Additional Filews

Title: Dysbiotic shifts in intestinal *Streptococcus*, *Veillonella*, and butyrate-producing bacteria are associated with disease severity in hepatitis E

The attached catalogue:

1. Supplementary Participants & Experimental Methods

Inclusion and Exclusion Criteria for Study Participants

This study enrolled patients with hepatitis E (HE) complicated with liver function abnormality. All participants were recruited from Zhejiang Province.

Inclusion Criteria

1. Aged between 45 and 65 years old.
2. Presented with abnormal liver function indicators, and liver damage was solely caused by hepatitis E virus (HEV) infection.
3. Ethnic Han residents in Zhejiang Province, with consistent daily diet and living patterns.

Exclusion Criteria

1. Co-infection with other hepatitis viruses apart from HEV.
2. Pregnant women and immunocompromised individuals, including solid organ transplant recipients and people living with HIV.
3. Diagnosed with autoimmune liver diseases, biliary tract diseases, gastrointestinal malignancies, hepatocellular carcinoma, alcoholic liver damage, non-alcoholic fatty liver disease, liver cirrhosis, parasitic liver disorders or drug-induced liver injury.
4. Long-term history of heavy drinking, tobacco abuse or illicit drug use.
5. Received antibiotic treatment within 12 weeks prior to enrolment.
6. Underwent any surgical operation within 5 years before participation.
7. Suffered from chronic nutritional, metabolic or immune-related disorders.

Experimental Protocol for Shotgun Metagenomic Sequencing

Total genomic DNA was extracted from human faecal specimens for subsequent metagenomic analysis. A total of 1 μg qualified DNA was taken and fragmented into DNA fragments of approximately 350 bp using an ultrasonic cell disruptor (SCIENTZ08-III, China).

DNA sequencing libraries were constructed using Hieff NGS Ultima Pro DNA Library Prep Kit for Illumina V2 (Yeasen Biotechnology Co., Ltd., Shanghai, China). Unique barcode sequences were appended to each sample for subsequent sample differentiation.

Library quality was verified via Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) and Agilent 2100 Bioanalyzer system. High-throughput sequencing was performed on the Illumina NovaSeq 6000 platform with NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles), following the official operating specifications of the manufacturer.

2. Participant Information Leaflet (Translated Version)

We belong to the State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, National Medical Center for Infectious Diseases and Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, School of Medicine, Zhejiang University. We plan to collect your faecal and peripheral blood samples to explore the correlation between intestinal flora imbalance and the aggravation of hepatitis E, and screen out specific microbial biomarkers associated with disease progression. The research results will provide new references for comprehensive treatment and prevention strategies of hepatitis E in the future.

Please provide samples in accordance with our requirements. All sample collection and related tests are free of charge. All collected data will be used exclusively for academic scientific research. Thank you for your cooperation.

Serial Number: _____

Participant Signature: _____

Date: _____

3. Formal Research Informed Consent Form

To All Potential Participants

This research is carried out by our research team. We will record your general health information, clinical data, collect your faecal and blood samples for scientific analysis. The research aims to clarify the dynamic changes of intestinal microbiota during the progression of hepatitis E.

Your participation is completely voluntary. You have the right to refuse to join this study or withdraw at any time without any negative impact on your routine medical treatment and medical benefits. Before formal enrolment, researchers will conduct eligibility screening. If you agree to participate, please complete the whole research process as far as possible.

The research subjects are patients infected with HEV without other severe concurrent illnesses. We will collect your health information through standardized questionnaires. Sample collection will be conducted at two time points: the day of enrolment and the one-year follow-up after hospital discharge. Faecal samples will be used for microbial DNA extraction and shotgun metagenomic sequencing. Blood samples will be applied for routine blood examination and biochemical detection.

No fees will be charged during the whole research process.

All personal information and research data will be strictly confidential. Your real name will be replaced with coded numbers. Only designated researchers of this project can access the relevant data, and no information will be disclosed to external personnel or institutions.

The research findings may be published in academic journals for academic communication, while all your private information will be fully protected. You can consult researchers about the progress of the study at any time. If you fully understand the above contents and agree to take part in this research, please sign this consent form. One copy of the signed form will be given to you for retention.

Participant's Declaration

I have fully read the content of this informed consent form, or the staff have read and explained all contents to me. I have raised questions about the research, and all doubts have been answered clearly. I understand that participation is voluntary, and refusal or withdrawal will not damage my legitimate rights and interests.

I authorize the research team to check my medical records and research data, and I confirm that all relevant information will be kept confidential.

I voluntarily agree to participate in this scientific research.

(Participants who cannot write shall press their thumbprint in addition.)

Printed Name of Participant: _____

Participant Signature: _____

Participant's Thumbprint: _____
Date (Day/Month/Year): _____

Witness Section

A literate witness is required to sign this form. The witness shall be selected by the participant and shall have no connection with the research team.

Printed Name of Witness: _____
Witness Signature: _____
Date (Day/Month/Year): _____

Researcher's Confirmation

I have fully interpreted all contents of the research notice to the participant, and confirmed that the participant clearly understands the following items:

1. We will collect and record your personal health information and clinical data.
2. All above data will be used for scientific research.
3. The collected faecal and blood samples will be used for experimental analysis in this study.

