**SUPPLEMENTARY MATERIAL**

# The Intra-hepatic Metataxonomic Signature in Nonalcoholic Fatty Liver Disease

Silvia Sookoian [1,2*], Adrian Salatino [1,3], Gustavo O Castaño [4], Maria Silvia Landa [1,3], Cinthia Fijalkowky [1,3], Martin Garaycoechea [5,6], Carlos J Pirola [1,3,6*]

[1] University of Buenos Aires, School of Medicine, Institute of Medical Research A Lanari, Ciudad Autónoma de Buenos Aires, Argentina.

[2] National Scientific and Technical Research Council (CONICET)−University of Buenos Aires, Institute of Medical Research (IDIM), Department of Clinical and Molecular Hepatology, Ciudad Autónoma de Buenos Aires, Argentina.

[3] National Scientific and Technical Research Council (CONICET)−University of Buenos Aires, Institute of Medical Research (IDIM), Department of Molecular Genetics and Biology of Complex Diseases, Ciudad Autónoma de Buenos Aires, Argentina.

[4] Liver Unit, Medicine and Surgery Department, Hospital Abel Zubizarreta, Ciudad Autónoma de Buenos Aires, Argentina.

[5] Department of Surgery and [6] Centro de Medicina Traslacional, Hospital de Alta Complejidad en Red "El Cruce", Florencio Varela, Pcia. de Buenos Aires, Argentina.

[*] Co-corresponding authorship



Sookoian S, *et al. Gut* 2020; 69:1483–1491. doi: 10.1136/gutjnl-2019-318811

**LIST OF SUPPLEMENTARY MATERIAL**

**Supplementary Methods**

**Supplementary Results**

**Supplementary Discussion**

**Supplementary Tables**



**Supplementary Figures**





Sookoian S, *et al. Gut* 2020; 69:1483–1491. doi: 10.1136/gutjnl-2019-318811

- **Supplementary Figure 5: The impact of obesity on the liver tissue metataxonomic profile**. Linear discriminant analysis score on the x-axis represents log changes in relative bacterial DNA according to BMI categories (<40 kg/m$^2$ vs. >=40 kg/m$^2$). This graphical output was generated by the LEfSe visualization modules publicly available at the Galaxy website.
- **Supplementary Figure 6: Results of 16S rRNA gene quantitative PCR (qPCR). A:** Graph shows in silico predicted coverage summary of qPCR using universal primers for all taxonomic units in the SILVA database. B: Liver content of 16S rRNA gene in NAFLD- morbid obesity cohort versus NAFLD-non-morbid obesity cohort. C and D: Liver content of 16S rRNA gene in NAFL and NASH in NAFLD-morbid obesity cohort versus NAFLD-non-morbid obesity cohort, respectively. Liver content of 16S rRNA gene is expressed as normalized by the amount of the liver albumin (ALB) exon 12. Graph shows results expressed as mean ± SD values. *P* value stands for statistical significance, as indicated by the nonparametric Mann–Whitney U test.
- **Supplementary Figure 7: LDA (linear discriminant analysis) score in the control group vs. NASH**. A**.** LDA score on the x-axis represents log changes in relative bacterial DNA in healthy controls compared to NASH patients. B. Cladogram representing the LDA LEfSe results. These graphical outputs were generated by the LEfSe visualization modules publicly available at the Galaxy website.
- **Supplementary Figure 8: Validation of results by qPCR in the NAFLD-morbid obesity cohort.** A nonparametric Mann-Whitney U test was used for pairwise comparisons between groups. Bars represent mean ±SEM. Data were analysed using GraphPad Prism version 4.00 for Windows (GraphPad Software, San Diego, CA, USA). The results were considered statistically significant at $P < 0.05$.Target gene content in the liver tissue was normalized to the amount of the single copy gene encoding human albumin (ALB) exon 12. Number of patients included in the analysis: NAFL: 23, NASH: 27. Lobular inflammation (Lob inf) no =17/yes n =33. Portal inflammation (Portal infl) no = 25/ yes = 25. Ballooning degeneration no = 23/ yes n= 27. Fibrosis no n =33/ yes n= 17.
- **Supplementary Figure 9: Validation of results by qPCR in the NAFLD-non-morbid obesity cohort.** A nonparametric Mann-Whitney U test was used for pairwise comparisons between groups. Bars represent mean ± SEM. Data were analysed using GraphPad Prism version 4.00 for Windows (GraphPad Software, San Diego, CA, USA). The results were considered statistically significant at $P < 0.05$. Target gene content in the liver tissue was normalized to the amount of the single copy gene encoding human albumin (ALB) exon 12. Number of patients included in the analysis: NAFL: 21, NASH: 26. Lobular inflammation (Lob inf) no =15/yes n =31. Portal inflammation (Portal infl) no = 24/ yes = 23. Ballooning degeneration no = 28/ yes n= 19. Fibrosis no n =34/ yes n= 13.
- **Supplementary Figure 10: Necrotic colonic polyp was used as positive control tissue of LPS immunostaining by IHQ.** A, B, C: Necrotic colon cancer (necrotic fragment from the peripheral area of the tumor, replaced by fibrous tissue and areas of bacterial colonies). D: Negative control that was carried out with normal goat serum diluted to the same concentration as the primary antibody.



