-HMDP mouse cohort. **B** Relative abundances of Firmicutes, Bacteroidetes, and Verrucomicrobia in the three enterotypes. **C** Correlation network of bacterial species (average relative abundance > 0.1% and present in at least 20% of samples) using CCREPE with a checkerboard score, indicating the co-occurrence or co-exclusion between species. Nodes represent the species and lines represent the similarity score. Solid lines represent co-occurrence bacterial species and dashed lines represent co-exclusion bacterial species. Significance was calculated by unpaired two-tailed Wilcoxon signed-rank test and is designated as follows: ***p* value < 0.001; *****p* value < 0.0001. ET-B, Bacteroidetes enterotype; ET-F, Firmicutes enterotype; ET-V, Verrucomicrobia enterotype

each dominated by a different phylum: (i) Firmicutes, (ii) Bacteroidetes, and (iii) Verrucomicrobia, respectively (Supplementary Fig. 2A). Because this stratification of microbial composition was observed, we next explored the existence of different enterotypes. Using partitioning around medoid (PAM) clustering of Bray– Curtis distance of species abundance, we detected three enterotypes (Fig. 1A), each of which was identifiable by the levels of Firmicutes, Bacteroidetes, and Verrucomicrobia respectively (Fig. 1B). Previous studies reported distinct enterotypes in the human gut that were dominated by *Bacteroides, Prevotella*, and *Ruminococcaceae* [7, 8]. We observed clusters dominated by different taxa as compared to humans, which is likely due to the anatomical, physiological, and biochemical differences in the gastrointestinal tract and the distinct overall gut microbiota composition detected in these two mammals.

Using microbial function abundance resulted in similar clustering each dominated by one of the three enterotype phyla mentioned above (Supplementary Fig. 2B). The first principal component (PC1) of the microbial functions, which explains the most variance for the functional profiles, was highly correlated with Firmicutes abundance (Spearman's ρ =0.86, *P*=7.87×10⁻¹⁰³) and Bacteroidetes abundance (Spearman's ρ =0.67, *P*=1.57×10⁻⁴⁷) (Supplementary Fig. 2C). The functions' PC1 was also highly correlated with core species including *Lachnospiraceae bacterium* (Spearman's ρ =0.72, *P*=1.29×10⁻⁵⁶), *Desulfovibrio SGB41239* (Spearman's ρ =0.7, *P*=2.36×10⁻⁵²), *Muribaculum gordoncarteri* (Spearman's ρ = -0.46,

 $P = 4.37 \times 10^{-20}$) and *Muribaculaceae SGB35953* (Spearman's $\rho = -0.49$, $P = 2.77 \times 10^{-22}$) (Supplementary Table 3).

To identify bacterial species that shape the enterotypes, we summarized the most abundant and prevalent species (relative abundance > 0.1% and present at least > 20% mice) into co-abundance groups (CAGs) using co-occurrence scores (Fig. 1C). The Bacteroidetes-dominated CAG contained mostly Bacteroidetes taxa including Muribaculaceae, the most abundant family in mouse gut [23] (Duncaniella muris, Duncaniella freteri, Duncaniella dubosii, Muribaculaceae CAG-495 sp., Muribaculaceae CAG-873 sp.), Bacteroides genus (Bacteroides acidifaciens, Bacteroides sp.), and Alistipes sp. The Firmicutesdominated CAG consisted mostly of Firmicutes taxa including Lachnospiraceae family (Roseburia sp., Dorea sp., Acetatifactor sp., Sporofaciens sp., Kineothrix sp., and Schaedlerella arabinosiphila), Oscillospiraceae family (Lawsonibacter sp., Enterenecus sp., and Pelethomonas sp.), and Proteobacteria including Bilophila sp.

Gut microbiome features are associated with host genetics

To identify associations between mouse genomic variation and gut microbiome features, we performed genome-wide association analysis and mapped the abundance of bacterial functions, taxa, and metabolic pathways to the mouse genome (Supplementary Fig. 3). We observed 646 functions, 138 species, and 109 pathways that were significantly associated with at least one locus by genome-wide significance cutoff ($P < 4 \times 10^{-6}$). We further examined the density of these significantly associated gut microbial traits over the whole mouse genome and identified a GWAS hotspot on chromosome 3 at 113-115 Mbp. This genomic locus was associated with 347 different bacterial functions. The three enterotypes were distinguished at microbial functions' PC1 and PC2 levels (Fig. 2A); thus, we reasoned that the genetic variants at chromosome 3 were associated with microbial enterotypes.

The genomic regions identified by GWAS most likely contain candidate genes/genes that are in strong linkage disequilibrium (LD) with the lead SNP. Nine proteincoding genes are in the LD region (determined by correlation r^2 with lead SNP > 0.8) at chromosome 3 GWAS hotspot, including mouse amylase cluster genes (*Amy1*, *Amy2a1*, *Amy2a2*, *Amy2a3*, *Amy2a4*, and *Amy2a5*), *Rnpc3*, *Col11a1*, and *Olfm3* (Fig. 2B). Gene expression data were available from a subset (n=4-8 female mice/ strain; 40 strains) of the same Ath-HMDP mice [20]. We conducted a correlation analysis between microbial functions' PC1 and PC2 with each gene within the chromosome 3 GWAS hotspot. We found significant correlations for microbial functions' PC1 with *Amy1* and *Rnpc3* gene expression levels (Supplementary Fig. 4A). The microbial functions' PC2 was not significantly correlated with expression levels of candidate genes. As mentioned above, the microbial functions' PC1 was associated with Firmicutes and Bacteroidetes enterotypes (Supplementary Fig. 2C); we also found that two major families in these enterotypes, *Muribaculaceae* and *Lachnospiraceae*, were significantly correlated with *Amy1* gene but not *Rnpc3* gene (Supplementary Fig. 4B). In humans, the salivary amylase gene (*AMY1*) copy number is associated with nearby SNPs, structural haplotypes of the amylase locus [24], and gut microbiome composition [15]. Our data suggest that genomic variants of the amylase gene locus in mice are associated with bacterial taxa that drive the Firmicutes and Bacteroidetes enterotypes.

