Genome-centric investigation of the potential succession pattern in gut microbiota and altered functions under high-protein diet

Yiming Zhao [a,b,1], Lulu Chen [a,c,1], Siqi Yao [b], Liyu Chen [b], Jing Huang [d], Shuijiao Chen [a,c,**], Zheng Yu [b,*]

[a] Department of Gastroenterology, Xiangya Hospital Central South University, Changsha, Hunan, China
[b] Department of Microbiology, School of Basic Medical Science, Central South University, Changsha, Hunan, China
[c] National Clinical Research Center for Geriatric Disorders, Xiangya Hospital Central South University, Changsha, Hunan, China
[d] Department of Parasitology, School of Basic Medical Science, Central South University, Changsha, Hunan, China

ARTICLE INFO

ABSTRACT

Excessive intake of protein has been considered as a factor leading to intestinal microecological disorder, but why and how intestinal microbes change under the high-protein diet (HPD) have yet to be fully elucidated. Here, we performed 16S rRNA gene amplicon sequencing and metagenomic sequencing on contents of cecum, colon and feces from two groups of mice with standard diet (SD) and HPD. And then the microbial alteration of composition and function were deeply analyzed by using several statistical models and bioinformatic methods. Among the three niches, the microbes in the colon are observed to show the most significant change with lower alpha-diversity and higher beta-diversity after HPD. In addition, this alteration of microbial structure may be related to the replacement process and co-occurring community. Most species are also enriched or impoverished in the colon during this process. After analyzing the functional genes related to protein and carbohydrate hydrolysis in different niches, we found that the carbon source provided by poor carbohydrates compared with the rich protein may be the potential factor driving the enrichment of mucin degraders and desulphaters in the colon under HPD. Therefore, our study provided a new insight to understand the underlying mechanism of HPD affecting intestinal health from the perspective of microbial functional ecology.

1. Introduction

The human gut harbors a diverse microbial communities, collectively known as the gut microbiota, which has co-evolved with the host over millions of years (Suzuki et al., 2022). Gut microbiota is involved in numerous physiological processes, including nutrient metabolism, energy extraction, immune system modulation, and protection against pathogenic invaders (Qin et al., 2010; Lozupone et al., 2012). It is increasingly recognized that alterations in the composition and functional potential of the gut microbiota can impact host health and contribute to the development of various diseases, including obesity, metabolic syndrome, and inflammatory disorders (Turnbaugh et al., 2006; Tilg et al., 2016).

Among the factors influencing gut microbiota composition and activity, diet has emerged as a critical determinant. Dietary patterns can shape the gut microbiota by selectively promoting the growth of specific microbial taxa and modulating their metabolic capabilities (David et al., 2014; Sonnenburg and Bäckhed, 2016), which in turn affects host health (Singh et al., 2017; Zmora et al., 2019). In this context, the high-protein diet (HPD) has gained attention due to its potential effects on weight management, muscle synthesis, and metabolic health. Proteins are mainly digested in the small intestine, and the digestibility of them can reach 95% under normal diet, but under HPD, much more incomplete digested proteins and peptides are transferred through the ileocecal junction to the large intestine which can be utilized and metabolized by the local microbiota (Blachier et al., 2019; Abdallah et al., 2020). However, despite the widespread adoption of HPD, the impact of such a kind of diet on the gut microbiota and their associated functional potentials remains poorly characterized.

The different positions within the digestive tract form unique

---

ecological niches that shape the microbial communities residing there. For instance, as food passes through the stomach, the low pH and presence of gastric acid create a hostile environment for most microbes, resulting in relatively lower microbial abundance and diversity compared to other gut regions (Lozupone et al., 2012; Koga, 2022). Upon reaching the small intestine, where nutrient absorption predominantly occurs, the gut environment changes in pH, bile acids, and nutrient availability. These factors influence the composition of the gut microbiota, with the small intestine typically harboring lower microbial diversity and distinct taxonomic profiles compared to the colon (Rajilić-Stojanović et al., 2007). The colon is characterized by a relatively stable pH, low oxygen levels, and abundant undigested carbohydrates, providing an anaerobic environment conducive to the proliferation of bacteria belonging to the Firmicutes and Bacteroidetes phyla (Qin et al., 2010).

The protein entering the large intestine will affect the metabolism and community structure of the microbiota, and the microbiota can also produce many metabolites by utilization of nitrogen, which in turn affect the host's health (Abdallah et al., 2020). It has been also found that fecal and colonic microbiota have completely different abilities to hydrolyze protein (Zhao et al., 2019). The degradation of protein in the large intestine is more active in the distal part than in the proximal part (Windey et al., 2012). Furthermore, excessive protein intake is always accompanied by low carbohydrate intake, and both carbon and nitrogen sources are important nutrients for the competitive growth of intestinal flora. In this study, the changes of large intestinal flora composition and potential function in different positions (cecum, colon and feces) under HPD were explored. By starting from the function of genomes, our research provided critical insights into the complex interplay between HPD and gut microbiota in composition, functional and their regional distribution within the gut. Moreover, this research may pave the way for the development of the microbiota-host variant dynamics under different diet strategy.

2. Methods

2.1. Animals and sample selection

3-week female BALB/c mice from Hunan SJA Laboratory Animal Co. Ltd (Changsha, China) were raised under specific pathogen-free condition-controlled environment (22 ± 2 °C, 40%–50% humidity, normal light-dark cycle) with free access to food and drink. Mice were randomly fed with SD (n = 16), and the other with HPD (n = 18) for 4 weeks. The SD was the AIN-93G diet which complied with the American Institute of Nutrition's food-grade ingredient formulation standard. The HPD were modified according to AIN-93G criteria. The details of diet composition are shown in Table 1. All food was produced by Beijing HFK Bioscience Co. Ltd (Beijing, China). At the end of 4 weeks, the mice were stimulated to defecate, and the fecal specimens were collected using a sterile cryopreservation tube. After that, all mice were sacrificed through the inhalation of isoflurane, the whole large intestine was cut longitudinally

Table 1
Composition of the experimental diets.

	SD	HPD
Casein [g/kg]	200	593
Cystine [g/kg]	3	3
Starch [g/kg]	397	67
Maltodextrin [g/kg]	132	69
Sucrose [g/kg]	100	100
Fiber [g/kg]	50	50
Fat [g/kg]	70	70
Antioxidants [g/kg]	0.014	0.014
Minerals [g/kg]	35	35
Vitamins [g/kg]	10	10
Choline bitartrate [g/kg]	2.5	2.5

by sterilized scissors and the content of cecum and colon was wiped with sterile cotton swabs separately and stored at −80 °C. The animal experiment was approved by the Institutional Ethics Committee for Animal Procedures of the Central South University (No. 2018syclwo0252).

