

Supplemental Digital Content

Title: Alteration of the fecal microbiome in patients with cholecystectomy: potential relationship with post-cholecystectomy diarrhea - before and after study -

Methods

1. Bacterial DNA extraction and 16S rRNA sequencing from fecal samples

DNA was extracted using a DNeasy PowerSoil Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The extracted DNA was quantified using a Quant-IT PicoGreen (Invitrogen). To amplify the V3 and V4 regions of the bacterial genomic DNA, the sequencing libraries were prepared according to Illumina 16S Metagenomic Sequencing Library protocols. The input gDNA 2 ng was polymerase chain reaction (PCR) amplified with 5x reaction buffer, 1 mM deoxynucleotide mix, 500 nM each of the universal forward/reverse PCR primers, and Herculase II fusion DNA polymerase (Agilent Technologies, Santa Clara, CA). The cycle conditions for the 1st PCR was 3 min at 95°C for heat activation, followed by 25 cycles each of 30 s at 95°C, 30 s at 55°C, and 30 s at 72°C, followed by a 5-min final extension at 72°C. The universal primer pair with Illumina adapter overhang sequences used for the first amplification was as follows: V3-F, 5'-TCGTCCGACGCTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3' and V4-R 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3'.

The 1st PCR product was purified using AMPure beads (Agencourt Bioscience, Beverly, MA). Following purification, 2 ul of the 1st PCR product was PCR amplified for final library construction containing the index using NexteraXT Indexed Primer. The cycle condition for the 2nd PCR was the same as the 1st PCR condition, except for 10 cycles. The PCR products were purified using AMPure beads. Subsequently, the final purified product was quantified using quantitative PCR (qPCR) according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using TapeStation D1000 ScreenTape (Agilent Technologies, Waldbronn, Germany). Paired-end (2 × 301 bp) sequencing was performed by Macrogen using the MiSeq platform (Illumina, San Diego, USA).

2. Data processing for microbial metagenome analysis

Raw data generated by the Illumina MiSeq platform were demultiplexed using index sequences. Adapter sequences and barcode primers were trimmed using Cutadapt (v3.2) program [1]. Preprocessing was performed using the DADA2 (v1.18.0) package in R (v4.0.3) to denoise sequencing errors and identify amplicon sequence variants (ASVs) [2]. Paired-end reads were truncated to 250 bp forward sequences and 200 bp reverse sequences. Reads with expected errors of >2 were discarded. Error-corrected reads were merged into one barcode sequence, and chimeric sequences were filtered

using the consensus method of DADA2. The resulting ASV reads of each sample were downsized to equate the minimum observed sampling depth using quantitative insights into microbial ecology (Quantitative Insights Into Microbial Ecology, QIIME, v1.9) by random selection of reads to compare microbial diversity [3]. Taxonomic classification was performed using the National Center for Biotechnology Information 16S ribosomal RNA DB using BLAST+ (v2.9.0) [4]. Each ASV was defined as the taxon of the best-hit subject from the blast results. When Query coverage was <85%, and the identity of the matched region was <85%, the ASVs were regarded as unassigned. To identify phylogenetic relationships between ASVs, multiple alignments of reads were conducted using the mafft (v7.475) program [5]. Subsequently, a phylogenetic tree was constructed using FastTreeMP (v2.1.10) [6].