We found that microbial functions' PC1 mapped to the same hotspot locus, as expected (Fig. 2C). The lead SNP rs31001780 has two alleles, A and G, which differed in prevalence between enterotypes (Fig. 2D, Supplementary Table 6). More specifically, mice with the allele A at SNP rs31001780 had higher Firmicutes and lower Bacteroidetes levels in the gut (Fig. 2E). The prevalence of allele A and allele G was similar in Firmicutes enterotype and Verrucomicrobia enterotype mice, but the allele G was more prevalent in Bacteroidetes enterotype mice (Fig. 2F) suggesting genomic variants of the amylase gene locus in mice are likely associated with bacteria species that drive the Bacteroidetes enterotype such as *Muribaculacea*.

We found that A. muciniphila is the only bacterial species detected within the Verrucomicrobia phylum; we then used the abundance of A. muciniphila species as the representative microbial feature to identify genetic locus associated with the Verrucomicrobia enterotype (SupplementaryFig. 5A). The significant SNPs were on chromosome 1, and this locus has the lead SNP rs31965376 with two alleles, A and T, which differed in prevalence between enterotypes (Supplementary Fig. 5B, Supplementary Table 6). Mice with the allele A at SNP rs31965376 had on average higher Verrucomicrobia levels in their gut (Supplementary Fig. 5C). The prevalence of allele A and allele T was similar in Firmicutes and Bacteroidetes enterotypes mice, but the allele A was more prevalent in Verrucomicrobia enterotype mice (Supplementary Fig. 5D).

Bacterial carbohydrate metabolism potentially mediates host amylase and gut microbial enterotypes

To explore potential bacterial functions and pathways that were modulated by enterotype-associated SNPs, we examined the predicted metabolic differences between the two identified CAGs. We first performed Spearman's correlation between the aggregated abundance of Firmicutes/Bacteroidetes CAGs with MetaCyc pathways



Fig. 2 Host genomic loci are associated with gut microbial enterotypes. **A** PCA plot of microbial functions with the assignment of gut microbial enterotypes for each mouse. **B** SNP associations for microbial functions' PC1 on chromosome 3. Protein coding genes are displayed for Chr3: 112–116 Mbp region. **C** Genomic locus at Chr3: 113–114 Mbp is associated with microbial functions' PC1. The lead SNP is an intergenic SNP rs31001780. Dashed lines represent significance thresholds determined by permutation tests ($P < 4 \times 10^{-6}$). **D** Mice with alleles AA or GG at SNP rs31001780 are visualized in PCoA of microbial species beta-diversity (left) and PCA of microbial functions (right). **E** Relative abundance of Firmicutes, Bacteroidetes, and Verrucomicrobia from mice with AA or GG at SNP rs31001780. **F** Number of mice that have AA or GG at SNP rs31001780 in each of three enterotypes. Statistical difference between groups was tested by unpaired two-tailed Wilcoxon signed-rank test

(present in at least 90% of mice). PWY-6737 (starch degradation V) is one of top pathways positively associated with Bacteroidetes CAGs but negatively associated with Firmicutes CAGs (adjusted P < 0.05) suggesting it may mediate the association between host amylase locus with Firmicutes and Bacteroidetes enterotypes (Supplementary Fig. 6A). We then compared functions from starch and sugar metabolism pathway with the species abundance from the two CAGs and noted that genes encoding enzymes predicted to target starch and other polysaccharides degradation were positively associated with Bacteroides CAGs and negatively associated with Firmicutes CAGs (Supplementary Fig. 6B). Glycosidases are enzymes that break down glycosidic linkages to liberate monosaccharides and oligosaccharides of lower molecular weight from carbohydrate substrates. We found glycosidases predicted to target starch and other polysaccharides, including alpha-amylase (K07405, predicted to degrade starch), endoglucanase (K01179, predicted to degrade cellulose), dextranase (K05988, predicted to degrade dextran), and pullulanase (K01200, predicted to degrade pullulan) were positively correlated with the Bacteroides CAGs, whereas glycosidases predicted to target mostly disaccharides were positively correlated with the Firmicutes CAGs.

To further examine whether polysaccharides degradation functions could mediate the associations between host amylase and Bacteroidetes CAG species, we next assembled shotgun sequencing data from each mouse and binned contigs into metagenome-assembled genomes (MAGs). From 436 high-quality non-redundant MAGs (completeness > 90% and contamination < 5%), 31 MAGs were annotated as Muribaculaceae (Supplementary Table 5). Bacterial genomes from the Bacteroidetes phylum encode polysaccharide-utilization loci (PULs), the products of which regulate the detection, transport, and degradation of glycans. We further identified PULs in MAGs (Methods) and found that the Muribaculaceae MAGs that have higher number of PULs were more negatively associated with Amy1 expression levels in Ath-HMDP mouse (Supplementary Fig. 6C). Glycoside hydrolase family 13 (GH13) is the major glycoside hydrolase family acting on substrates containing α -glucoside linkages present in starch. We found that an abundance of Muribaculaceae MAGs that have PULs containing GH13 were negatively associated with Amy1 expression levels in Ath-HMDP mice (Supplementary Fig. 6D).

We reasoned that variation within the host amylase genes locus could affect amylase activity, which in turn would result in different availability of carbohydrates for microbes in the gut. This could in turn lead to distinct gut microbial communities (enterotypes). To further examine this hypothesis, we characterized the gut metagenomes from six commercially available genetically diverse mouse strains, C57BL/6 J (B6), A/J (A/J), 129S1/ SvImJ (129), NOD/ShiLtJ (NOD), NZO/HLtJ (NZO), and CAST/EiJ (CAST), which harbor diverse genetic background (Fig. 3A). These mice (4-6 animals per)strain) were fed a high carbohydrate diet (Supplementary Table 7) for 12 weeks. We first measured amylase gene copy numbers (CN) using digital PCR and found Amy1 gene copy number was detected at similar levels among all mouse strains (CN = 1-2); however *Amy2* gene copy number showed dramatic variation among the mouse strains (CN=9-16) (Fig. 3B). We further found the abundance of *Muribaculaceae* family was significantly higher in low amylase gene CN mice (e.g., CAST) comparing to high amylase gene CN mice (e.g., B6) (Fig. 3C). In addition, bacterial alpha-amylase abundance was also significantly higher in low amylase gene CN mice compared to high amylase gene CN mice (Fig. 3D). These results suggested that a lower amylase gene copy number in the mouse genome resulted in increased *Muribaculaceae* abundance in the gut supporting the hypothesis that genetic variation in the amylase gene region could be a causal driver for enterotype associated variants on chromosome 3 in mice.