2.2. DNA extraction and 16S rRNA gene sequencing

All samples were accessed for 16S rRNA gene sequencing. In detail, Total genomic DNA in sample was extracted using the OMEGA Soil DNA Kit (M5635-01) (Omega Bio-Tek, Norcross, GA, USA), following the manufacturer's instructions. The PCR amplification of the bacterial 16S rRNA genes V3–V4 region was performed using the forward primer 338F (5′-ACTCCTACGGGAGGCAGCA-3′) and the reverse primer 806R (5′-GGACTACHVGGGTWTCTAAT-3′). Sample-specific 7 bp barcodes were incorporated into the primers for multiplex sequencing. The PCR components contained 5 μl of buffer (5 ×), 0.25 μl of Fast pfu DNA Polymerase (5 U/μl), 2 μl (2.5 mM) of dNTPs, 1 μl (10 μM) of each Forward and Reverse primer, 1 μl of DNA Template, and 14.75 μl of ddH$_2$O. Thermal cycling consisted of initial denaturation at 98 °C for 5 min, followed by 25 cycles consisting of denaturation at 98 °C for 30 s, annealing at 53 °C for 30 s, and extension at 72 °C for 45 s, with a final extension of 5 min at 72 °C. PCR amplicons were purified with Vazyme VAHTSTM DNA Clean Beads (Vazyme, Nanjing, China) and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA). After the individual quantification step, amplicons were pooled in equal amounts, and pair-end 2 × 250 bp sequencing was performed using the Illlumina NovaSeq platform with NovaSeq 6000 SP Reagent Kit (500 cycles).

2.3. 16S rRNA gene sequencing data processing

Briefly, to ensure the accuracy of the data, a series of processing steps on the raw sequence data were performed. The primers of sequence data were trimmed with cutadapt (Martin, 2011) (version 3.4). After that the sequences were merged and truncated to 400 bp using Vsearch (Rognes et al., 2016) (version 2.7.0). We performed quality control in Qiime2 (Estaki et al., 2020) (version 2021.4.0) with 'quality-filter' function using default parameters. Next, the reads were filtered chimera, denoised, clustered to Operational Taxonomic Units (OTUs), and profiled with 'deblur' function in Qiime2. To reduce false positives of obtaining data, the OTUs of which mapped reads less than 5 were discarded and rarefied to the minimum read counts using the R package 'picante' (Kembel et al., 2010). Meanwhile, the representative sequence was selected from each OTU to perform taxonomic classification by feature-classifier function in Qiime2 against the Silva database (Quast et al., 2013).

2.4. Microbial community composition and assembly analysis

To investigate the variation of microbial communities across samples, the Shannon–Wiener index and Abundance-based Coverage Estimator (ACE) index were calculated. At the same time, the Bray-Curtis distance metrics were calculated based on relative abundance of OTUs and visualized via nonmetric multidimensional scaling (NMDS) by the R package 'vegan' (Dixon, 2003). In addition, the Random Forest model was performed to determine affected sizes from different conditions (diets, positions and individuals) on microbial composition using the R package 'randomForest'. To test for differences among groups based on multivariate data, the permutational multivariate analysis of variance (PERMANOVA) was also applied. To estimate the dissimilarity among different groups, the jaccard distance metrics were calculated and partitioned into similarity, replacement and richness difference components using the R package 'adespatial'.

The null model approach described by Stegen et al. (2013) was applied to quantify community assembly processes in microbial

communities across niches. Using a null distribution of β-mean nearest taxon distance values where taxa are randomly shuffled across the phylogeny with 1,000 permutations, the beta-nearest taxon index (βNTI) was determined. The community assembly process was divided into five types (Hete-selection, Hom-selection, Drift, Dispersal limitation and Hom-dispersal) according to βNTI and Bray-Curtis-based Raup-Crick metrics (RCBray) indices using the 'qpen' function in the R package 'iCAMP' (Ning et al., 2020).

2.5. Universality of microbial interaction and co-occurrence network analysis

The dissimilarity-overlap curve (DOC) can be used to check whether different groups of samples have the same community dynamics. A higher fraction of data points for which the DOC displays a negative slope (*fns*) represents the stronger the microbial interaction influence (Bashan et al., 2016). The DOC analysis was performed with the R package 'DOC'.

To explore the microbial correlations between two diet groups, the SparCC (Friedman and Alm, 2012) co-occurrence networks were constructed using FastSpar (Watts et al., 2019) (Version 1.0) with OTUs presented in greater than or equal to 30% of the samples in each group based on different thresholds (FDR adjusted p-value and r-value), and the network complexity was also calculated according to methods in Wagg et al. (2019). To better understand co-occurrence microbes, a co-occurrence network was re-built utilizing all OTUs from two diet groups. To reduce false positives in detecting microbial co-occurrence relationships, only statistically significant (FDR, adjusted $p < 0.05$) correlations were incorporated into network analysis, and the random matrix theory (RMT) was used to select the appropriate r-value. Network visualization, modularity and parameters analysis were performed with the igraph (version 0.9.10) package in python. The greedy algorithm was applied to cluster all nodes in the RMT based co-occurrence network into four main modules (M0, M1, M2 and M3). In addition, we performed Linear discriminant analysis Effect Size (Lefse) analysis (Segata et al., 2011) to determine biomarkers in different groups.

2.6. Metagenomic sequencing and processing

To study the alteration of microbial functions at gene level, some samples in each group were selected randomly to perform metagenomic sequencing. Mice were divided into two groups according to different diets (SD and HPD). The samples of mice fed with SD were sequenced based on samples from different gut parts, in which the number of samples w as follows: cecum: 3, colon: 3, and feces: 4. And for mice fed with HPD, the selected number of samples for metagenomic sequencing was all three. There were total 19 samples selected to perform metagenomic sequencing. Total microbial genomic DNA was extracted using the QIAGEN DNeasy PowerWater Kit (14900-100-NF). The extracted microbial DNA was processed to construct metagenomic sequencing libraries with insert sizes of 400 bp by using Illumina TruSeq Nano DNA LT Library Preparation Kit. Each library was sequenced by Illumina HiSeq X-ten platform (Illumina, USA) with PE150 strategy.

To ensure data accuracy and obtain clean reads, raw sequences were processed the quality control steps. Low-quality reads (with an average quality <35 or containing more than 16 unclear bases) were removed using Fastp (Chen et al., 2018) (version 0.21.0). Duplicate reads were eliminated using FastUniq (Xu et al., 2012) (version 1.1.0). Sequences of human origin were filtered out by mapping to the human reference genome (hg38) using Bowtie2 (Langdon, 2015) (version 2.3.5) with the sensitive mode. Afterwards, the clean reads were assembled to contigs using Megahit (Li et al., 2015) (version 1.2.9) with default settings. Contigs longer than 500 bp were used to perform the gene prediction with MetaGeneMark (Zhu et al., 2010) (version 3.38). After that, the hmmer (Mistry et al., 2013) (version 3.3.1) based on the AntiFam database (Eberhardt et al., 2012) was applied to remove spurious genes.

A gene catalog was established using mmseq2 (Steinegger and Söding, 2017) (version 13.45111) with the minimum sequence identity >95% and clustering threshold >90%. The function annotation for gene catalog was performed by eggNOG-mapper (Huerta-Cepas et al., 2017) (version 2.0.1). Only genes annotated within the taxonomy scopes of bacteria, fungi, and viruses were retained to ensure the accuracy of annotations and prevent contamination of host genes. For each gene, reads from the respective sample were mapped using BWA (Li and Durbin, 2009) (version 0.7.17), and the gene abundance was calculated according to the mapped read number divided by gene length. Finally, the abundance of genes annotated as the same ortholog (OG) were added up and classified to different OG functional categories.