Sookoian S, *et al. Gut* 2020; 69:1483–1491. doi: 10.1136/gutjnl-2019-318811

- **Supplementary Figure 11: Predictive functional profiling of liver bacterial DNA in NAFLD patients.** Functional profiling was performed by PICRUSt from the 16S rRNA gene sequencing amplicon sequencing. Heatmap shows MetaCyc reference pathways that were used to compare relative abundance determined from both cohorts of patients. The OTU table was normalized by dividing each OTU by the known/predicted 16S copy number abundance. Thousands of predicted functions were collapsed into higher categories of MetaCyc pathways.
- **Supplementary Figure 12: Liver bacterial DNA profiling shows distinctive predicted functional signatures.** The top metabolic pathways (MetaCyc) for each phenotype (morbid obesity vs. non-morbidly obese patients with NAFLD) are represented with effect sizes, confidence intervals, and their associated *p* value. Bars indicate mean proportions (%). STAMP was used to determine differentially enriched metabolic pathways ($p < 0.05$) and their effect size ($\eta 2$). Bootstrap test was performed to assess statistical significance. Corrected *p* values were calculated using Storey's FDR approach.
- **Supplementary Figure 13: Principal coordinate (PC) analysis plot of the two cohorts examined by the Piphillin package.**
- **Supplementary Figure 14: Detection of significantly different functions by KEGG KO Piphillin**
- **Supplementary Figure 15: Results in Piphillin BioCyc reaction annotations**



Sookoian S, *et al. Gut* 2020; 69:1483–1491. doi: 10.1136/gutjnl-2019-318811

# SUPPLEMENTARY METHODS

## Patients and Methods

### Study design, selection of patients, and phenotypic characterization

Patients and controls were selected from two independent hospital-based settings, whereby individuals diagnosed with NAFLD and non-morbid obesity were recruited from Hospital Abel Zubizarreta, Ciudad Autónoma de Buenos Aires, Argentina, whereas morbidly obese patients that underwent BS were recruited from Hospital de Alta Complejidad en Red "El Cruce", Florencio Varela, Provincia de Buenos Aires, Argentina. Patients were considered for inclusion if they had histopathologic evidence of fatty liver disease, either simple steatosis (non-alcoholic fatty liver −NAFL) or nonalcoholic steatohepatitis (NASH), on liver biopsy performed within the study period.

Demographic and clinical data were collected and subjects were matched by age and sex. Liver tissue samples were collected at the time of liver biopsy.

The case participants and control subjects were consecutively selected during the same study period from the same population of patients attending the participating institutions located in Argentina, ensuring that all shared the same demographic characteristics (occupation, educational level, place of residence, and ethnicity) as the matched patients.

### Control liver specimens of individuals without NAFLD (control non-NAFLD group)

As control liver tissue (non-NAFLD) we included samples from 19 subjects, including 9 healthy patients without either NAFLD or features of the metabolic syndrome who were selected from patients attending the Liver Unit, whose age and sex matched the NAFLD patients. These patients presented near normal liver histology in the liver specimens obtained by percutaneous liver biopsy. The reason for performing a liver biopsy in these subjects was based on the



Sookoian S, *et al. Gut* 2020; 69:1483–1491. doi: 10.1136/gutjnl-2019-318811

presence of persistently mildly elevated serum liver enzymes activity. In all the near normal liver histology subjects, all causes of common liver disease were ruled out, and they were included in the study if they did not have histological evidence of fatty change; the histological diagnosis of control livers was minimal changes. In the population of morbid obese patients, control subjects ($n$ =10) were obese patients who also underwent bariatric surgery and had not features of NAFLD demonstrated in the liver biopsy.