Enterotype species are associated with host traits

Because of a large variation in host cardiometabolic phenotypes (Supplementary Fig. 7), we next explored the associations between CAG species with these host phenotypes. Species from the two CAGs discussed above showed distinct associations (Fig. 4A). Atherosclerotic lesion size was positively correlated with Bacteroides sp. (Spearman's $\rho = 0.35$, $P = 4.1 \times 10^{-6}$) and Bilophila sp. (Spearman's $\rho = 0.17$, $P = 4.4 \times 10^{-2}$) and negatively correlated with *Roseburia* sp. (Spearman's $\rho = -0.31$, $P = 5.6 \times 10^{-5}$), a finding that was also reported in a previous study [25]. We further identified positive associations for Turicimonas muris (Spearman's $\rho=0.18$, P=0.03), Atopobiaceae bacterium sp. (Spearman's $\rho = 0.2$, P = 0.015) and negative associations for Dorea sp. (Spearman's $\rho = -0.29$, $P = 2 \times 10^{-4}$) and Enterenecus sp. (Spearman's $\rho = -0.16$, P = 0.05) with atherosclerotic lesion size. Plasma levels of HDL and LDL/VLDL were positively correlated with bacteria from Firmicutes CAG and negatively correlated with bacteria from Bacteroidetes CAG. Liver fibrosis was negatively correlated with Bacteroidetes CAG species including Bacteroides sp. and taxa within the *Muribaculaceae* family and positively correlated with Firmicutes CAG species, such as Sporofaciens sp., Enterenecus sp., and Kineothrix sp. These results suggested that host physiology phenotypes are associated with gut microbiome enterotypes (Firmicutes and Bacteroidetes CAG) in the Ath-HMDP mice.

We next performed bi-directional Mendelian randomization (MR) to assess whether Firmicutes and Bacteroidetes enterotypes causally contribute to host traits (Supplementary Table 8). We observed significant causal relationships between microbial functions' PC1 and liver fibrosed area ($P=7.3 \times 10^{-3}$) and HDL cholesterol ($P=4.6 \times 10^{-4}$) (Fig. 4B, Supplementary Fig. 8). When we tested MR considering clinical traits as the exposure and microbial functions' PC1 as the outcome, we did not observe significant causal effects (Fig. 4C).

Enterotype functions are associated with host traits

We found that bacterial flagellar functions were overrepresented among the functions that mapped at chromosome 3 locus (Supplementary Fig. 9A), and they are noteworthy characteristic that distinguish Firmicutes and Bacteroidetes CAGs (Supplementary Fig. 9B). We found the abundance of *Lachnospiraceae*, a family



Fig. 3 *Muribaculaceae* and gut bacterial alpha-amylase genes are more abundant in mouse strains with low-amylase gene copy numbers. A Phylogenetic relationship among the six mouse strains tested, C57BL/6 J (B6), A/J (A/J), 129S1/SvImJ (129), NOD/ShiLtJ (NOD), NZO/HLtJ (NZO), and CAST/EiJ (CAST). B Amylase gene (*Amy1* and *Amy2*) copy number in the six mouse strains. C Relative abundance of *Muribaculaceae* family across the six mouse strains fed a high carbohydrate diet and its correlation with host amylase gene copy number. D Relative abundance of gut bacterial alpha-amylase gene (K07405) across the six mouse strains and its correlation with host amylase gene copy number. Statistical difference among groups was tested by unpaired two-tailed Wilcoxon signed-rank test

containing many flagellated bacteria, was significantly higher in higher amylase gene CN mice comparing to low amylase gene CN mice (Supplementary Fig. 10A). We observed a strong association between *fliC* abundance (K02406; encoding flagellar filament structural protein) and LDL+VLDL cholesterol (Spearman's ρ =0.31, *P*=5.8×10⁻¹⁰), HDL cholesterol (Spearman's ρ =0.17, *P*=1.8×10⁻³), liver fibrosis area (Spearman's ρ =0.17, *P*=2.5×10⁻³), liver cholesterol levels (Spearman's ρ =0.24, *P*=4.1×10⁻⁵), and liver triglycerides (Spearman's $\rho = -0.27$, $P = 4.3 \times 10^{-6}$) (Supplementary Fig. 11A). Sequence polymorphisms in flagellin gene can result in structural variants that influence host recognition and immune responses [26, 27]. We examined the most abundant (relative abundance > 0.01%) flagellin gene variants from the *Lachnospiraceae* (*Roseburia, Dorea*) and *Desulfovibrionaceae* (*Desulfovibrio*) families and found that variants in the nD1 TLR5 epitope motif in the *fliC* gene were associated differently with host physiology phenotypes (Supplementary Fig. 11B). These



Fig. 4 Host metabolic phenotypes are associated with gut microbial enterotypes. A Spearman's correlation between bacteria species from Firmicutes CAG and Bacteroidetes CAG with host metabolic phenotypes. B Bidirectional MR analysis between host metabolic traits and gut microbial functions' PC1, which is the representative microbial features that distinguish Firmicutes and Bacteroidetes enterotypes. MR beta effect sizes using gut microbial functions' PC1 as exposure and host metabolic traits as an outcome. C MR beta effect sizes using host metabolic traits as exposure and gut microbial functions' PC1 as outcome

results suggest that flagella may influence the progression of liver disease and, consistent with previous work, show that genes involved in flagellar assembly were enriched in fecal microbiomes of patients with moderate to severe fibrosis [28].