2.7. Binning, genome quality assessment and SGBs profiling

Metagenome assembled genomes (MAGs) were generated using MetaBAT2 (Kang et al., 2019) (version 2.12.1). The quality of MAGs was evaluated with CheckM (Parks et al., 2015) (version 1.2.0) for completeness and contamination as microbial genomes. The high or median quality MAGs (HM-MAGs) with >50% completeness and <5% contamination were retained. The retained HM-MAGs were clustered to species-level genome bins (SGBs) based on average nucleic acid identity $\geq$ 95% using dRep (Olm et al., 2017) (version 3.4.0). The GTDB-tk (Chaumeil et al., 2019) (version 1.7.0) was used for taxonomic identification with the Genome Taxonomy Database (r202). A phylogenic analysis was inferred from the concatenation of 120 ubiquitous single-copy marker genes (bac120), and the bac120 reference tree was then inferred with FastTree2 (Price et al., 2010) (version 2.1.10). ITol (Letunic and Bork, 2019) (version 4.3.1) was used for visualization of the phylogenic tree. The relative abundance of each SGB was calculated using CoverM (version 0.6.1).

2.8. Functional analysis of SGBs

All HM-MAGs in each SGB were annotated using Prokka (Seemann, 2014) (version 1.13), and protein hydrolase (PE) sequence of each MAG were extracted from the results of Prokka by searching 'protease' or 'proteinase'. The Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) (Gerlt et al., 2015) was employed to construct a sequence similarity network (SSN) for PE homologs, which were also predicted signal peptides and transmembrane structures using SignalP (Nielsen, 2017) (version 6.0) and TMHMM2 (version 6.0), respectively. If the a PE has signal peptides but no transmembrane structure, it was determined to be the potential secretory PE. The functions of each PE were determined by selecting the most annotations.

The carbohydrate-active enzyme (CAZyme) profile of each SGB was predicted according to the CAZyme database using dbCAN2 (Zhang et al., 2018) which utilized Diamond (Buchfink et al., 2015), HMMER (Potter et al., 2018), and eCAMI (Xu et al., 2020) to predict CAZymes. To make the results accurate, only the same annotations predicted by all three methods were retained. Referring to the research of Tailford et al. (2015), all CAZymes were grouped into two categories ('Mucin degradation' and 'Others'). Annotated genes involved in sulfate conversion pathway were also selected according to Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000). Based on the information available for the genes, their original SGBs were determined.

2.9. Other statistical analysis and plot visualization

Procrustes Analysis applied by the R package 'vegan' of the correlation between SGBs and OGs was performed based on the matrix obtained from the package UMAP (version 0.5.2) in python. All statistical tests were conducted using 'scipy' or 'statsmodels' package in python, and data visualization was applied using 'ggplot2' in R or 'matplotlib' in python.

3. Results

3.1. High-protein diet effects on gut microbial diversity and assembly in different niches

The mice fed with HPD and SD were dissected, and the contents of their gut were sequenced to find out the microbial composition in different niches. The results showed that the high-protein diet significantly changed the microbial composition of the intestine (Fig. 1A, HPD vs. SD $p < 0.0001$). The microbiome of different niches was different, but the difference among individuals was almost negligible (Fig. 1A). Next, the diversity of different niches was explored. The alpha-diversity of microbial communities in different parts (cecum, colon and feces) altered to some extent. Here, the Shannon-Wiener index was higher in the colon, but the ACE index was lower. These two alpha diversity indexes are considered from different angles. The former takes into account the richness and evenness, while the latter only focuses on the richness of microbes. Therefore, it showed that the impact of HPD led to the change of alpha-diversity of intestinal microbial communities, and the richness of microbial communities in the colon decreased

significantly with the evenness increased (Fig. 1B and C). In addition, the beta-diversity in the colon increased significantly (Fig. 1D, $p < 0.01$), and this change was most obvious in the colon (Fig. 1E, $p < 0.001$).

By decomposing the beta-diversity index, we found that the formation of microbial communities in the three niches was dominated by the replacement process (with an average explanation of 72%), especially in the colon (Fig. 1F). In order to explore the specific process of community construction, we quantified the influence of selection, dispersal and drift on the construction of intestinal microbial community according to the model of Stegen et al. (2013). This model provided a method to better understand how microorganisms form different communities in the changing intestinal environment from the ecological point of view, and provided quantifiable values to evaluate the effects of different community assembly processes. In our study, the intestinal microbes in different niches are determined by both deterministic and stochastic processes, in which the stochastic process is dominant, but the construction process in different niches is slightly different (Fig. 1G). Overall, all three niches are dominated by the drift process, but under the impact of HPD, the microbial community in the colon is more determined by species replacement, which leads to the microbial
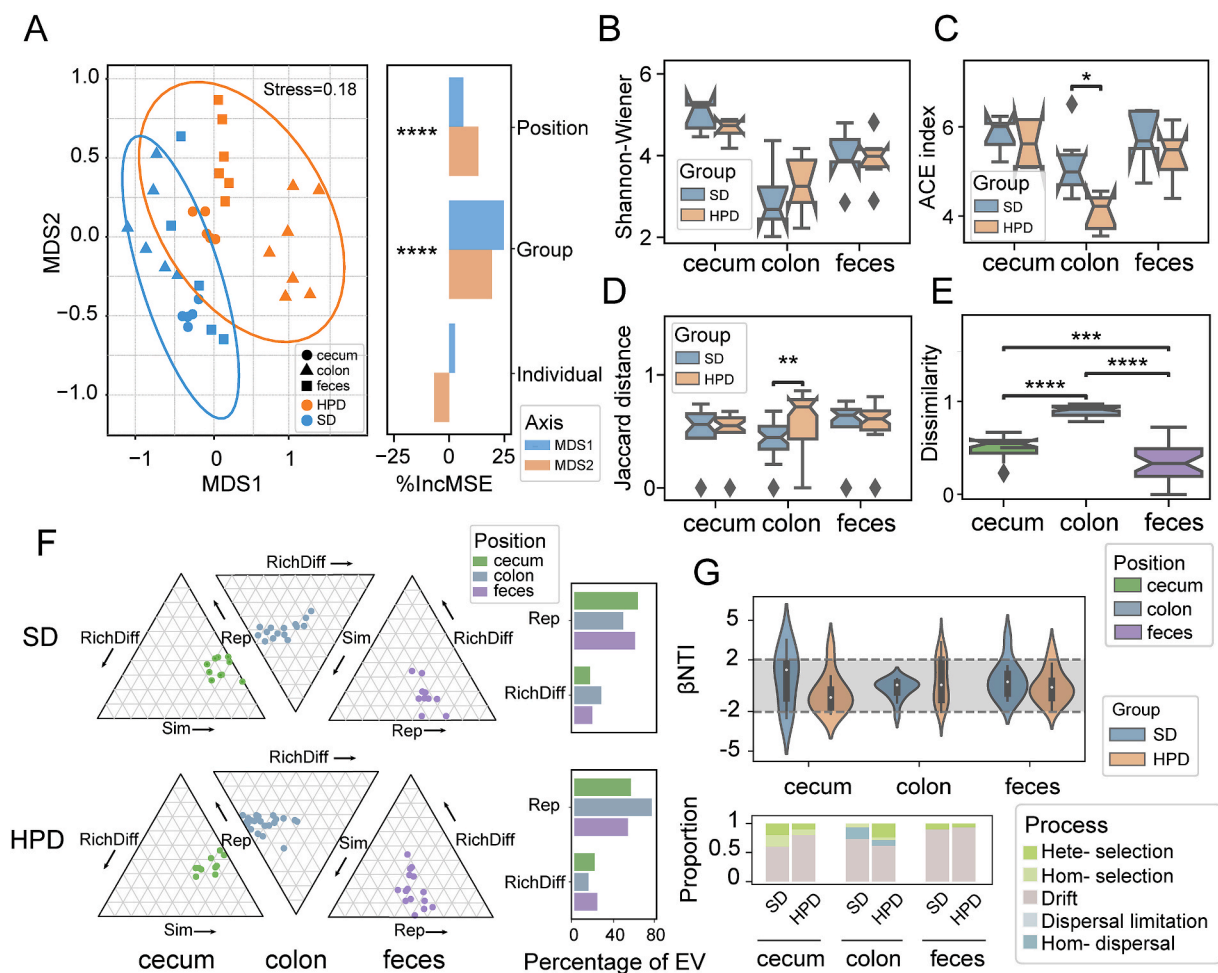
Fig. 1. Alteration of intestinal microbes in different niches under high-protein diet. (A) Left: Microbial composition in three positions (cecum, colon and feces) from two diet groups (HPD and SD) with NMDS. Right: The Mean Decrease Accuracy (%IncMSE) indicates the affected size of each feature in distinguishing the different point at axis of MDS1 and MDS2 (**** represents p < 0.0001, permutational multivariate analysis of variance by Adonis, PERMANOVA). **(B and C)** The Shannon-Wiener and ACE index are calculated to estimate the alpha-diversity, respectively. **(D)** The Jaccard distance was used to estimate the beta-diversity and **(E)** dissimilarity of microbiome between two diet groups in different position (Independent *t*-test with Bonferroni correction was used. *, **, *** and **** represent p < 0.05, p < 0.01, p < 0.001 and p < 0.0001 respectively). **(F)** Triangular plots of beta-diversity comparisons (using Jaccard dissimilarity index) for microbial communities. The position of each point is determined by a triplet of values from the Sim (similarity), Rep (replacement) and RichDiff (richness difference) matrices. The percentages of explained variance (EV) for these matrices were also shown. **(G)** The Beta Nearest Taxon Index (βNTI) distributions and proportions of community assembly processes of two diet groups in different positions are shown. The shade of grey represents neutral processes (|βNTI | < 2). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