3. Analysis of bacterial composition and diversity

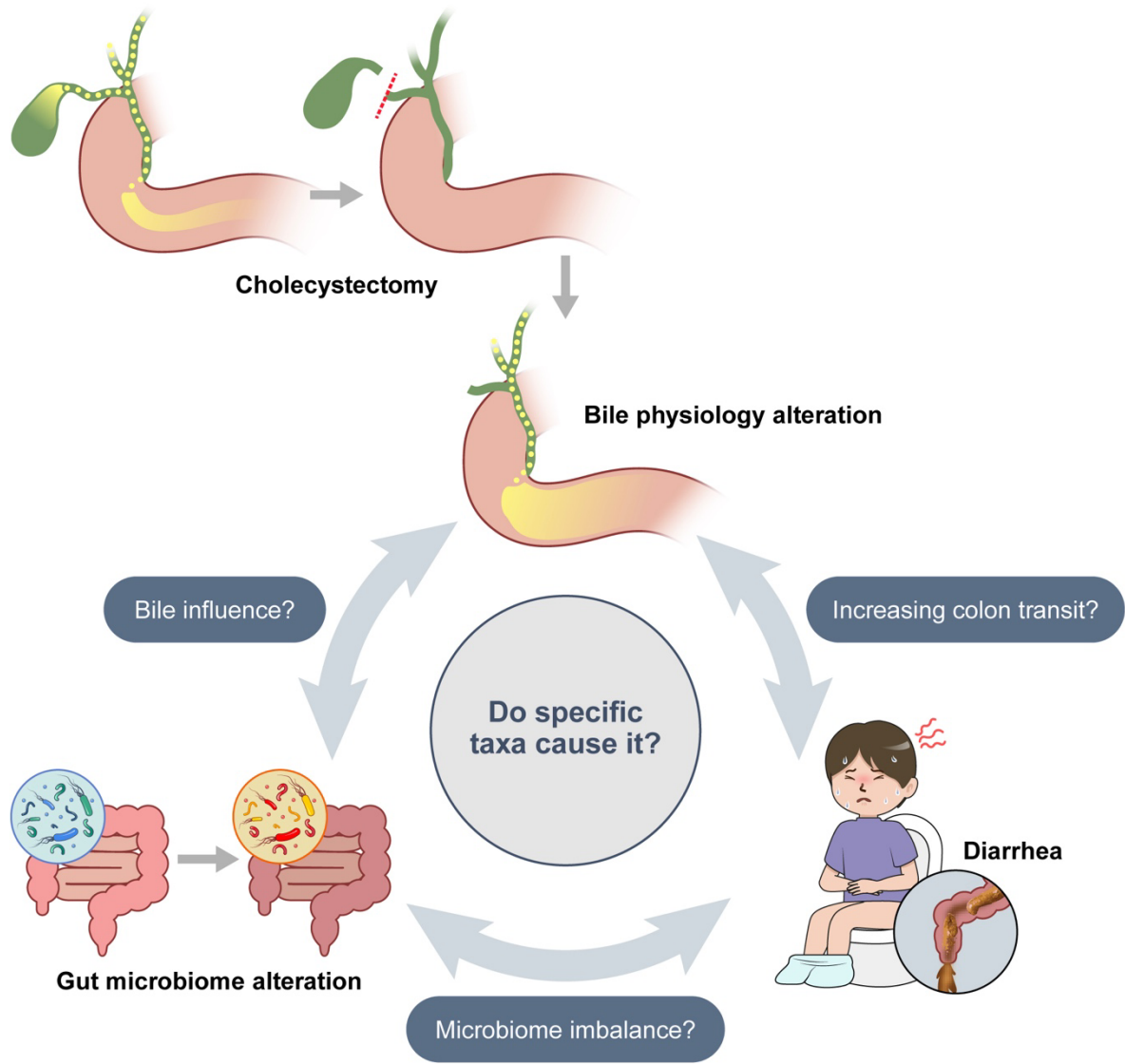
All taxonomic analyses used relative abundances, defined as ratios, to even sampling depth. Abundances at the phylum, genus, and species levels were averaged to determine the taxonomic composition of each group with a bar graph. The Firmicutes to Bacteroidetes (F/B) ratio was calculated based on their abundance for each patient. The Krona chart, which visualizes complex hierarchies of metagenomic classifications, was constructed using Krona Tools (v2.8.1) [7]. The Krona charts illustrate the average in each group at the species level.

Community diversity was analyzed using QIIME (v1.9.0) [3]. To check the diversity and evenness of the microbial community, the Shannon and Gini–Simpson indices were calculated [8]. In addition, we checked that the rarefaction curves were saturated in all samples to confirm that sequencing depths were sufficient to capture the real diversity [9]. Beta diversities were determined based on weighted and unweighted UniFrac distance metrics, which compared microbial communities by measuring phylogenetic distances [10]. The microbial network was inferred using the SpiecEasi package (v1.1.2) in R (v4.1.2) [11]. The inferred networks were then transformed into an igraph (v1.2.11) object to visualize topological properties [12].