Discussion

Genome-wide association studies have identified multiple host genomic loci associated with the gut microbiome in humans and mice; however, most of these previous efforts focused on organismal composition, and

there is limited evidence linking specific functions and pathways with host genetic variation [16]. Additionally, gut microbiome enterotypes were described in human cohorts where a small number of bacterial taxa determines the stratification of the whole community, but there was limited evidence of enterotypes in genetically diverse mouse cohorts [29]. The work presented here comprehensively characterized gut microbiome composition, functions, and metabolic pathways in the Ath-HMDP mouse cohort. In this population, we identified three enterotypes dominated by different phyla including Firmicutes, Bacteroidetes, and Verrucomicrobia. We also found that the enterotypes were associated with bacterial taxa (Firmicutes and Bacteroidetes CAGs) and microbial functions (starch and sugar metabolism). The Bacteroidetes CAG included key drivers for these associations from Muribaculaceae, the most abundant family in mouse gut [23].

Using genetic mapping, we identified host genomic loci associated with bacterial taxa, functions, and pathways, and two of these loci were associated with enterotypes. The genetic variant rs31001780 (A/G) at the Chr3 locus was significantly associated with Firmicutes and Bacteroidetes enterotypes and genetic variant rs31965376 (A/T) at the Chr1 locus was significantly associated with the Verrucomicrobia enterotype. Importantly, the expression level of the Amy1 gene, which spans in the LD region of the Chr3 locus, was significantly correlated with Lachnospiraceae and Muribaculaceae abundances. Given the same starch content, genetic variants of the Amy1 gene may allow different carbohydrate accessibility for gut microbiota and shape the different enterotypes. In humans, individuals with higher salivary amylase gene (AMY1) copies show higher levels of many genera within the Firmicutes in the gut [15], which aligns with our results in mice (i.e., Amy1 gene expression is positively associated with Firmicutes abundance). We reasoned that the differences in sugar and starch metabolism between Firmicutes and Bacteroidetes CAG species may explain how genetic amylase locus variants shaped the enterotypes in the mouse gut. Previous work showed that mice treated with acarbose, which inhibits amylase activity, have an increased abundance of the Muribacu*laceae* family in gut [30–32]. Members of this family possess starch utilization genes [31]. Consistent with these findings we found Muribaculaceae were more abundant in low amylase gene copy number mice. This is a different amylase gene than the one that was identified to be associated with gut microbiome in humans (salivary amylase gene *AMY1*) [15, 33]. We found salivary (*Amy1*) amylase gene expression level and pancreatic (Amy2) amylase gene copy number were associated with gut microbiome. Both salivary and pancreatic amylases would impact carbohydrate availability and potentially influence the gut microbiome; thus, loss-of-function mouse experiments are needed to directly test the effects of each amylase on the gut microbiome.

MR analysis showed that the Firmicutes and Bacteroidetes enterotypes represented by the microbial functions' PC1 were associated with increased liver fibrosed area and HDL cholesterol levels. Previous studies showed that gut microbiome partially explained variations of plasma triglyceride and HDL cholesterol levels in humans [34]. Another study showed that a high-fat diet increased flagellated bacteria in the gut, which increased apolipoprotein A1 (ApoA1) production and HDL cholesterol levels in mice [35]. MR seeks to infer causal effects of modifiable exposure using measured variation in genes of known function. Successful applications of MR in humans have revealed relationships between gut microbiome and other molecular traits, including blood metabolites [36], short-chain fatty acids [37], and host metabolic traits [38, 39]. To the best of our knowledge, our study is the first MR application of gut microbiome in a genetically diverse mouse cohort. Furthermore, our MR results provide evidence for the potentially causal relationship between gut flagellated bacteria and plasma HDL cholesterol levels. However, the sample size in this study is smaller than that in large-scale human studies, which often include thousands of participants. Therefore, further research with larger sample sizes will be necessary to validate these findings. Additionally, follow-up studies will need to be conducted to test this causal relationship in mice directly.

We found genetic variants at the amylase gene locus were not only associated with *Muribaculaceae* but also associated with *Lachnospiraceae*, a family containing many flagellated bacteria. We further reasoned that the amylase gene copy number could perhaps indirectly affect the abundance of flagellated bacteria because of lacking starch degradation functions. Interestingly, previous work showed that treatment with acarbose attenuates experimental non-alcoholic fatty liver disease [40] and non-alcoholic steatohepatitis in mice [41]. The causal role of flagellated bacteria and the mechanism by which it might exacerbate liver fibrosis warrant further investigation.

A recent study showed that bacteria flagellin gene variants from *Lachnospiraceae* family were associated with differential TLR5 activation [26]. We also found that variants in the nD1 TLR5 epitope motif in the *fliC* gene were associated differently with host physiology phenotypes, including atherosclerotic lesions. This underscores the importance of bacterial genetic variation in gut microbiome association studies. Recent studies have linked bacterial SNPs in the human gut microbiome with host phenotypes [42], but investigations of bacterial SNPs in the mouse gut microbiome have been limited by a lack of host genetic diversity. We demonstrate here that using a genetically diverse cohort of mice, such as the HMDP, is a effective approach for uncovering gene-level microbiome associations with the host.

Together, this work highlights how host genetics can shape different microbial enterotypes in the mouse gut and identifies potential candidate host genes involved. The presented results also suggest that the microbial enterotype-associated functions and pathways that are the consequence of host genetic variantion may influence host cardio-metabolic phenotypes.

Methods

HMDP mouse cohort

Male and female mice from the F1 Ath-HMDP panel were maintained in a temperature and humidity-controlled environment under a 12-h light/dark cycle (lights on at 6:00 and off at 18:00). Mice were housed by strain. All mice were fed a high-fat diet (33 kcal % fat from cocoa butter) supplemented with 1% cholesterol (Research Diets D10042101) for 16 weeks. Cecal contents were collected immediately after the animals were euthanized. Among the 918 animals generated for the original study, 356 mice had cecal samples available for metagenomic sequencing. Clinical phenotypes for this cohort were described in a previous study [20].