Exclusion criteria: Secondary causes of steatosis, including alcohol abuse (≥ 30 g for men and ≥ 20 g for women, of alcohol daily), total parenteral nutrition, hepatitis B and hepatitis C virus infection, and the use of drugs known to precipitate steatosis were excluded. In addition, patients with any of the following diseases were excluded: autoimmune liver disease, metabolic liver disease, Wilson's disease, and α-1-antitrypsin deficiency. Patients under treatment with antibiotics, immunosuppressive medication or proton-pump bomb inhibitors were also excluded.

**Physical and anthropometric evaluation**

Health examinations included anthropometric measurements, a questionnaire on health-related behaviors, and biochemical determinations. Anthropometric measurements and blood samples were obtained from each patient at the time of liver biopsy or DNA sample collection, and before any intervention. The body mass index (BMI) was calculated as weight/squared height ($kg/m^2$) and used as an index for relative weight.

**Biochemical determinations**

Blood was drawn from 12-hour fasting subjects who had been in a supine resting position for at least 30 min. Laboratory evaluation included serum ALT and AST, gamma glutamyl transferase (γGT), alkaline phosphatase (AP), glucose and insulin, total cholesterol, high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterol, and plasma triglycerides (TG). All

biochemical determinations were measured using a Hitachi-912 Autoanalyzer (Roche, Diagnostic, Buenos Aires, Argentina) or Immulite 1000 (DPC, Buenos Aires, Argentina). Homeostasis Model Assessment (HOMA-IR) was used to evaluate an insulin resistance index and was calculated as follows: Fasting serum insulin (μU/ml) × Fasting plasma glucose (mmol/l) / 22.5. Anthropometric measurements and blood samples were obtained from each patient at the time of liver biopsy and before any intervention.

## Histological assessment of NAFLD

NAFLD disease severity was assessed by liver biopsy that was performed before any intervention with ultrasound guidance using a modified 1.4-mm-diameter Menghini needle (Hepafix, Braun, Germany) under local anesthesia on an outpatient basis, or during bariatric surgery (surgically excised samples from the left lobe were immediately collected after the abdomen was opened and before organs were manipulated).

A portion of each liver biopsy specimen was routinely fixed in 40 g/l formaldehyde (pH:7.4), embedded in paraffin, and stained with hematoxylin and eosin, Masson trichrome, and silver impregnation for reticular fibers. All biopsies were at least 3 cm in length and contained a minimum of eight portal tracts.

The degree of steatosis was assessed according to the system developed by Kleiner et al. [1], based on the percentage of hepatocytes containing macrovesicular fat droplets [1]. NASH and NAFLD Activity Score (NAS) [1;2] were defined as reported previously, and a NAS threshold of 5 was used for further comparisons with variables of interest; NASH was defined as steatosis plus mixed inflammatory-cell infiltration, hepatocyte ballooning and necrosis, Mallory's hyaline, and any stage of fibrosis, including absent fibrosis [1;2].

Intra-acinar (lobular) inflammation was defined as the presence of cellular components of inflammation (polymorphonuclear leukocytes, lymphocytes and other mononuclear cells, eosinophils and microgranulomas) located in sinusoidal spaces, surrounding Mallory's hyaline or in hepatocellular necrosis [1;2]. It was graded 0–3 and was defined as 0 (absent) = no foci; 1 = < 2 foci per 200 × field; 2 = 2 to 4 foci per 200 × field; and 3 = >4 foci per 200 × field. Ballooning was scored as: 0 = none; 1 = rare or few; and 2 = many (1). The severity of fibrosis was expressed on a 4-point scale, as follows: 0 = none; 1 = perivenular and/or perisinusoidal fibrosis in zone 3; 2 = combined pericellular portal fibrosis; 3 = septal/ bridging fibrosis; and 4 = cirrhosis.

**Immunohistochemistry (IHQ) for LPS (Lipopolysaccharide)**

Four-micrometer sections were mounted onto silane coated glass slides to ensure section adhesion through subsequent staining procedures. Briefly, sections were deparaffinized, rehydrated, washed in phosphate buffer solution (PBS), and treated with 3 % $H_2O_2$ in PBS for 20 min at room temperature to block endogenous peroxidase. Following heat-induced epitope retrieval in 0.1M citrate buffer at pH 6.0 for 20 min, the slides were incubated with a dilution of 1:200 of a goat polyclonal antibody for Lipid A LPS Antibody Antibody (NB100-64484), Novus Biologicals CO, USA). NB100-64484 recognizes the Lipid A or endotoxin region of LPS (Lipopolysaccharide) from E. coli. Lipid A is a glucosamine-based phospholipid which forms part of the bacterial outer membrane and anchors LPS to the surface of the bacteria. This antibody crosses react with P. aeruginosa, K. pneumoniae, S. enteriditis, E. aerogenes, E. hermanii, Y. enterocolitica and S. sonnei.