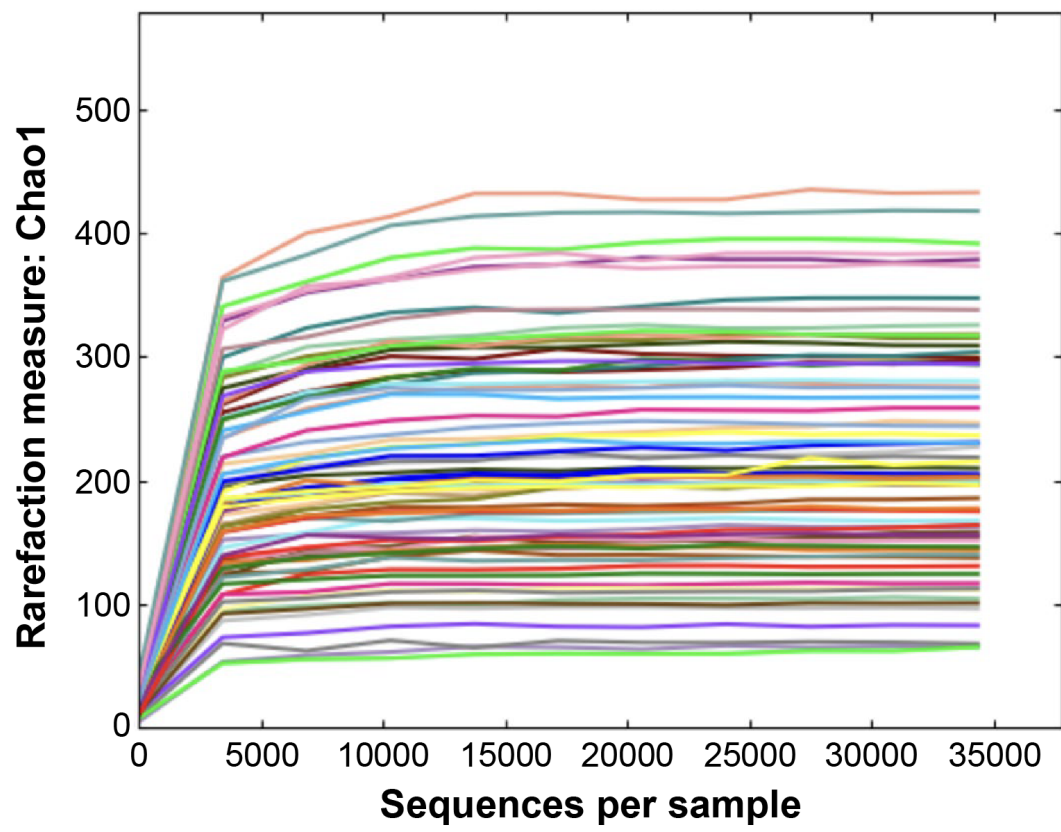
4. Statistical analysis

The Wilcoxon Rank Sum test and Kruskal–Wallis test were conducted to compare the Shannon and Gini–Simpson indices between groups [13, 14]. The false discovery rate of the Benjamini–Hochberg method was performed with a cut-off value of 0.05 to correct errors of multiple testing [15]. The ggpubr R package (v0.4.0) was used to visualize box plots to compare differences in groups for each index [16]. Firmicutes to Bacteroidetes (F/B) ratio is the proportion of two major phyla in the intestinal microflora. An increase or decrease in the F/B ratio indicates an imbalance of the microbiome [17]. The F/B ratio was examined using the Kruskal–Wallis and Dunn tests as a post hoc test with a *P*-value of 0.05 [18]. In addition, compositional differences among samples were visualized by principal coordinate analysis

(PCoA) [19] and the unweighted pair group method with arithmetic mean tree [20] using unweighted/weighted UniFrac distance metrics. In addition, an analysis of similarities (ANOSIM) statistical test was performed with QIIME script to confirm the significance of differences between groups [3]. Microbes that showed significant differences between groups were identified using the Kruskal–Wallis test and their differences were estimated with a cut-off linear discriminant analysis (LDA) score (\log_{10}) of >2 with linear discriminant analysis effect size (LEfSe) [21]. Furthermore, a cladogram with GraPhlAn (v1.1.3.1) [22] and bar plots were generated using the ggplot R package (v3.3.2). Microbes with a mean abundance of all samples $\geq 1\%$ and significant lists observed from LEfSe were visualized using a heatmap to identify associations between sample groups. Relative abundance data were transformed to log-scale form. Significant taxa from LEfSe were indicated with asterisks. The hierarchical structure calculated using the Euclidean distance was shown as a dendrogram at the top of the plot. Moreover, visualization was performed using ggplot2 (v3.3.5) R package [23]. The accuracy of taxa scoring with high LDA was assessed using the area under the receiver operating curve (AUROC). The pROC R package (v1.18.0) was used to plot the Receiver operating characteristic (ROC) and calculate the AUROC values [24]. Subsequently, the significance of the AUROC was verified using the verification R package [25].