Validation mouse cohort

Male A/J, C57BL/6 J, 129S1/SvImJ, NOD/ShiLtJ, NZO/ HLtJ, and CAST/EiJ mice were maintained in a temperature and humidity controlled environment under a 12-h light/dark cycle (lights on at 6:00 and off at 18:00). Mice were single housed. All mice were fed a high-carbohydrate and low-fat diet (Envigo Teklad TD.200339) for 12 weeks. Cecal contents were collected immediately after the animals were euthanized.

Metagenomic sequencing

Cecal DNA was extracted from individual mice using the PowerSoil DNA Isolation Kit. DNA concentration was verified using the Qubit[®] dsDNA HS Assay Kit (Life Technologies, Grand Island, NY). Samples were prepared using the Illumina NexteraXT library preparation kit. The quality and quantity of the finished libraries were assessed using an Agilent bioanalyzer and Qubit[®] dsDNA HS Assay Kit, respectively. Libraries were standardized to 2 nM. Paired end, 150 bp sequencing was performed using the Illumina NovaSeq6000. Images were analyzed using the standard Illumina Pipeline, version 1.8.2. Low-quality reads were filtered out from raw DNA reads using Trimmomatic [43] (v.0.39) with parameters (SLIDINGWINDOW:4:20 MINLEN:50). To identify and eliminate host sequences, reads were aligned against the mouse genome (mm10/GRCm38) using Bowtie2 [44] (v.2.3.4) with default settings and microbial DNA reads that did not align to the mouse genome were identified using samtools (v.1.3; samtools view -b -f 4 -f 8). Samples with a total read depth of < 10 million were excluded from downstream analyses.

Profiling microbiome composition

Gut microbial taxon was profiled by MetaPhlAn4 pipeline (v.4.0.2) using the MetaPhlAn database (mpa_ vOct22) and the ChocoPhlAn pan-genome database (mpa_vOct22_CHOCOPhlAnSGB_202212) that contains a collection of around 1 million prokaryotic metagenome-assembled genomes [45]. The taxonomy clades with average relative abundance > 0.01% and present in at least > 20% of samples were kept as microbial taxons for downstream analyses. The unclassified SBG taxa were further annotated to the Genome Taxonomy Database (GTDB) using mpa_vOct22_CHOC-OPhlAnSGB_202212_SGB2GTDB.tsv data from the MetaPhlan4 pipeline.

Profiling microbiome function and pathway

Quantification of microbial genes was done by aligning clean paired-end reads to a previously published mouse gut microbiome non-redundant gene catalog [19] using Bowtie2 [44] (v.2.3.4) and default parameters. RSEM [46] (v.1.3.1) was used to estimate microbial gene abundance. The relative abundance of microbial gene counts per million (CPM) was calculated using microbial gene expected counts divided by gene effective length and then normalized by the total sum. To obtain abundance information for microbial functions, the CPM of genes with the same KEGG orthologous (KO) annotation was summed together. In case there were multiple KO annotations for a single gene, we used all KO annotations. Gut microbial pathways were profiled using the HUMAnN3 pipeline (v.3.0.0), the MetaPhlAn database (mpa_v20_m200), the ChocoPhlAn pan-genome database (v296_v201901b), and the UniRef90 protein database [47] (v.0.1.1). Pathways with average relative abundance > 0.01% detected in at least>20% samples were used for downstream analyses.

Enterotype clustering and microbial co-abundance groups Gut microbiome data was clustered using partitioning around medoid (PAM) clustering via pam() function from the R package cluster (v.2.1.2). The Bray–Curtis dissimilarity matrix of 166 bacteria species abundance was used for PAM clustering. The optimal cluster number was estimated using the Calinski–Harabasz index via pamk() function from R package fpc (v.2.2–13). In addition, we tested different parameters *k* using the pam() function, and the optimal cluster number k=3 gave the highest average silhouette width value. The dominant phylum in each cluster was determined by the highest abundant taxon compared to other clusters. Similarity scores between species were calculated using CCREPE (compositionality corrected by renormalization and permutation) package [48] (v.1.1.3). The species network was visualized using ggnet2() function from R package ggnet (v.0.1.0).

Metagenome-assembled genome

Shotgun sequencing reads were assembled for all 356 samples using SPAdes [49] (v.3.15.5; metaspades.py -k 21,33,55,77), the assembled contigs were quantified by mapping shotgun reads to contigs using Bowtie2 (v.2.3.4). Contigs that were less than 500 bp were excluded for downstream analyses. Contigs were then binned into MAGs using MetaBAT2 [50] (v.2.17). The quality of all MAGs was assessed by genome completeness and contamination using CheckM [51] (v.1.2.2), and high-quality MAGs were kept (completeness>90% and contamination < 5%). The high-quality MAGs were dereplicated using dRep [52] (v.3.5.0; -pa 0.9 -sa 0.99). The final dataset contains 436 MAGs. To quantify the abundance of each MAG in each sample, shotgun reads from each cecal sample were mapped to MAGs using Bowtie2 [44] (v.2.3.4) and RSEM [46] (v.1.3.1). Taxonomic assignments of these 436 MAGs were using the Genome Taxonomy Database Toolkit (GTDB-Tk [53]; v.2.3.2) and the GTDB database (ver. R214). Genes from MAGs were predicted using Prodigal [54] (v.2.6.3) and annotated to the KEGG Orthology database using kofamscan [55] (v.1.3.0) and annotated to carbohydrate-active enzymes (CAZymes) database using dbcan3 [56] (v.4.1.4) and database "dbCAN3_db_v12_20240415". PULs were identified followed by previous study [57], by combining marker genes (SusC-SusD), CAZyme annotation, and operon structure.