Sookoian S, *et al. Gut* 2020; 69:1483–1491. doi: 10.1136/gutjnl-2019-318811

Immunostaining was performed using the VECTASTAIN® Universal Quick HRP Kit (Peroxidase), R.T.U. (Ready-to-Use) (Vector Lab. CA, USA) detection system (PK-7800). Subsequently, slides were immersed in a 0.05% 3,3'-diaminobenzidine solution in 0.1 M Tris buffer, pH 7.2, containing 0.01 % $H_2O_2$. After a brown color developed, slides were removed and the reaction was stopped by immersion in PBS. Negative controls were carried out with goat serum diluted to the same concentration as the primary antibody. LPS immunostaining was evaluated in a blinded fashion regarding any of the histological and clinical characteristics of the patients. The specificity of immunostaining was demonstrated by omission of primary Ab. Sections were counter-stained with Harris hematoxylin and examined by light microscopy in a blinded fashion (H-3401-500. Vector Lab. CA, USA).

The extent of staining was scored according to its amount and intensity by a 4-point scoring system as follows: 0 = no staining, 1 = positive staining in less than 20% of cells and /or tissue area of the portal tracts, 2 = 21-50 % of cells and /or tissue area of the portal tracts, and 3 = positive staining in more than 50% of cells and /or tissue area of the portal tracts. Histological specimens were assessed by a LEICA DM 2000 (Leica, Germany) trinocular microscope equipped with a high-definition camera (Leica MC190 HD*)*; all images were recorded using the Leica Application Suite (LAS) software.

## Liver DNA extraction method

Total DNA was extracted from human liver specimens by a manual protocol modified from Kawasaki et al. [3] after two mechanical lyses steps. Our DNA extraction method remained constant throughout the whole process and the extraction protocol was employed consistently in all samples by using systematically same reagents across the study project.

Liver samples collected from patients were first manually homogenized with a tissue micro pestle (Sigma Aldrich Inc. MI, USA). In order to break the cell membranes of some species and to obtain intact genomic DNA, samples were immediately subjected to an extra cell lysis step that involved bead beating tissue homogenization by Bioprep-6 Homogeneuzer (Hangzhou Allsheng Instruments CO, LTD, Hangzhou, China); a protocol of 3 cycles of 20 second shaking of ceramic beads (1.4 mm) with 1-min intervals on ice was used to physically disrupt cells walls/membranes, which ensures releasing DNA while still retaining molecular integrity. The DNA concentration (ng/ul) extracted from liver tissue was 85.29± 69.25 ng/ul (mean±SD). The OD260/280 ratio was used as an indicator of DNA quality; samples with values in the 1.8–2.0 range were sequenced. As a part of the initial quality control, we determined the size, purity, and concentration of the DNA sample (Agilent Technologies 2100 Bioanalyzer). The quality and quantity of DNA were assessed by NanoDrop 2000 UV spectrophotometer (Thermo Scientific, Waltham, MA, USA) and by agarose gel electrophoresis. DNA samples were immediately frozen and stored at −70°C before analysis.

## 16S rRNA gene metagenomic sequencing, sequence data analysis, and operational taxonomy unit (OTU) clustering

16S rRNA gene sequencing data were filtered, denoised, and processed on the QIIME2 (version 2018.11) platform (http://qiime2.org/index.html). High-quality sequences were assigned to operational taxonomic units (OTUs) using the QIIME pipeline [4]; default parameters were used in the selection of OTUs that would be used in constructing the OTU table. After importing and demultiplexing, sequences were denoised using the 16S deblur algorithm trimming at 210 bp. A total of 1,870,504 high-quality reads were retained.

High-quality amplicon sequence variants were classified using vsearch algorithm with default parameters and SILVA 16S-only 99% identity database (release 137) to build our BIOM feature table of operational taxonomic units (OTUs). Sequence data was defined at different levels of resolution (kingdom, phylum, class, order, family, genus, and species. Reads corresponding to Chloroplasts were filtered out.

## Taxonomic analysis

Liver metataxonomic data analysis, including compositional profiling, statistical and differential analysis, visual exploration, and data integration, was performed by the MicrobiomeAnalyst web server (https://www.microbiomeanalyst.ca) [5] and the web-application Calypso (http://cgenome.net/calypso/) [6]. Data filtering procedures, including filtering of low abundance features (<0.001%) and low variance features putatively caused by sequencing errors were systematical applied to all analysis.