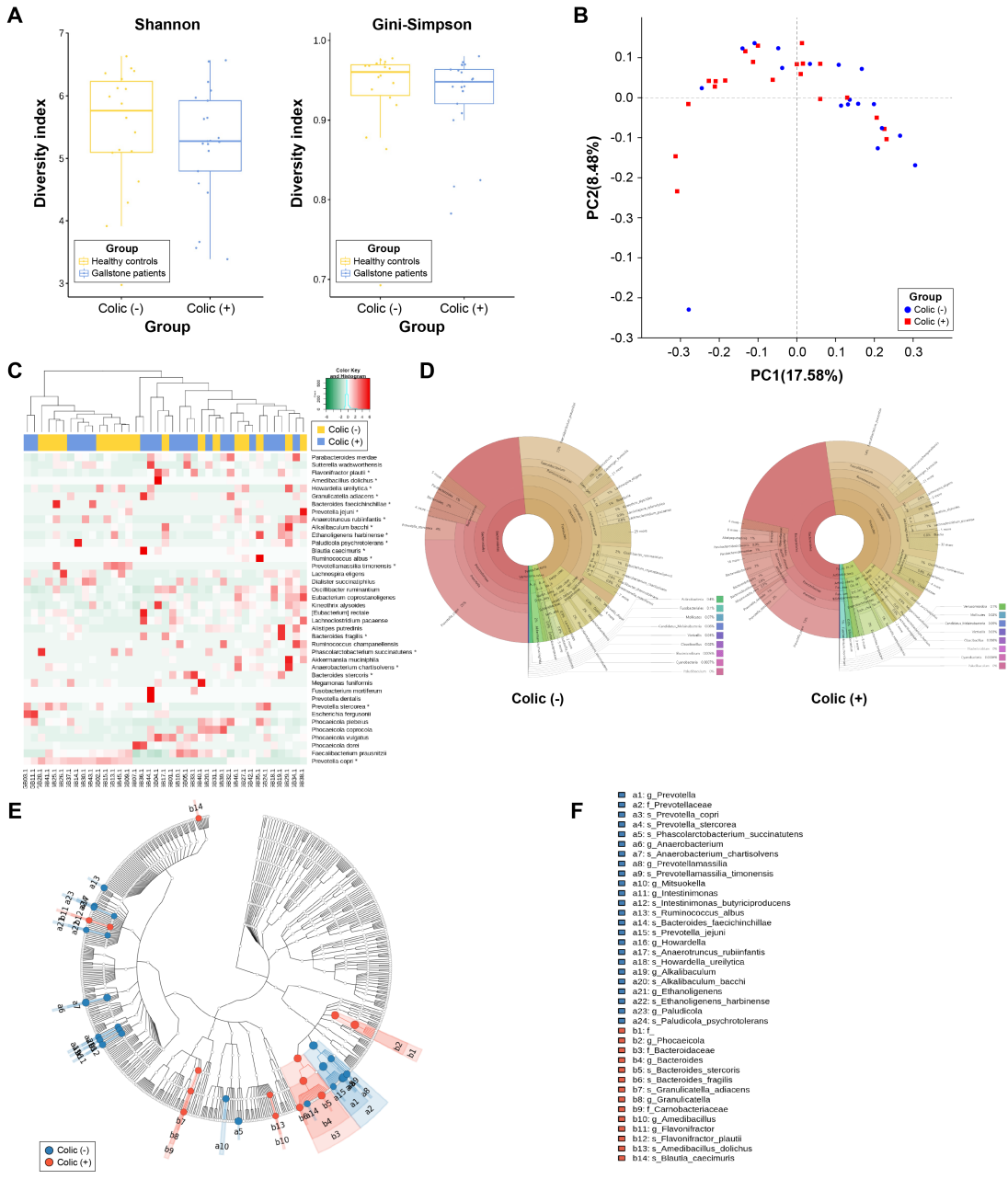
Supplementary Figure 1. Schematic representation of the research hypothesis.