Genome-wide association of gut microbiome

The mouse diversity genotyping array was used for genotyping and gave approximately 450,000 SNPs. SNPs used for each trait were filtered by the following criteria: the minor allele frequency (MAF) > 5% and missing genotype frequency < 10%. GWAS analysis was performed using FaST-LMM [58] (Python v.3.7.4). When testing SNPs on chromosome N, all SNPs from other chromosomes besides N were used for kinship matrix construction, that is the leave out one chromosome (LOOC) approach. The kinship matrix was used in the regression model (the argument "K0" from fastImm algorithm). Any bias from measuring different numbers of mice for a given trait was accounted for by the kinship matrix. Sex was used as a covariate in the regression model. All individual mice were included in the GWAS regression model. GWAS significance thresholds were determined by permutation tests [20]. A genome-wide significance threshold of $P < 4 \times 10^{-6}$ was used. We defined a study-wide significance threshold of $P < 4 \times 10^{-6}$ was used. We defined a study-wide significance threshold of $P < 4 \times 10^{-6}/(2127 + 108 + 300) = 1.5$ 8×10^{-9} . Broad sense heritability for each trait was estimated using "repeatability()" function from "heritability" (v.1.3) R package. Narrow sense heritability for each trait was calculated using all filtered SNPs to estimate the proportion to explain total variations for each trait.

MR analysis

Bidirectional MR analyses were performed to first test if microbiome traits causally affect a host phenotype and then test if the host phenotypes can causally affect the microbiome traits. We identified independent genetic variants with $P < 1 \times 10^{-4}$ as instrument variables in MR. We carried out an inverse variance weighted (IVW) test using R package TwoSampleMR [59] (v.0.5.6). Egger regressions were performed for analyses with more than two instrument variables available to test sensitivity and horizontal pleiotropy.

Amylase gene copy number analysis

Digital PCR (dPCR) was used to experimentally determine *Amy1* and *Amy2* gene copy numbers. dPCR was run on the Applied Biosystems QuantStudio Absolute Q digital PCR system using Absolute Q DNA Digital PCR Master Mix. The *Tfrc* gene was used as a control (TaqMan Copy Number Reference Assay, mouse). Primers and probes targeting exon 5 of the *Amy1* gene (Applied Biosystems TaqMan Copy number assay, assay ID Mm00085061_cn) were used to determine the *Amy1* gene copy number. Primers and probes for the *Amy2* gene were described in a previous study [14]. One microliter of DNA was added per reaction. Data was collected and analyzed by the QuantStudio Absolute Q digital PCR system software.

Data and statistical analysis

All data integration and statistical analysis were performed in R (v.3.6.3). Differences between groups were evaluated using an unpaired two-tailed Wilcoxon signedrank test. Enrichment analysis was performed with Fisher's exact test using a custom R function. Correlation analysis was performed with two-sided Spearman's correlation using the R function 'cor.test()'. For multiple testing, the Benjamini–Hochberg false discovery rate (FDR) procedure was used to adjust *P* values. Data integration was performed using R packages dplyr (v.1.0.6), tidyr (v.1.1.3), reshape2 (v.1.4.4), and data.table (v.1.14.0). Heatmaps were generated using the R package pheatmap (v.1.0.12). Other plots were created using the R packages ggplot2 (v.3.3.3).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-025-02093-y.

Additional file 1. Supplementary Fig. 1 Gut microbiota composition and top bacteria species in Ath-HMDP mouse cohort. A. The relative abundance of microbiota composition at phyla level in Ath-HMDP mouse cohort consuming a high fat diet (33 kcal % fat from cocoa butter) supplemented with 1% cholesterol (Research Diets D10042101) for 16 weeks. B. The relative abundance of top 50 abundant bacteria species and their prevalence in Ath-HMDP mouse cohort. Supplementary Fig. 2 Gut microbial species and functions and their correlations with three enterotype phyla. A. PCoA showing community-level difference among mice in the cohort. Three major phyla show large inter-individual variation and stratify the mouse samples. B. PCA visualizing the microbial functions in the cohort. Three major phyla show large inter-individual variation and stratify the mouse samples. C. Spearman's correlation between microbial functions' PC1 with abundance of Firmicutes, Verrucomicrobia and Bacteroidetes. D. Spearman's correlation between microbial functions' PC2 with abundance of Firmicutes, Verrucomicrobia and Bacteroidetes. Supplementary Fig. 3 Genetic associations of gut microbial traits in Ath-HMDP mice. Genetic associations of gut microbial functions, metabolic pathways and species abundance. The density of associated gut bacterial functions in the mouse genome is showed on the top. Dashed lines represent significance thresholds determined by permutation tests (P < 4 x 10-6). Supplementary Fig. 4 Amy1 gene expression level correlates with microbial enterotypes and associated species. A. Spearman's correlation between candidate genes (Amy1, Rnpc3, Col11a1 and Olfm3) in Chr3 locus with microbial functions' PC1 and PC2. B. Spearman's correlation between candidate genes (Amy1, Rnpc3, Col11a1 and Olfm3) in chromosome 3 locus and abundances of Muribaculaceae and Lachnospiraceae families. Supplementary Fig. 5 Host genomic loci are associated with gut A. muciniphila abundance. A. Genomic locus at Chr1 193–194 Mbp are associated with Akkermansia muciniphila abundance. The lead SNP is rs31965376. Dashed lines represent significance thresholds determined by permutation tests (P < $4 \times 10-6$). B. Mice with allele AA or TT at SNP rs31965376 visualized in PCoA of species beta-diversity (left) and functions beta-diversity (right). C. Relative abundance of Firmicutes, Bacteroidetes, and Verrucomicrobia from mice with AA or TT at SNP rs31965376. D. Number of mice with AA or TT at SNP rs31965376 in each of three enterotypes. Statistical difference between groups was tested by unpaired two-tailed Wilcoxon signed-rank test. Supplementary Fig. 6 Bacterial starch and sugar metabolism is associated with microbial enterotypes' species. A. Top bacterial pathways significantly and oppositely associated with Firmicutes CAG and Bacteroidetes CAG (adjusted P < 0.05). B. Spearman's correlation between bacterial species and bacterial functions involved in starch and sugar metabolism. Bacterial species from Firmicutes CAG and Bacteroidetes CAG showed opposite associations. C. Spearman's correlation between the number of PUL in 31 Muribaculaceae MAG and the correlation coefficient of corresponding MAG's abundance with Amy1 gene expression levels: The value in x axis represents how strong, positively or negatively, a MAG is associated with Amy1 gene expression, the value in y axis represents how many PULs in that MAG. D. Spearman's correlation between the number of PUL that contain GH13 gene in 31 Muribaculaceae MAG and the correlation coefficient of corresponding MAG's abundance with Amy1 gene expression levels. Supplementary Fig. 7 Summary of host metabolic phenotypes in Ath-HMDP. Boxplots show the distribution of data for each host metabolic phenotype in the Ath-HMDP population. The coefficient of variation (CV) for each phenotype was calculated. Supplementary Fig. 8 Scatter plots of SNP effect on exposure and outcome. A. Microbial functions' PC1 as exposure and plasma HDL cholesterol as outcome. B. Microbial functions' PC1 as exposure and liver fibrosed area as outcome. Supplementary Fig. 9 Bacterial flagellar assembly genes are associated with microbial enterotype species. A. Enrichment analysis