community construction in the colon (Fig. 1G).

3.2. Alteration of co-occurring microbes under high-protein diet

To further understand the influence of microbial interaction, the co-occurring microbes were analyzed in the gut. Firstly, we measured the universality of microbial interaction in two diet groups using DOC model (Bashan et al., 2016). This model can detect host-independent or host-specific microbial co-occurring pattern. The inter-individual variability may reflect host-specific dynamics. If a negative correlation is detected between the overlap and dissimilarity of high overlap values, it indicates that the community has similar (universal) dynamics in interspecific interaction. And the *fns* is the score in this model to quantify the strength of the microbial interaction influence. The higher *fns* means more universal co-occurring microbial community. Therefore, the results showed that the universal dynamics of microbial interaction in intestinal tract did not disappear under HPD (Fig. S1). However, the universal dynamic under HPD was significantly lower than that of SD, and the size of consistent symbiotic flora was reduced (Fig. 2A and B, *p* < 0.05). In addition, the complexity of co-occurrence network of intestinal microbes is obviously reduced under HPD, which confirms the shrinkage of the symbiotic system of intestinal microorganisms and the change of symbiotic mode under HPD from another angle (Fig. 2C).

In order to further study the influence of HPD on intestinal microbial symbiotic community, a random matrix theory based (RMT-based) microbial co-occurrence network was constructed with the greedy algorithm to determine the module (Fig. S2). There are 105 connected nodes in this network, and these nodes are mainly positively correlated, accounting for 69% (178/257) of all connections. The network is divided into four large modules, of which M3 contains the most nodes and connections (Fig. 2D). The Lefse algorithm was used to obtained a total of 205 marker OTUs in different niches with HPD and SD (Fig. S3). In addition, these marker OTUs were significantly enriched in the linked part of the network, which suggested that the microbial changes caused by HPD might have a great influence on microbial interaction (Fig. 2E, *p* < 0.001). Next, we paired these marker OTUs with modules in networks, and found that marker OTUs in colon under HPD and SD were enriched in M1 and M2 respectively, while that in cecum under HPD and SD were enriched in M0 and M3 respectively (Fig. 2F, Table S1). This implies the relationship between different co-occurring microbes and diets (Fig. 2F).