Supplementary Figure 2. Rarefaction curve of all fecal samples.



Supplementary Figure 4. Comparison of the fecal microbiome in patients with gallstones (GS) after cholecystectomy and healthy controls (HC). (A) Comparison of alpha diversity (Shannon and Gini-Simpson indices) **(B)** Unweighted UniFrac principal coordinate analysis. GS After cholecystectomy (red dot) vs. HC (green dot). **(C)** Heat map of taxonomic assignment of fecal samples. The colored columns in the upper part of the heat map indicate GS after cholecystectomy and HC, and those in the lower part of the heat map indicate each participant. Taxonomic abundance is proportional to color intensity (color scale in the upper-left panel of the figure). **(D)** Krona chart illustrating the differential abundance of bacteria in HC and GS after cholecystectomy. **(E)** Cladogram highlighting the distribution of the fecal microbiome with differential abundance. **(F)** Index of the cladogram.



Supplementary Figure 5. Comparison of the gut microbiome in patients with gallstones with typical biliary colic [i.e., colic (+)] and those without symptoms [colic (-)]. (A) Comparison of alpha diversity (Shannon and Gini–Simpson indices) (B) Unweighted UniFrac principal coordinate analysis. Colic (-) (red dot) vs. colic (+) (blue dot). (C) Heat map of taxonomic assignment of fecal samples. The colored columns in the upper part of the heat map indicate patients with colic (-) and colic (+), and those in the lower part of the heat map indicate each participant. Taxonomic abundance is proportional to color intensity (color scale in the upper-left panel of the figure). (D) Krona chart illustrating the differential abundance of bacteria in colic (-) and colic (+). (E) Cladogram highlighting the distribution of the fecal microbiome with differential abundance. (F) Index of the cladogram.

- a1: f_Prevotellaceae
- a2: s_Prevotella_copri
- a3: g_Prevotella
- a4: s_Prevotella_stercorea
- a5: g_Oscillibacter
- a6: f_Eubacteriaceae
- a7: g_Eubacterium
- a8: s_Oscillibacter_ruminantium
- a9: s_Lacrimispora_xylyanolytica
- a10: g_Lacrimispora
- a11: s_Eubacterium_coprostanoligenes
- a12: g_Flintibacter
- a13: s_Flintibacter_butyricus
- a14: s_Intestinimonas_butyriciproducens
- a15: g_Intestinimonas
- a16: g_Sporobacter
- a17: s_Sporobacter_terminidis
- a18: g_Duodenibacillus
- a19: g_Christensenella
- a20: f_Christensenellaceae
- a21: s_Duodenibacillus_massiliensis
- a22: g_Anaerobacterium
- a23: s_Anaerobacterium_chartisolvens
- a24: s_Duncaniella_freteri
- a25: g_Duncaniella
- a26: f_Muribaculaceae
- a27: s_[Eubacterium]_siraeum
- a28: g_Holdemanella
- a29: s_Holdemanella_biformis
- a30: s_Coprococcus_eutactus
- a31: s_Vallitalea_pronyensis
- a32: f_Vallitaleaceae
- a33: g_Vallitalea
- a34: s_Eubacterium_ruminantium
- a35: s_Phocaea_massiliensis
- a36: g_Phocaea
- a37: f_Odoribacteraceae
- a38: g_Desulfohalotomaculum
- a39: s_Anaerotruncus_rubiinfantis
- a40: g_Butyricimonas
- a41: s_Pseudobutyrvibrio_ruminis
- a42: s_Paludicola_psychrotolerans
- a43: g_Paludicola
- a44: g_Pseudobutyrvibrio
- a45: s_Alistipes_finegoldii
- a46: g_Anaerotruncus
- a47: s_Fournierella_massiliensis
- a48: s_Butyricimonas_virosa
- a49: g_Fournierella
- a50: f_Victivallaceae
- a51: s_Anaerotaenia_torta
- a52: o_Victivallales
- a53: s_Victivallis_vadensis
- a54: p_Lentisphaerae
- a55: g_Anaerotaenia
- a56: g_Victivallis
- a57: c_Lentisphaeria
- a58: f_Clostridiales_Family_XIII_Incertae_Sedis
- a59: g_Ethanoligenens
- a60: s_Ethanoligenens_harbinense
- b1: f_
- b2: g_Phocaeicola
- b3: s_Phocaeicola_vulgatus
- b4: f_Sutterellaceae
- b5: o_Burkholderiales
- b6: c_Betaproteobacteria
- b7: g_Sutterella
- b8: g_Blautia
- b9: s_Blautia_wexlerae
- b10: s_Dialister_invisus
- b11: p_Actinobacteria
- b12: o_Bifidobacteriales
- b13: f_Bifidobacteriaceae
- b14: g_Bifidobacterium
- b15: c_Actinomycetia
- b16: s_Sutterella_stercoricanis
- b17: s_Hungatella_effluvii
- b18: s_Bifidobacterium_longum
- b19: g_Tyzzereella
- b20: g_Collinsella
- b21: s_Collinsella_aerofaciens
- b22: o_Coriobacteriales
- b23: f_Coriobacteriaceae