Data normalization was performed to account for uneven sequencing depths across samples. Specifically, we used Total Sum Scaling (TSS) normalization, involving taking the square root of the ratio of the OTU count and the total number of counts in each individual sample, resulting in 4,622 of 16,490 data points per sample. The coverage of the original microbial communities by metagenomic sequence data was estimated by rarefaction analysis. Microbial sequences were randomly drawn from each sample. For each subsample, the number of observed species was counted and plotted as a function of the number of sampled sequences. Consequently, the slope of the rarefaction curve signifies whether the underlying microbial DNA is well represented by the sequence data.

As an estimator of microbial DNA diversity we used the alpha-diversity Shannon index (H) [7]. Significance in taxa comparisons of alpha diversity was determined by nonparametric one-way ANOVA (Wilcoxon rank test). Beta diversity was calculated for every pair of samples to generate a distance or dissimilarity matrix via principal coordinate analysis (PCoA) using weighted UniFrac distances and Bray-Curtis index; PERMDISP2 procedure was used for the analysis of multivariate homogeneity of group dispersions (variances). PCoA was conducted for visual representation at low-dimensional space (2D plots). The Kruskal-Wallis test was also performed to compare abundance across the three groups, whereas the Mann-Whitney test was used for pairwise comparisons. A false discovery rate (FDR) of 0.05 was adopted as an indicator of statistical significance.

**16S rRNA gene quantitative PCR (qPCR)**

The DNA extracted from the liver specimens was evaluated by qPCR using universal primers based on conserved sequences as described earlier by Fuller et al. [8]. 16S rRNA gene qPCR data were used to determine total bacterial amount within each sample; measurements were done in triplicate. A no template negative control (sterile nuclease free water) was added to all qPCR runs. The Applied Biosystems Step One Plus Real-Time PCR System was used for amplification and real-time fluorescence detections. The cycling conditions consisted of 5 min of initial denaturation at 95 °C, followed by 35 cycles consisting of 30-s denaturation at 95 °C, 30-s annealing at the 55 °C annealing temperature, and a 1 min extension period at 72 °C with a final period of 10-min extension. Applied Biosystems SYBR Green PCR Master Mix (Thermo Fisher Scientific, Inc, Waltham, Massachusetts, USA) was used to amplify the target amplicons.



Sookoian S, *et al. Gut* 2020; 69:1483–1491. doi: 10.1136/gutjnl-2019-318811

The specificity of amplification and absence of primer dimers were confirmed using the melting curve analysis at the end of each run. The amplification products were visualized after electrophoresis on a 3% agarose gel.

The 16S rRNA gene content in the liver tissue was normalized to the amount of the single copy gene encoding human albumin (ALB) exon 12 that was included in each qPCR series. The liver 16S rRNA amplicon levels were expressed relative to the levels of the albumin exon 12 amplicon using fluorescence threshold cycle values (Ct) calculated for each sample using the delta Ct method (16S rRNA/ALB ratio) using the PCR efficiency calculated for each reaction as follows: raw data were exported from the StepOne PCR system and imported into the LinRegPCR program. The estimated efficiency of the PCR for each product was expressed as the average of all sample efficiency values obtained [9]. The primers used for the quantitative PCR assay are listed below:

| **Primer name** | **Primer sequence** | **Amplicon size** | **Annealing Temperature** |
|---|---|---|---|
| Universal 16S Forward | GTGSTGCAYGGYYGTCGTCA | 147 bp | 55 °C |
| Universal 16S Reverse | ACGTCRTCCMCNCCTTCCTC | | |
| Human albumin exon 12 Forward | TGTTGCATGAGAAAACGCCA | 72 bp | 55 °C |
| Human albumin exon 12 Reverse | GTCGCCTGTTCAACCAAGGAT | | |

The performance of the universal primer pairs was assessed by running an in silico PCR on the SILVA database with the TestPrime resource (https://www.arb-silva.de/search/testprime/). TestPrime computes coverage for each taxonomic group in all of the taxonomies offered by SILVA.

**Validation of main results by real time qPCR**

Validation of main findings of bacterial DNA composition that were associated with the disease severity was performed by real time qPCR method, and target gene content in the liver tissue was normalized to the amount of the single copy gene encoding human albumin (ALB) exon 12 that was included in each qPCR series as above described. Measurements were done in duplicate. Negative controls were added to each plate with template DNA being replaced with PCR-grade water. RT-qPCR was achieved with a STEP one PLUS instrument and software (Applied Biosystems, Foster City, CA, USA) using SYBR Green PCR MasterMix (Applied Biosystems, Foster City, CA, USA).