3.3. High-protein diet mainly reshape the specific colonic microbial species

To explore the effect of HPD on the microbiome from different niches at the species level, the reads of metagenomic sequencing were assembled to contigs, and clustered to a total of 129 species-level genome bins (SGBs, Fig. 3A). After quantitative analysis of these SGBs, the highest number of different species was observed in the colon site (n = 83), followed by feces (n = 19), and the least in the cecum (n = 5), which is consistent with the previous result (Fig. 1E) that the microbiome in the colon was most affected by HPD (Fig. 3A and B). Although the different SGBs were dominated by Firmicutes overall (Fig. 3B), the dissimilarity still existed at different niches. For example, the species that had significantly different abundance in the colon and feces partially overlapped (13/83), while this overlap didn't appear between the cecum and feces, which may be influenced by the distance of the digestive tract (Fig. 3B). Alter HPD, the most decreased species were concentrated in
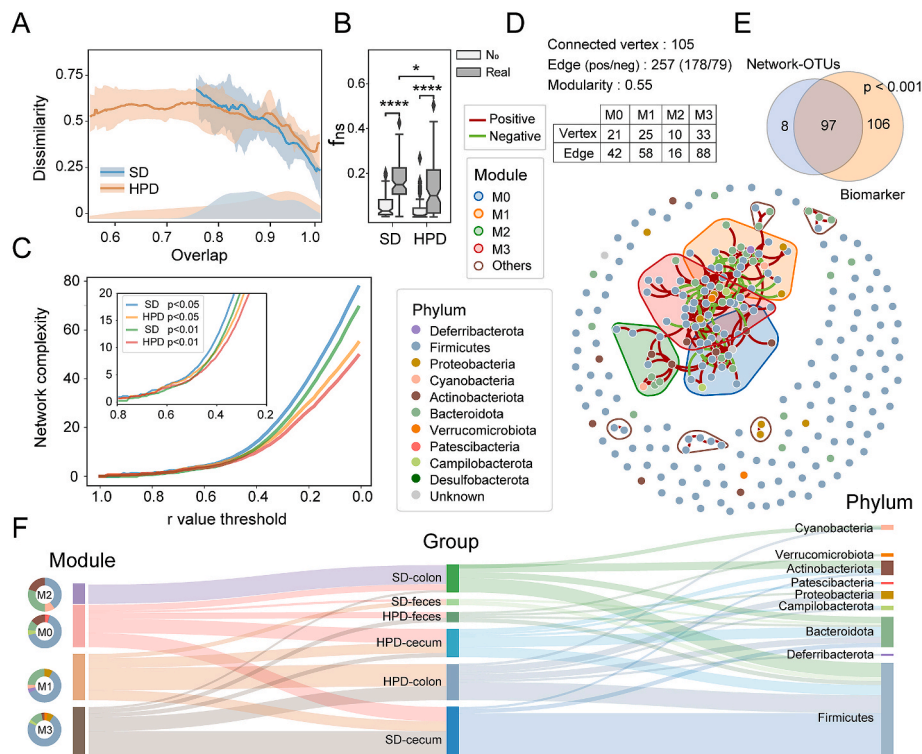
Fig. 2. Variation of co-occurrence microbes under high protein diet. (A) Dissimilarity–overlap curve (DOC) of two diet groups, **(B)** and the *fns* (fraction of data points for which the DOC displays a negative slope) is used to quantify the universality of microbial interaction. The N_0 and Real mean Random null-model and real-data model, respectively (* and **** represent p < 0.05 and p < 0.0001, two-sided Mann-Whitney-Wilcoxon test with Bonferroni correction). **(C)** Network complexity of two diet groups with different thresholds. **(D)** RMT (Random matrix theory) is used to build a robust RMT-based co-occurrence network with Fruchterman-Reingold layout, and the network was divided into four main modules which marked different covering colors. **(E)** The Venn plot shows the overlap of the linked-OTU in **D** and the biomarkers detected by Lefse. The Fisher exact test was used to calculate the p value. **(F)** The correspondence in four modules in **D** and different groups was shown in the Sankey diagram. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
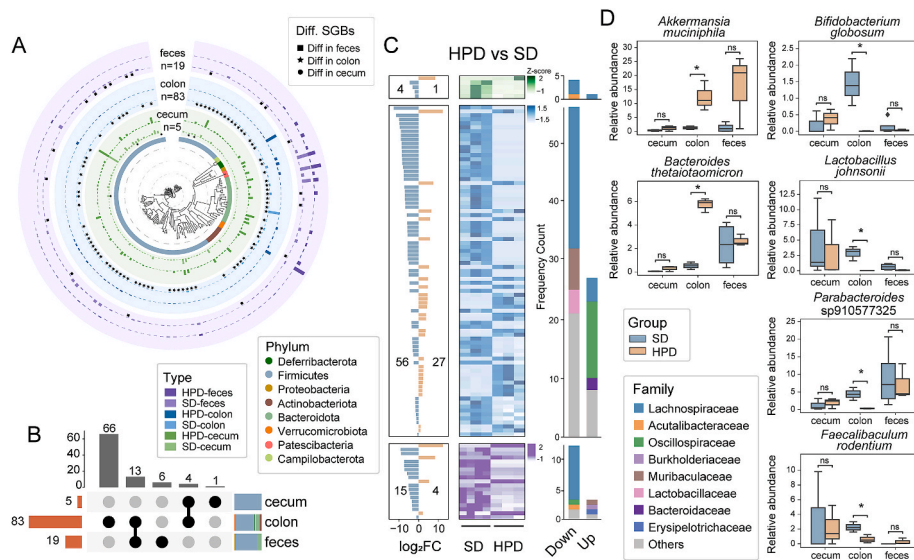
Fig. 3. Difference of species-level microbes in two diet groups. (A) The phylogenetic tree of the gut microbiome built with 129 bacterial representative genomes. Two sided Mann-Whitney rank test (p < 0.05) was used to detect significant different species between HPD and SD groups in different position. The columns in the outer circle represent the relative abundance of each species. **(B)** The Upset plot shows the overlap of different species from three positions. The red column on the left shows the number of different species, and the histogram on the right shows the composition of these different species at phylum level. **(C)** The different species in these different parts are divided into two categories by the log₂FC (log₂ Fold Change). On the right, the family-level subordination of microbial species in different parts is shown. Only the top three are displayed, and the rest are classified as 'Others'. **(D)** The relative abundances of six species were compared using the Mann-Whitney rank test, with significant increases or decreases indicated between the HPD and SD groups. (* represents p value < 0.05 and ns represents non-significant). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Lachnospiraceae, while the most increased species were the member of Oscilospiraceae and Bateroidaceae (Fig. 3C).

Monitoring the microbiome changes in the target species, several essential species were altered after HPD. *Akkermansia muciniphila* and *Bacteroides thetaiotaomicron*, two members related to degrade mucus, accumulated rapidly and maintained significant enrichment in the colon under HPD (Fig. 3D). After HPD, the relative abundance of *Bifidobacterium globosum* and *Lactobacillus johnsonii*, as potential probiotics in the gut decreased. Moreover, other common intestinal commensal bacteria from Parabacteroides or Faecalibaculum, such as *Parabacteroides* sp910577325 and *Faecalibaculum rodentium* also decreased after HPD (Fig. 3D). These results suggested that microbial species in the colonic site underwent major alterations after HPD and this kind of change may relate to intestinal health.

3.4. Various microbes in the intestine provide the substitutability and richness of proteases

The Procrustes analysis showed an agreement between the abundance of Orthologous Groups (OGs) from the microbiome and composition of SGBs (Fig. 4A, p < 0.05). Therefore, the functional differences of samples can be well explained by the various genotypes of SGBs. In addition, the colon had the most functional changes after HPD, while the cecum had the least, which is consistent with the results of microbial species alterations in these sites (Figs. 3A and 4B). To further understand the relationship between altered microbiota and protein metabolism, we constructed a sequence similarity network (SSN) with 1,795 protein hydrolases (PEs) which extracted from pan-genomes of these SGBs (Fig. 4C). And the major 36 of these clusters (>40% sequence identity and >10 PEs) were retained (Table S2). The analysis revealed that the same clusters in SSN representing different functions can be provided by multiple microorganisms (Fig. 4C). Next, we screened these PEs according to the properties of signal peptides and transmembrane structures to count the distribution as potential secretory PEs in different species (Fig. 4D). We found that most of potential secretory PEs are affiliated with Muribaculaceae and Bacteroidaceae. Notably, these two

families are significantly reduced and elevated at the colonic site after HPD, respectively (Fig. 3C).

3.5. Potential carbohydrate competition pattern and related hydrogen sulfide metabolism of colon microbes

In the colon, the enrichment of mucin degraders was observed (Fig. 3D). As the carbohydrates accounted for up to 80% of the total mucin mass (Gendler and Spicer, 1995), we further constructed a pan-carbohydrate degrader profile, and analyzed the distribution of the species involved in mucin degradation (Fig. 5A). Based on the distribution of Carbohydrate-Active enzymes (CAZymes) genes in different species, all species were divided into three patterns (P1, P2 and P3), and each pattern has the potential to metabolize similar carbohydrates. We found that microbial pairs in the same pattern may be subject to microbial replacement in the colonic site. For example, Muribaculaceae has been replaced by Bacteroidaceae after HPD in the colon, and they were both main species in P3 (have the ability to hydrolyze various polysaccharides) (Fig. 5A). Since gut microorganisms that break down mucopolysaccharides can transform sulfate to sulfide in general (Koropatkin et al., 2012), we examined the genes involved in pathway of sulfate conversion in the genomes of all species (Fig. 5B). Notably, over 65% of related genes in this pathway originate from altered species in the colon after HPD (Fig. 5B), and these species also included mucin degraders such as Oscilospiraceae, *B. thetaiotaomicron* and *A. muciniphila* (Fig. 5C).