Supplementary Figure 6. The index of Cladogram in PCD (-) vs. PCD (+) patients. Abbreviation: PCD, post-cholecystectomy diarrhea.

References

1. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10-12.
2. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13(7):581-583.
3. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7(5):335-336.
4. Popovici V, Goldstein DR, Antonov J, Jaggi R, Delorenzi M, Wirapati P. Selecting control genes for RT-QPCR using public microarray data. *BMC Bioinformatics.* 2009;10:42.
5. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772-780.
6. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5(3):e9490.
7. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics.* 2011;12:385.
8. Shannon CE, Weaver W. *The mathematical theory of communication.* Urbana, IL: The University of Illinois Press; 1964. Available at: https://pure.mpg.de/rest/items/item_2383164/component/file_2383163/content. Accessed May 18, 2022.
9. Gotelli NJ, Colwell RK. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett.* 2001;4(4):379-391.
10. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol.* 2005;71(12):8228-8235.
11. Kurtz ZD, Muller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol.* 2015;11(5):e1004226.
12. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal.* 2006; *Complex Systems:* 1695. https://www.researchgate.net/profile/Gabor-Csardi/publication/221995787_The_Igraph_Software_Package_for_Complex_Network_Research/links/0c96051d301a30f265000000/The-Igraph-Software-Package-for-Complex-Network-Research.pdf. Accessed May 18, 2022.
13. Bauer DF. Constructing confidence sets using rank statistics. *J Am Stat Assoc.* 1972;67(339):687-690.
14. Hollander M, Wolfe DA, Chicken E. *Nonparametric statistical methods.* 3rd ed. John Wiley & Sons; 2013.
15. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful

- approach to multiple testing. *J R Stat Soc Series B Methodol.* 1995;57(1):289-300.
16. Kassambara A, Kassambara MA. ggpubr: 'ggplot2' Based publication ready plots. The Comprehensive R Archive Network. Available at: <https://cran.r-project.org/web/packages/ggpubr/index.html>. Accessed May 18, 2022.
 17. Ojima M, Motooka D, Shimizu K, et al. Metagenomic analysis reveals dynamic changes of whole gut microbiota in the acute phase of intensive care unit patients. *Digestive diseases and sciences.* 2016; 61:1628-1634.
 18. Dunn OJ. Multiple comparisons using rank sums. *Technometrics.* 1964;6(3):241-252.
 19. Gower JC. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika.* 1966;53(3-4):325-338.
 20. Sokal RR. A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull.* 1958;38:1409-1438.
 21. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011;12(6):R60.
 22. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ.* 2015;3:e1029.
 23. Wickham H. Data analysis. In: *ggplot2: elegant graphics for data analysis.* 2nd ed. Springer; 2016:189-201.
 24. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77.
 25. NCAR - Research Applications Laboratory. verification: Weather forecast verification utilities. The Comprehensive R Archive Network. Available at: <https://cran.r-project.org/web/packages/verification>. Accessed May 18, 2022.