using Fisher's exact test for gut bacterial functions mapping at the hotspot locus on Chr3: 112–116 Mbp. B. Spearman's correlation between bacterial species and bacterial functions involved in flagellar assembly. Bacterial species from Firmicutes CAG and Bacteroidetes CAG showed opposite associations. Supplementary Fig. 10 Lachnospiraceae family is associated host amylase gene copy number and host metabolic phenotypes. A. Relative abundance of the Lachnospiraceae family across six mouse strains fed a high carbohydrate diet and its correlation with host amylase gene copy number. B. Bidirectional MR analysis between host metabolic traits and gut bacterial flagellin gene abundance. MR beta effect sizes using gut bacterial flagellin gene abundance as exposure and host metabolic traits as outcome. C. MR beta effect sizes using host metabolic traits as exposure and gut bacterial flagellin gene abundance as outcome. Supplementary Fig. 11 Gut bacterial flagellin gene association with host phenotypes. A. Spearman's correlation between host metabolic traits and gut bacterial flagellin gene abundance. B. Spearman's correlation between host metabolic traits and most abundant individual gut bacterial flagellin gene abundance from Lachnospiraceae and Desulfovibrionaceae family. The flagellin gene sequences were aligned and conserved motifs were identified

Additional file 2. Supplementary Table 1. Summary of host physiologic traits and gut metagenome in Ath-HMDP mice. Host cardio-metabolic phenotypes in Ath-HMDP mice were obtained from previous studies. Gut metagenomes were characterized from 356 Ath-HMDP mice, encompassing 90 strains (190 male mice and 166 female mice). Supplementary Table 2. Gut microbiome traits in the Ath-HMDP mice using MetaPhlAn4. Supplementary Table 3. Spearman's correlation between the microbial functions' PC1 and bacteria species. Supplementary Table 4. Spearman's correlation between the microbial functions' PC2 and bacteria species. Supplementary Table 5. Ath-HMDP MAG summary and annotation. Supplementary Table 6. SNP information for top enterotype associated SNPs. Supplementary Table 7. High-carbohydrate diet used in amylase validation experiment. Supplementary Table 8. Instrument variables used in MR analysis and Egger regressions test for sensitivity and horizontal pleiotropy.

Acknowledgements

We thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing sequencing and support services; the University of Wisconsin Biotechnology Center BioGarage for providing dPCR services; the University of Wisconsin Center for High Throughput Computing (CHTC) in the Department of Computer Sciences for providing computational resources, support and assistance.

Authors' contributions

F.E.R. and A.J.L. conceived the study. Q.Z. designed the experiments and data analyses. Q.Z. and E.R.H. contributed to sample processing for metagenomic sequencing. Q.Z. performed metagenomic, network and GWAS analyses. C.P. assisted with GWAS analyses. M.F.W., M.P.K. and A.D.A. assisted with high carbohydrate mouse experiment. Q.Z. and F.E.R. wrote the manuscript. All authors approved the final manuscript.

Funding

This work was supported by the National Institutes of Health (NIH) grants HL144651 (F.E.R., A.J. L), HL148577 (F.E.R., A.J. L), and RC2DK125961 (A.D.A.). E.R.H. and M.F.W. were supported in part by the Metabolism and Nutrition Training Program NIH T32 (DK007665). E.R.H. was supported by the University of Wisconsin–Madison Food Research Institute (Robert H. and Carol L. Deibel Distinguished Graduate Fellowship in Probiotic Research). This work was also supported by a grant from a Transatlantic Networks of Excellence Award from the Leducq Foundation (17CVD01).

Data availability

Shotgun metagenomic data are available from the Sequence Read Archive (SRA) under accession number PRJNA1078419. Microarray data for live expression are deposited in the NCBI GEO under the accession number GSE66570. Mouse genotype data is available at Dryad https://doi.org/10.5061/dryad. cz8w9gjb9. Code and metadata used in this study is available in GitHub repository https://github.com/gijunz/Zhang_HMDP_paper.

Declarations

Ethics approval and consent to participate

Animal care and study protocols were approved by the AAALAC-accredited Institutional Animal Care and Use Committee of the College of Agricultural Life Sciences at the University of Wisconsin-Madison (UW-Madison). All experiments with mice were performed under protocols approved by the UW-Madison Animal Care and Use Committee.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 March 2024 Accepted: 16 March 2025 Published online: 21 April 2025