4. Discussion

More and more studies have shown that the gut microbiome plays an important mediating role between diet and human health (Singh et al., 2017; McQuade et al., 2020; Daniel and McQuade, 2019). Because of high accessibility and noninvasiveness, fecal samples are the most commonly selected specimens in animal and human studies of gut microbiota. However, a recent study identified significant differences between samples of different intestinal regions and feces in bacteria,
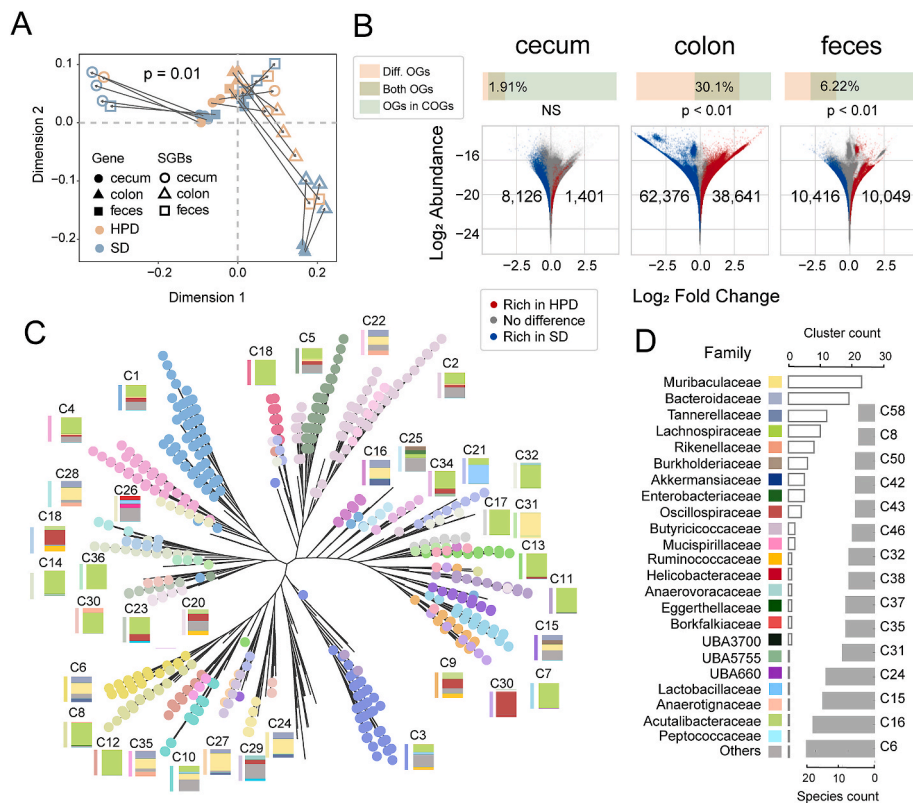
Fig. 4. Functional differences of the gut microbial community under high-protein diet and the substitution of proteases. (A) Procrustes Analysis of the correlation between SGBs and gene functions based on the UMAP results of their abundances ($M^2 = 0.69$, $p = 0.01$, 9,999 permutations). **(B)** The different OGs (Mann-Whitney rank test $p < 0.05$) from three positions between HPD and SD groups were displayed. The top bar plots represent the proportion of these different OGs related to protein metabolism in the COG databases (Hypergeometric distribution test, NS: $p > 0.05$) **(C)** 36 clusters with 10 or more PEs based on a sequence similarity network (SSN, >40% sequence identity) are shown. The histogram next to each cluster in the figure consists of two parts: The thinner column on the left corresponds to the cluster in the phylogenetic tree, and the composition of different families is displayed on the right. Only the top three families are displayed in each cluster, and the rest are classed as 'Others'. **(D)** Diversity of secretory protein types of different families and source species diversity of different clusters in the SSN. The color boxes represent different family corresponds ono-by-one in **C**. On the right side of the figure, 15 clusters with the largest number of species are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

host proteins and metabolites by using ingestible sampling devices (Shalon et al., 2023). In this study, we explored the microbiome in three gut positions (cecum, colon and feces) under HPD and SD from the perspectives of community composition and function. The change of α and β diversity were mainly detected in the colon between HPD and SD. The analysis of the altered microbial community under HPD suggested that the heterogeneous selection and replacement process were possiblely responsible for the alterd microbial community. Similar results were also reported in a previous study that the intestinal microbial community was affected by the substrate nutrition to drive microbial deterministic behavior (Oliphant et al., 2019). In addition, the lower universality and complexity of co-occurrence microbes under HPD demonstrated that the changed diet was also likely related to microbial interaction which was included in the deterministic process (Fargione et al., 2003; Chesson, 2000).

We then examined the relative abundance of species-level microbiome under HPD from different niches using metagenomic sequencing, and counted the significantly different species at the family level. Not surprisingly, the largest number of different species was observed in the colon position. Carbohydrate is both the main component of mucin in the host (Gendler and Spicer, 1995), and the most important carbon source of gut microbiota (Wardman et al., 2022). In this study, HPD led to not enough carbohydrate left in the diet for microbes. Therefore, the ability to degrade the intestinal mucin of the host become the survival superiority for the microbes. The HPD contributed to the increased abundance of mucin degraders such as *A. muciniphila* (Derrien et al., 2004) and *B. thetaiotaomicron* (Mahowald et al., 2009). We also

identified that some essential commensal bacteria such as *B. globosum* and *L. johnsonii* were reduced in the colon. This difference in different positions of the intestine may be due to the construction of different niches with the gradual change of food in the digestive tract (Ailioaie and Litscher, 2021).

To further study the alteration of microbial functions from SD to HPD, a series of analyses were performed about functional genes. The results showed that the altered functional genes relative to protein metabolism were mainly enriched in the colon, and species replaced within similar CAZymes distributed pattern. As one of the substrate to protect intestinal mucosa, mucin of the host can be used as an alternative nutritional carbon basis for microbes (Koropatkin et al., 2012). In the Pattern 3 (P3), the enrichment of mucin degraders indicating the species replacement process dominated by mucopolysaccharide due to scanty carbon sources. Previous study reported that HPD can increase the content of hydrogen sulfide in the large intestine and cause damage to intestinal cells (Beaumont et al., 2016). Our data also showed the potential of producing hydrogen sulfide from altered microbes under HPD. The genomic structure that can predict metabolite dynamics (Gowda et al., 2022), the substitutability of protease from different microbes, and the potential laws of microbial replacement about CAZymes prompted us to put forward such a hypothesis:

Excessive intake of protein led to the entry of protein residues into the colon. The selecteive pressure of microbes from protein decreased, while the metabolic capacity of carbohydrates became more dominant selective factors (Fig. S4). This shift of the colon enviroment can cause the replacement of co-occurrence microbial species. The species