Primers were obtained from previous publications or their design was performed by Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/); specificity was checked in silico on the SILVA databases by the TestPrime resource (https://www.arb-silva.de/search/testprime/) and http://insilico.ehu.es/PCR/.

The primers used for the validation quantitative PCR assays are listed below:

| Primer name | Primer sequence | Amplicon size | Annealing Temperature | Reference |
|---|---|---|---|---|
| Lachnospiraceae F | AAACAGCTTAGTGGCGGACG | 203 | 60°C | This study |
| Lachnospiraceae R | GGCTACTGATCGTCGCTTTG | | | |
| Actinobacteria F | TACGGCCGCAAGGCTA | 300 | 65°C | [10] |
| Actinobacteria R | TCRTCCCCACCTTCCTCCG | | | |
| Alphaproteobacteria F | CGGTAATACGRAGGGRGYT | 142 | 60°C | [11] |
| Alphaproteobacteria R | CBAATATCTACGAATTYCACCT | | | |
| Gammaproteobacteria F | CMATGCCGCGTGTGTGAA | 350 | 65°C | [12] |
| Gammaproteobacteria R | ACTCCCCAGGCGGTCDACTTA | | | |
| Bacteroides/ Prevotella F | AAGGTCCCCCACATTG | 104 | 60°C | [13] |
| Bacteroides/ Prevotella R | CACGCTACTTGGCTGGTTCAG | | | |
| Veillonellaa F | AYCAACCTGCCCTTCAGA | 343 | 55°C | [14] |
| Veillonellaa R | CGTCCCGATTAACAGAGCTT | | | |
| Peptostreptococcus-F | TCGAGCGCGTCTGATTTGAT | 71 | 58°C | This study |
| Peptostreptococcus-R | TTACCCACGCGTTACTCACC | | | |
| Verrucomicrobia-F | GAATTCTCGGTGTAGCA | 551 | 55°C | [15] |
| Verrucomicrobia-R | GGCATTGTAGTACGTGTGCA | | | |

## The workflow for 16S rRNA gene amplicon sequencing

The Illumina 16S rRNA amplicon sequencing workflow included 4 steps:

1) Sample Preparation: For library construction, DNA was extracted from each sample. After performing quality control (QC), qualified samples proceed to library construction.

2) Library Construction: The sequencing library was prepared by random fragmentation of the DNA sample, followed by 5' and 3' adapter ligation. Adapter-ligated fragments are then PCR amplified and gel purified.

3) Sequencing: For cluster generation, the library was loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment was then amplified into distinct, clonal clusters through bridge amplification. When cluster generation was complete, the templates were ready for sequencing. Illumina SBS technology utilizes a proprietary reversible terminator-based method that detects single bases as they are incorporated into DNA template strands. As all 4 reversible, terminator-bound dNTPs are present during each sequencing cycle, natural competition minimizes incorporation bias and greatly reduces raw error rates compared to other technologies. The result is highly accurate base-by-base sequencing that virtually eliminates sequence-context-specific errors, even within repetitive sequence regions and homopolymers.

4) Raw data: Sequencing data was converted into raw data for the analysis.

**Library construction:**



Sookoian S, *et al. Gut* 2020; 69:1483–1491. doi: 10.1136/gutjnl-2019-318811

The method for preparing samples for sequencing the variable V3 and V4 regions of the 16S rRNA gene is described below. The Herculase II Fusion DNA Plymerase Nextera XT Index Kit V2 was used.

**1.** The protocol includes the primer pair sequences for the V3 and V4 region that create a single amplicon of approximately ~460 bp. The protocol also includes overhang adapter sequences that were appended to the primer pair sequences for compatibility with Illumina index and sequencing adapters. See Amplicon Primers: Bakt_341Forward: CCTACGGGNGGCWGCAG and Bakt_805Reverse: GACTACHVGGGTATCTAATCC [16].

**2.** Library preparation: using a limited cycle PCR, Illumina sequencing adapters and dual-index barcodes were added to the amplicon target. Using the full complement of Nextera XT indices, up to 96 libraries can be pooled together for sequencing.