I confirm that the participant has had sufficient time to ask questions, and all questions have been answered properly. No coercion or inducement was used to obtain this consent.

Printed Name of Researcher: _____
Researcher Signature: _____
Date (Day/Month/Year): _____

4. Study Questionnaire

General Information (For participants aged over 18)

Name: _____ Age: _____ Gender: _____

Height: _____ cm Weight: _____ kg

1. Place of residence
   □ Urban area □ Rural area
2. Serological test results (Tick √ for positive, × for negative)
   □ HAV-IgM □ HAV-IgG
   □ HBV-IgM □ HBV-IgG
   □ HCV-IgM □ HCV-IgG
   □ HDV-IgM □ HDV-IgG
   □ HEV-IgM □ HEV-IgG
3. Daily dietary habits (Multiple choices allowed)
   □ Staple food: rice
   □ Staple food: noodles
   □ Regular intake of meat and vegetables
   □ Regular intake of yoghurt and other dairy products
   □ Long-term excessive drinking
   □ Heavy smoking
   □ History of drug abuse
4. History of antibiotic use
   □ Never used antibiotics
   □ Used antibiotics before, no medication within the latest 12 weeks
   □ Used antibiotics within the latest 12 weeks
5. Regular late-night staying
   □ Yes □ No
6. Digestive tract discomfort (Multiple choices allowed)
   □ Recurrent abdominal distension
   □ Frequent diarrhoea
   □ Acid regurgitation
   □ Recurrent abdominal pain
   □ Constipation
   □ Other digestive diseases: _____
   □ No digestive system discomforts
7. Special health conditions (Multiple choices allowed)
   □ Received solid organ transplantation
   □ HIV infection
8. History of liver diseases (Multiple choices allowed)

□ Autoimmune hepatitis
□ Primary biliary cholangitis
□ Primary sclerosing cholangitis
□ Hepatocellular carcinoma
□ Alcoholic liver disease
□ Drug-induced liver injury
□ Viral hepatitis
□ Non-alcoholic fatty liver disease
□ Liver cirrhosis
□ Parasitic liver disease

5. Extended Data Figures S1–S5

Fig. S1 PERMANOVA and ANOSIM Analysis of Bacterial, Viral, and Archaea Community Composition (Microbial Relative Abundance) in the ANIH, AIH, ALF Groups, and Age-Matched Healthy Controls Based on Jaccard Binary Distance Dissimilarities at the Species Level. Panels **A**, **B**, and **C** show principal coordinate analysis (PCoA) for bacteria, viruses, and archaea, respectively. PERMANOVA was used to assess the separation between groups, with ellipses representing 80% confidence intervals (i.e., encompassing 80% of the points for each group). *P*-values were calculated to determine the significance of group separation after 999 permutations, with $P < 0.05$ indicating statistical significance. Each point represents a sample and is color- and shape-coded by group. Panels **D**, **E**, and **F** show the analysis of similarities (ANOSIM) for bacteria, viruses, and archaea, respectively, using Jaccard binary distance. Box plots represent the interquartile range (25th to 75th percentiles), with the center line indicating the median. Abbreviations: PERMANOVA, Permutational Multivariate Analysis of Variance; ANOSIM, Analysis of Similarities; ALF, Acute Liver Failure; ANIH, Acute Non-Icteric Hepatitis E; AIH, Acute Icteric Hepatitis E; HC, Healthy Controls

Fig. S2 Differentially Abundant Genera Associated with Hepatitis E (HE) Severity.

A Linear discriminant analysis (LDA) effect size (LEfSe) results showing genera that were differentially represented across the four groups (LDA score ≥ 2.5). **B** Results from the Kruskal-Wallis test with Bonferroni correction for differentially abundant genera. Boxplots display the median (black solid line within the box) and the interquartile range (25th to 75th percentiles). The relative abundance of each genus was $\log_{10}$-transformed, and *P*-values were adjusted using the Holm-Bonferroni method for multiple comparisons. Statistical significance is indicated as ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$. Group color coding: Purple boxes: ALF (Acute Liver Failure); Blue boxes: ANIH (Acute Non-Icteric Hepatitis E); Green boxes: AIH (Acute Icteric Hepatitis E); Red boxes: HC (Healthy Controls).

Fig. S3 Differentially Abundant Species Among the ALF, AIH, ANIH, and Healthy Control Groups. **A** Linear discriminant analysis (LDA) effect size (LEfSe) results, showing species that were differentially represented across the four groups (LDA score ≥ 2.5). **B** Results of the Kruskal-Wallis test with Bonferroni correction for differentially abundant species, with average relative abundance for each genus and Wilcoxon matched-pairs signed rank test statistics provided in Supplementary Data S2e.