Fig. 5. Characterization of species which metabolize carbohydrate or produce hydrogen sulfide. **(A)** The carbohydrate enzymes (columns) colored by the number of genes per genome (rows) with $\log_{10}$ scaling of each species annotated with the CAZy database. The rows and columns are clustered by 'single' method using Euclidean metric. According to the clustering results of the row data, species are divided into three patterns (P1, P2 and P3). Above and to the right are frequency distribution histograms with data in columns or rows. On the far right of the plot, some microbes closely related to gut mucosal health are marked. **(B)** The pathway about sulfide metabolism from KEGG is drawn. The intersection of the species involved in this pathway, and the different species in the colon between two diet groups are shown on the right. **(C)** The genes involved in the pathway in **B** and the species to which they belong are shown. The species with a significant decrease or increase status in the colon after high-protein diet are marked in blue or red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

replacement process was driven by three patterns based on the distribution of CAZymes. In the first pattern (P1), *Adlercreutzia* (anti-inflammatory, reduced in the ulcerative colitis patients (Galipeau et al., 2021)) was replaced by *M. schaedleri* (Herp et al., 2021) and *Helicobacter* (Zhao et al., 2022) which were related to Crohn's disease. In the second pattern (P2), the common intestinal symbiotic bacteria Lachnospiraceae and Lactobacillaceae were replaced by Oscillospiraceae (considered to decompose the intestinal mucosa (Raimondi et al., 2021)) and *E. coli* (increased the permeability of intestinal epithelial cells (Ma et al., 2017)). In the third pattern (P3), most kinds of carbohydrates can be utilized, and Muribaculaceae, which is the main protein decomposing ability, is replaced by Bacteroidaceae and *A. muciniphila*, which are considered as the main intestinal mucosal decomposers (Derrien et al., 2004). With the enrichment of intestinal mucosal bacteria which can generate hydrogen sulfide, the intestinal mucosa becomes thinner and damaged (Koropatkin et al., 2012).

Collectively, this study highlighted the relationship between HPD and the perturbation of the colon microbiome. The relative abundance of mucin degraders increased in the colon after HPD, which may cause mucosal injuries in the gut (Yoshihara et al., 2020; Slomiany and Slomiany, 1991; Chen et al., 2021a). Additionally, the colon-increased microbes can encode genes involved in the hydrogen sulfide metabolism, suggesting the potential accumulation of hydrogen sulfide, which was unanimous with the previous report (Beaumont et al., 2016). Interestingly, this kind of community re-assembly behavior under HPD was likely to be driven by CAZymes, rather than proteinases. Therefore, we offered a novel opinion to understand the comprehensive alteration of the gut microbial structure adaptation under HPD in terms of genomic functions. A variety of CAZymes and proteases provided a wide range of adaptability for intestinal microbes, while redundant nitrogen sources promoted carbon sources to become the primary selective nutrient sources. Under this mode the process of microbial replacement with the same carbon-based niche occurred, and the newly assembled

community may destroy the original microbial homeostasis, causing intestinal damage. Our research lays the foundation for a better understanding of microbial adaption in the large intestine under the high-protein diet, providing insight into the manipulation of intestinal microbiota to improve the host health. It also helps to further elucidate the etiology of gut disorders.

5. Conclusion

In this study, we investigated the microbiome present in three different sections of the gut: cecum, colon, and feces under high-protein diet (HPD) and standard diet (SD). We found significant changes in α and β diversity primarily in the colon when comparing HPD and SD. In addition, the lower universality and complexity of co-occurring microbes were observed under HPD. Upon analyzing functional genes associated with protein and carbohydrate hydrolysis, the results showed that the carbon source derived from low-carbohydrate content, in contrast to high-protein content, may be a potential factor contributing to the enrichment of mucin degraders and desulphation processes in the colon under HPD. This work will help to deeply understand the relationship between high-protein diet and gut microbes, and take a step toward a hypothesis-driven, theoretical understanding of the human gut microbiome.

CRediT authorship contribution statement

Zheng Yu and **Shuijing Chen**: conceived the study. **Yiming Zhao**: analyzed the data. **Lulu Chen**: performed the experiments. **Yiming Zhao, Huang Jing, Siqi Yao, Liyu Chen** and **Lulu Chen**: wrote the manuscript. All authors contributed to the article and approved the submitted version.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crfs.2023.100600.

References

Abdallah, A., Elemba, E., Zhong, Q., Sun, Z., 2020. Gastrointestinal interaction between dietary Amino acids and gut microbiota: with special Emphasis on host nutrition. Curr. Protein Pept. Sci. 21, 785–798.

Ailioaie, L.M., Litscher, G., 2021. Probiotics, photobiomodulation, and disease management: Controversies and challenges. Int. J. Mol. Sci. 22, 4942.

Bashan, A., Gibson, T.E., Friedman, J., Carey, V.J., Weiss, S.T., Hohmann, E.L., et al., 2016. Universality of human microbial dynamics. Nature 534, 259–262.

Beaumont, M., Andriamihaja, M., Lan, A., Khodorova, N., Audebert, M., Blouin, J.-M., et al., 2016. Detrimental effects for colonocytes of an increased exposure to luminal hydrogen sulfide: the adaptive response. Free Radic. Biol. Med. 93, 155–164.

Blachier, F., Beaumont, M., Portune, K.J., Steuer, N., Lan, A., Audebert, M., et al., 2019. High-protein diets for weight management: interactions with the intestinal microbiota and consequences for gut health. A position paper by the my new gut study group. Clin. Nutr. 38, 1012–1022.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.

Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36, 1925–1927.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.

Chen, L., Wang, J., Yi, J., Liu, Y., Yu, Z., Chen, S., et al., 2021. Increased mucin-degrading bacteria by high protein diet leads to thinner mucus layer and aggravates experimental colitis. J. Gastroenterol. Hepatol. 36, 2864–2874.

Chesson, P., 2000. Mechanisms of maintenance of species diversity. Annu. Rev. Ecol. Systemat. 31, 343–366.

Daniel, C.R., McQuade, J.L., 2019. Nutrition and cancer in the microbiome era. Trends in Cancer 5, 521–524.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., et al., 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505, 559–563.

Derrien, M., Vaughan, E.E., Plugge, C.M., de Vos, W.M., 2004. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int. J. Syst. Evol. Microbiol. 54, 1469–1476.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14, 927–930.

Eberhardt, R.Y., Haft, D.H., Punta, M., Martin, M., O'Donovan, C., Bateman, A., 2012. AntiFam: a tool to help identify spurious ORFs in protein annotation. Database, 2012.

Estaki, M., Jiang, L., Bokulich, N.A., McDonald, D., González, A., Kosciolek, T., et al., 2020. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. Curr. Protocol. Bioinformat. 70, e100.

Fargione, J., Brown, C.S., Tilman, D., 2003. Community assembly and invasion: an experimental test of neutral versus niche processes. Proc. Natl. Acad. Sci. USA 100, 8916–8920.

Friedman, J., Alm, E.J., 2012. Inferring correlation networks from genomic survey data. PLoS Comput. Biol. 8, e1002687.