**3.** Sequence on MiSeq: using paired 300-bp reads, and MiSeq v3 reagents, the ends of each read are overlapped to generate high-quality, full-length reads of the V3 and V4 region in a single 65-hour run. The MiSeq run output is approximately > 2 million reads, commonly recognized as sufficient for metagenomic surveys. Illumina adapter overhang nucleotide sequences are added to the gene-specific sequences

**Amplicon PCR**

| **Item** | **Quantity** |
|---|---|
| Microbial Genomic DNA (5 ng/µl in 10 mM Tris pH 8.5) | 2.5 µl per sample |
| Amplicon PCR Reverse Primer (1 µM) | 5 µl per sample |
| Amplicon PCR Forward Primer (1 µM) | 5 µl per sample |
| 2x KAPA HiFi HotStart ReadyMix | 12.5 µl per sample |

PCR thermal cycler program:

• 95°C for 3 minutes

• 25 cycles of:

— 95°C for 30 seconds

— 55°C for 30 seconds

— 72°C for 30 seconds

• 72°C for 5 minutes

• Hold at 4°C

1 µl of the PCR product was run on a Bioanalyzer DNA 1000 chip to verify the size.

**PCR Clean-Up:** At this step we used AMPure XP beads to purify the 16S V3 and V4 amplicon away from free primers and primer dimer species.

| **Item** | **Quantity** |
|---|---|
| 10 mM Tris pH 8.5 | 52.5 µl per sample |
| AMPure XP beads | 20 µl per sample |
| Freshly Prepared 80% Ethanol (EtOH) | 400 µl per sample |
| 96-well 0.2 ml PCR plate | 1 plate |

Index PCR: This step attaches dual indices and Illumina sequencing adapters using the Nextera XT Index Kit. The following reaction of DNA was performed as follows:

| **Item** | **Volume** |
|---|---|
| DNA | 5 µl |
| Nextera XT Index Primer 1 (N7xx) | 5 µl |
| Nextera XT Index Primer 2 (S5xx) | 5 µl |
| 2x KAPA HiFi HotStart ReadyMix | 25 µl |

| | |
|---|---|
| PCR Grade water | 10 µl |

PCR was performed on a thermal cycler using the following program:

• 95°C for 3 minutes

• 8 cycles of:

— 95°C for 30 seconds

— 55°C for 30 seconds

— 72°C for 30 seconds

• 72°C for 5 minutes

• Hold at 4°C

**PCR Clean-Up 2:** This step uses AMPure XP beads to clean up the final library before quantification

| **Item** | **Quantity** |
|---|---|
| 10 mM Tris pH 8.5 | 27.5 µl per sample |
| AMPure XP beads | 56 µl per sample |
| Freshly Prepared 80% Ethanol (EtOH) | 400 µl per sample |
| 96-well 0.2 ml PCR plate | 1 plate |

**Library validation:** 1 µl of a 1:50 dilution of the final library was run on a Bioanalyzer DNA 1000 chip to verify the size. Using the V3 and V4 primer pairs in the protocol, the expected size on a Bioanalyzer trace of the final library is ~630 bp.

**Library Quantification, Normalization, and Pooling:** Libraries were quantified using a fluorometric quantification method that uses dsDNA binding dyes. DNA concentration was

calculated in nM, based on the size of DNA amplicons as determined by an Agilent Technologies 2100 Bioanalyzer trace.

**Library Denaturing and MiSeq Sample Loading:** In preparation for cluster generation and sequencing, pooled libraries were denatured with NaOH, diluted with hybridization buffer, and then heat denatured before MiSeq sequencing. Each run included a minimum of 5% PhiX to serve as an internal control for these low diversity libraries.

| Item | Quantity |
|---|---|
| 10 mM Tris pH 8.5 or RSB (Resuspension Buffer) | 6 µl |
| HT1 (Hybridization Buffer) | 1540 µl |
| 0.2 N NaOH (less than a week old) | 10 µl |
| PhiX Control Kit v3 (FC-110-3001) | 4 µl |
| MiSeq reagent cartridge | 1 cartridge |

Denature DNA

1. The following volumes of pooled final DNA library and freshly diluted 0.2 N NaOH were combined in a microcentrifuge tube:

• 4 nM pooled library (5 µl)

• 0.2 N NaOH (5 µl)

2. The remaining dilution of 0.2 N NaOH was set aside to prepare a PhiX control within the next 12 hours.

3. The sample solution was vortexed briefly and then centrifuged the sample solution at 280 × g at 20°C for 1 minute.

4. Incubated for 5 minutes at room temperature to denature the DNA into single strands.

5. the following volume of pre-chilled HT1 was added to the tube containing denatured DNA:

• Denatured DNA (10 µl). Pre-chilled HT1 (990 µl). Adding the HT1 results in a 20 pM denatured library in 1 mM NaOH.