C Species that were significantly altered between admission and one-year post-discharge follow-up, as determined by the Wilcoxon matched-pairs signed rank test, with average relative abundance for each genus and Wilcoxon test statistics detailed in Supplementary Data S2f. Genera enriched in different groups are color-coded as follows: Purple: Enriched in ALF (Acute Liver Failure), Green: Enriched in AIH (Acute Icteric Hepatitis E), Blue: Enriched in ANIH (Acute Non-Icteric Hepatitis E). Boxplots show the median (solid black line within the box) and the interquartile range (25th to 75th percentiles). The relative abundance of each taxon was log10-transformed, and *P*-values were adjusted using the Holm-Bonferroni method for multiple comparisons. Statistical significance is indicated as ***, $P < 0.001$; **, $P < 0.01$; *, *P*

< 0.05. ALF, Acute Liver Failure; ANIH, Acute Non-Icteric Hepatitis E; AIH, Acute Icteric Hepatitis E; HC, Healthy Controls.

Fig. S4 Differential Carbohydrate-Active Enzyme (CAZYmes) Genes Among Patient Groups and Healthy Controls. **A** The mean relative abundance (left) and associated *P*-values (right) for selected differentially expressed CAZYmes genes in the ALF group compared to healthy controls. **B** Correlations between differentially expressed CAZYmes genes and bacterial species. The boxes on the left represent selected genes that are upregulated in the ALF group relative to healthy controls, while the boxes on the right show the bacterial species identified as discriminatory using LEfSe analysis. The lines connecting the boxes indicate the relationship between each gene and species across the groups. Color coding represents the following groups: Purple: ALF (Acute Liver Failure), Green: AIH (Acute Icteric Hepatitis E), Blue: ANIH (Acute Non-Icteric Hepatitis E), Red: Healthy Controls, Light Blue: Follow-up visits. PL, Polysaccharide Lyases; GT, Glycosyltransferases; GH, Glycoside Hydrolases; CE, Carbohydrate Esterases; CBM, Carbohydrate-Binding Module; ALF, Acute Liver Failure; ANIH, Acute Non-Icteric Hepatitis E; AIH, Acute Icteric Hepatitis E; HC, Healthy Controls; FU, Follow-up.

Fig. S5 Spearman's Correlation Between the Relative Abundance of Fecal Bacterial Species and Major Serum Indicators of Liver Dysfunction. The size and color intensity of the squares represent the magnitude of the correlation. Red squares indicate positive correlations, while green squares indicate negative correlations. Solid red and light blue lines represent significant correlations at the $P \leq 0.01$ level, whereas broken red and light blue lines indicate significant correlations at the $0.01 < P \leq 0.05$ level (red lines for positive correlations and light blue lines for negative correlations). Species colored as follows: Purple: Bacteria enriched in ALF, Green: Bacteria enriched in AIH, Blue: Bacteria enriched in ANIH, Red: Bacteria enriched in HC. Note: The prothrombin time (PT) is mathematically converted to the international normalized ratio (INR). TB, Total Bilirubin; CHE, Cholinesterase; GGT, Gamma-Glutamyl Transferase; DBil, Direct Bilirubin; TBA, Total Bile Acid; AST, Aspartate Aminotransferase; ALB, Albumin; ALP, Alkaline Phosphatase; INR, International Normalized Ratio; ALT, Alanine Aminotransferase.

6. Supplementary Datasets S1–S3

Supplementary Data S1

S1a. Details of the comparisons in the alpha diversity of bacterial, viral, and archaeal communities among the ALF, AIH, ANIH, and healthy control groups using the Kruskal-Wallis rank sum test.

S1b. The r-values of SparCC correlations between microbial genera and results from permutational multivariate analysis of variance (PERMANOVA).

S1c. The p-values of SparCC correlations among the microbial genera.

S1d. The genera information in each co-abundance group (CAG).

S1e. Statistical comparisons summarizing the abundance of co-abundance groups (CAGs) among the ALF, AIH, ANIH, and healthy control groups using Mann–Whitney–Wilcoxon tests.

S1f. Statistical comparisons summarizing the abundance of each differential genus between the intestinal microbiota at admission and that at the one-year post-discharge follow-up visit, using post hoc pairwise Mann-Whitney U tests with Bonferroni correction.

Supplementary Data S2

S2a. The r-values of SparCC correlations between microbial species and results from permutational multivariate analysis of variance (PERMANOVA).

S2b. The *P*-values of SparCC correlations among microbial species.

S2c. Statistical comparisons summarizing the abundance of species co-abundance groups (s-CAGs) among the ALF, AIH, ANIH, and healthy control groups using Mann–Whitney–Wilcoxon tests.

S2d. The species information in each s-CAG.

S2e. Statistical comparisons of each species between the ALF, AIH, ANIH, and control groups using Mann–Whitney–Wilcoxon tests.

S2f. Statistical comparisons summarizing the abundance of each differential species between the intestinal microbiota at admission and that at the one-year post-discharge follow-up visit, using post hoc pairwise Mann-Whitney U tests with Bonferroni correction.

Supplementary Data S3

S3a. The r- and *P*-values for the correlations between viruses and the selected discriminatory bacterial species in LEfSe analysis.

S3b. The abundance and *P*-value shown in Fig. 3B.

S3c. A summary of the statistical information for differential KEGG genes using paired comparisons of the patient groups and healthy controls.