Galipeau, H.J., Caminero, A., Turpin, W., Bermudez-Brito, M., Santiago, A., Libertucci, J., et al., 2021. Novel fecal biomarkers that precede clinical diagnosis of ulcerative colitis. Gastroenterology 160, 1532–1545.

Gendler, S.J., Spicer, A.P., 1995. Epithelial mucin genes. Annu. Rev. Physiol. 57, 607–634.

Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., et al., 2015. Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. Biochim. Biophys. Acta Protein Proteomics 1854, 1019–1037.

Gowda, K., Ping, D., Mani, M., Kuehn, S., 2022. Genomic structure predicts metabolite dynamics in microbial communities. Cell 185, 530, 46. e25.

Herp, S., Durai Raj, A.C., Salvado Silva, M., Woelfel, S., Stecher, B., 2021. The human symbiont *Mucispirillum schaedleri*: causality in health and disease. Med. Microbiol. Immunol. 210, 173–179.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., Von Mering, C., et al., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34, 2115–2122.

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

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al., 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7, e7359.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., et al., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464.

Koga, Y., 2022. Microbiota in the stomach and application of probiotics to gastroduodenal diseases. World J. Gastroenterol. 28, 6702.

Koropatkin, N.M., Cameron, E.A., Martens, E.C., 2012. How glycan metabolism shapes the human gut microbiota. Nat. Rev. Microbiol. 10, 323–335.

Langdon, W.B., 2015. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. BioData Min. 8, 1–7.

Letunic, I., Bork, P., 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47, W256-W9.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., Knight, R., 2012. Diversity, stability and resilience of the human gut microbiota. Nature 489, 220–230.

Ma, N., Tian, Y., Wu, Y., Ma, X., 2017. Contributions of the interaction between dietary protein and gut microbiota to intestinal health. Curr. Protein Pept. Sci. 18, 795–808.

Mahowald, M.A., Rey, F.E., Seedorf, H., Turnbaugh, P.J., Fulton, R.S., Wollam, A., et al., 2009. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc. Natl. Acad. Sci. USA 106, 5859–5864.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10–12.

McQuade, J.L., Ologun, G.O., Arora, R., Wargo, J.A., 2020. Gut microbiome modulation via fecal microbiota transplant to augment immunotherapy in patients with melanoma or other cancers. Curr. Oncol. Rep. 22, 1–9.

Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., Punta, M., 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41, e121–e.

Nielsen, H., 2017. Predicting Secretory Proteins with SignalP. Protein Function Prediction: Methods and Protocols, pp. 59–73.

Ning, D., Yuan, M., Wu, L., Zhang, Y., Guo, X., Zhou, X., et al., 2020. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. Nat. Commun. 11, 4717.

Oliphant, K., Parreira, V.R., Cochrane, K., Allen-Vercoe, E., 2019. Drivers of human gut microbial community assembly: coadaptation, determinism and stochasticity. ISME J. 13, 3080–3092.

Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 11, 2864–2868.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055.

Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., Finn, R.D., 2018. HMMER web server: 2018 update. Nucleic Acids Res. 46, W200–W204.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., et al., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596.

Raimondi, S., Musmeci, E., Candeliere, F., Amaretti, A., Rossi, M., 2021. Identification of mucin degraders of the human gut microbiota. Sci. Rep. 11, 11094.

Rajilić-Stojanović, M., Smidt, H., De Vos, W.M., 2007. Diversity of the human gastrointestinal tract microbiota revisited. Environ. Microbiol. 9, 2125–2136.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahe, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., et al., 2011. Metagenomic biomarker discovery and explanation. Genome Biol. 12, 1–18.

Shalon, D., Culver, R.N., Grembi, J.A., Folz, J., Treit, P.V., Shi, H., et al., 2023. Profiling the human intestinal environment under physiological conditions. Nature 1–11.

Singh, R.K., Chang, H.-W., Yan, D., Lee, K.M., Ucmak, D., Wong, K., et al., 2017. Influence of diet on the gut microbiome and implications for human health. J. Transl. Med. 15, 1–17.

Slomiany, B., Slomiany, A., 1991. Role of mucus in gastric mucosal protection. J. Physiol. Pharmacol. 42.

Sonnenburg, J.L., Bäckhed, F., 2016. Diet–microbiota interactions as moderators of human metabolism. Nature 535, 56–64.

Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., et al., 2013. Quantifying community assembly processes and identifying features that impose them. ISME J. 7, 2069–2079.

Steinegger, M., Söding, J., 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. 35, 1026–1028.

Suzuki, T.A., Fitzstevens, J.L., Schmidt, V.T., Enav, H., Huus, K.E., Mbong Ngwese, M., et al., 2022. Codiversification of gut microbiota with humans. Science 377, 1328–1332.

Tailford, L.E., Crost, E.H., Kavanaugh, D., Juge, N., 2015. Mucin glycan foraging in the human gut microbiome. Front. Genet. 6, 81.

Tilg, H., Cani, P.D., Mayer, E.A., 2016. Gut microbiome and liver diseases. Gut 65, 2035–2044.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., Gordon, J.I., 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444, 1027–1031.

Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E.E., van der Heijden, M.G., 2019. Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. Nat. Commun. 10, 4841.

Wardman, J.F., Bains, R.K., Rahfeld, P., Withers, S.G., 2022. Carbohydrate-active enzymes (CAZymes) in the gut microbiome. Nat. Rev. Microbiol. 20, 542–556.

Watts, S.C., Ritchie, S.C., Inouye, M., Holt, K.E., 2019. FastSpar: rapid and scalable correlation estimation for compositional data. Bioinformatics 35, 1064–1066.

Windey, K., De Preter, V., Louat, T., Schuit, F., Herman, J., Vansant, G., et al., 2012. Modulation of protein fermentation does not affect fecal water toxicity: a randomized cross-over study in healthy subjects. PLoS One 7, e52387.

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., et al., 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS One 7, e52249.

Xu, J., Zhang, H., Zheng, J., Dovoedo, P., Yin, Y., 2020. eCAMI: simultaneous classification and motif identification for enzyme annotation. Bioinformatics 36, 2068–2075.

Yoshihara, T., Oikawa, Y., Kato, T., Kessoku, T., Kobayashi, T., Kato, S., et al., 2020. The protective effect of *Bifidobacterium bifidum G9-1* against mucus degradation by *Akkermansia muciniphila* following small intestine injury caused by a proton pump inhibitor and aspirin. Gut Microb. 11, 1385–1404.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al., 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46, W95–W101.

Zhao, J., Zhang, X., Liu, H., Brown, M.A., Qiao, S., 2019. Dietary protein and gut microbiota composition and function. Curr. Protein Pept. Sci. 20, 145–154.

Zhao, Y., Chen, L., Chen, L., Huang, J., Chen, S., Yu, Z., 2022. Exploration of the potential relationship between gut microbiota remodeling under the influence of high-protein diet and crohn's disease. Front. Microbiol. 13, 70.

Zhu, W., Lomsadze, A., Borodovsky, M., 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 38, e132–e.

Zmora, N., Suez, J., Elinav, E., 2019. You are what you eat: diet, health and the gut microbiota. Nat. Rev. Gastroenterol. Hepatol. 16, 35–56.