References

- 1. Lynch SV, Pedersen O. The human intestinal microbiome in health and disease. N Engl J Med. 2016;375:2369–79.
- Wang Z, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature. 2011;472:57–63.
- Tremaroli V, Bäckhed F. Functional interactions between the gut microbiota and host metabolism. Nature. 2012;489:242–9.
- Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell. 2014;157:121–41.
- Kurilshikov A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 2021;53:156–65.
- Gacesa R, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature. 2022;604:732–9.
- MetaHIT Consortium (additional members), et al. Enterotypes of the human gut microbiome. Nature. 2011;473:174–180.
- Costea PI, et al. Enterotypes in the landscape of gut microbial community composition. Nat Microbiol. 2017;3:8–16.
- 9. Murga-Garrido SM, et al. Gut microbiome variation modulates the effects of dietary fiber on host metabolism. Microbiome. 2021;9:117.
- Hutchison ER, et al. Dissecting the impact of dietary fiber type on atherosclerosis in mice colonized with different gut microbial communities. Npj Biofilms Microbiomes. 2023;9:31.
- 11. Perry GH, et al. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 2007;39:1256–60.
- 12. Bolognini D, et al. Recurrent evolution and selection shape structural diversity at the amylase locus. Nature. 2024;634:617–25.
- Linnenbrink M, Ullrich KK, McConnell E, Tautz D. The amylase gene cluster in house mice (Mus musculus) was subject to repeated introgression including the rescue of a pseudogene. BMC Evol Biol. 2020;20:56.
- 14. Pajic P, et al. Independent amylase gene copy number bursts correlate with dietary preferences in mammals. eLife. 2019;8:e44628.
- Poole AC, et al. Human salivary amylase gene copy number impacts oral and gut microbiomes. Cell Host Microbe. 2019;25:553-564.e7.
- 16. Sanna S, et al. Challenges and future directions for studying effects of host genetics on the gut microbiome. Nat Genet. 2022;54:100–6.
- 17. Kemis JH, et al. Genetic determinants of gut microbiota composition and bile acid profiles in mice. PLOS Genet. 2019;15: e1008073.
- Org E, et al. Genetic and environmental control of host-gut microbiota interactions. Genome Res. 2015;25:1558–69.
- 19. Zhang Q, et al. Genetic mapping of microbial and host traits reveals production of immunomodulatory lipids by Akkermansia muciniphila in the murine gut. Nat Microbiol. 2023;8:424–40.
- Bennett BJ, et al. Genetic architecture of atherosclerosis in mice: a systems genetics analysis of common inbred strains. PLOS Genet. 2015;11: e1005711.
- 21. Hui ST, et al. The genetic architecture of diet-induced hepatic fibrosis in mice. Hepatology. 2018;68:2182–96.

- 22. Kasahara K, et al. Gut bacterial metabolism contributes to host global purine homeostasis. Cell Host Microbe. 2023;31:1038-1053.e10.
- Lkhagva E, et al. The regional diversity of gut microbiome along the GI tract of male C57BL/6 mice. BMC Microbiol. 2021;21:44.
- Usher CL, et al. Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. Nat Genet. 2015;47:921–5.
- Kasahara K, et al. Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model. Nat Microbiol. 2018;3:1461–71.
- Clasen SJ, Bell MEW, Borbón A, Lee DH, Henseler ZM. Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. Sci Immunol. 2023;8(79):7001.
- 27. Bourgonje AR, et al. Phage-display immunoprecipitation sequencing of the antibody epitope repertoire in inflammatory bowel disease reveals distinct antibody signatures. Immunity. 2023;56:1393-1409.e6.
- Schwimmer JB, et al. Microbiome signatures associated with steatohepatitis and moderate to severe fibrosis in children with nonalcoholic fatty liver disease. Gastroenterology. 2019;157:1109–22.
- 29. Hildebrand F, et al. Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. Genome Biol. 2013;14: R4.
- Smith BJ, et al. Changes in the gut microbiome and fermentation products concurrent with enhanced longevity in acarbose-treated mice. BMC Microbiol. 2019;19:130.
- 31. Smith BJ, Miller RA, Schmidt TM. Muribaculaceae genomes assembled from metagenomes suggest genetic drivers of differential response to acarbose treatment in mice. mSphere. 2021;6:e00851-21.
- 32. Baxter NT, Lesniak NA, Sinani H, Schloss PD, Koropatkin NM. The glucoamylase inhibitor acarbose has a diet-dependent and reversible effect on the murine gut microbiome. mSphere. 2019;4:e00528-18.
- Hasegawa T, et al. Impact of salivary and pancreatic amylase gene copy numbers on diabetes, obesity, and functional profiles of microbiome in Northern Japanese population. Sci Rep. 2022;12:7628.
- 34. Fu J, et al. The Gut microbiome contributes to a substantial proportion of the variation in blood lipids. Circ Res. 2015;117:817–24.
- Yiu JHC, et al. Gut microbiota-associated activation of TLR5 induces apolipoprotein A1 production in the liver. Circ Res. 2020;127:1236–52.
- Liu X, et al. Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. Nat Genet. 2020;54:52–61.
- Sanna S, et al. Causal relationships among the gut microbiome, shortchain fatty acids and metabolic diseases. Nat Genet. 2019;51:600–5.
- Qin Y, et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. Nat Genet. 2020;54:134–42.
- Rühlemann MC, et al. Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. Nat Genet. 2021;53:147–55.
- 40. Nozaki Y, et al. Long-term combination therapy of ezetimibe and acarbose for non-alcoholic fatty liver disease. J Hepatol. 2009;51:548–56.
- Lieber CS, et al. Acarbose attenuates experimental non-alcoholic steatohepatitis. Biochem Biophys Res Commun. 2004;315:699–703.
- 42. Zahavi L, et al. Bacterial SNPs in the human gut microbiome associate with host BMI. Nat Med. 2023;29:2785–92.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
- 44. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
- Blanco-Míguez A, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhIAn 4. Nat Biotechnol. 2023;41:1633–44.
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:1–16.
- Beghini F, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife. 2021;10:e65088.
- Gevers D, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15:382–92.
- 49. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34.

- Kang DD, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7: e7359.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
- Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11:2864–8.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics. 2020;36:1925–7.
- 54. Hyatt D, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11: 119.
- Aramaki T, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2.
- 56. Zheng J, et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023;51:W115–21.
- Terrapon N, Lombard V, Gilbert HJ, Henrissat B. Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. Bioinformatics. 2015;31:647–55.
- Lippert C, et al. FaST linear mixed models for genome-wide association studies. Nat Methods. 2011;8:833–5.
- 59. Hemani G, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.