**6** denatured DNA was placed on ice until you are ready to proceed to final dilution.

**Denature and Dilution of PhiX Control**

The following instructions were followed to denature and dilute the 10 nM PhiX library to the same loading concentration as the Amplicon library. The final library mixture must contain at least 5% PhiX.

**1** The following volumes were combined to dilute the PhiX library to 4 nM:

• 10 nM PhiX library (2 µl)

• 10 mM Tris pH 8.5 (3 µl)

**2** The following volumes were combined of 4 nM PhiX and 0.2 N NaOH in a microcentrifuge tube:

• 4 nM PhiX library (5 µl)

• 0.2 N NaOH (5 µl)

**3** Vortex briefly to mix the 2 nM PhiX library solution.

**4** Incubated for 5 minutes at room temperature to denature the PhiX library into single strands.

**5** the following volumes of pre-chilled HT1 were added to the tube containing denatured PhiX library to result in a 20 pM PhiX library:

• Denatured PhiX library (10 µl)

• Pre-chilled HT1 (990 µl)

**6** the denatured 20 pM PhiX library was diluted to the same loading concentration as the Amplicon library as follows:

| Final Concentration | 2 pM | 4 pM | 6 pM | 8 pM | 10 pM |
|---|---|---|---|---|---|
| 20 pM denatured library | 60 µl | 120 µl | 180 µl | 240 µl | 300 µl |
| Pre-chilled HT1 | 540 µl | 480 µl | 420 µl | 360 µl | 300 µl |

**Amplicon Library and PhiX Control**

The following volumes of denatured PhiX control library were combined to the denatured amplicon library in a microcentrifuge tube:

• Denatured and diluted PhiX control (30 µl)

• Denatured and diluted amplicon library (570 µl)

**3** Using a heat block, the combined library and PhiX control tube were incubated at 96°C for 2 minutes.

**Dual Indexing Principle**

The dual indexing strategy uses two 8 base indices, Index 1 (i7) adjacent to the P7 sequence, and Index 2 (i5) adjacent to the P5 sequence. Dual indexing is enabled by adding a unique Index 1 (i7) and Index 2 (i5) to each sample. The 96 sample Nextera XT Index Kit (FC-131–1002) use 12 different Index 1 (i7) adapters (N701–N712) and 8 different Index 2 (i5) adapters (S501–S508). The 24 sample Nextera XT Index Kit (FC-131–1001) uses 6 different Index 1 (i7) adapters (N701–N706) and 4 different Index 2 (i5) adapters (S501–S504). In the Index adapter name, the N or S refers to Nextera XT sample preparation, 7 or 5 refers to Index 1 (i7) or Index 2 (i5), respectively. The 01–12 refers to the Index number.

A list of index sequences is provided for generating sample sheets to demultiplex the samples:

| Index 1 (i7) | Sequence | Index 2 (i5) | Sequence |
|---|---|---|---|
| N701 | TAAGGCGA | S501 | TAGATCGC |
| N702 | CGTACTAG | S502 | CTCTCTAT |
| N703 | AGGCAGAA | S503 | TATCCTCT |
| N704 | TCCTGAGC | S504 | AGAGTAGA |
| N705 | GGACTCCT | S505 | GTAAGGAG |
| N706 | TAGGCATG | S506 | ACTGCATA |
| N707 | CTCTCTAC | S507 | AAGGAGTA |
| N708 | CAGAGAGG | S508 | CTAAGCCT |
| N709 | GCTACGCT | | |
| N710 | CGAGGCTG | | |
| N711 | AAGAGGCA | | |
| N712 | GTAGAGGA | | |

## Linear Discriminant Analysis (LDA) Effect Size (LEfSe)

LEfSe analysis was performed on the OTU table using the online Galaxy web application (http://huttenhower.sph.harvard.edu/galaxy/) to identify features (taxa) that were differentially abundant across biological conditions of interest. To select the most predictive/discriminative taxa classifying the samples according to the two patient cohorts, Sparse Partial Least Squares Linear Discriminant Analysis (sPLS-DA) was conducted using the Mixomics software package. The non-parametric factorial Kruskal-Wallis (KW) sum-rank test was used to detect features with significant differential abundance with respect to the class of interest; biological significance was subsequently investigated using a set of pairwise tests among subclasses based on the (unpaired) Wilcoxon rank-sum test. Linear Discriminant Analysis (LDA) was performed to estimate the effect size of each differentially abundant feature. Graphical reports were generated, including LDA bar plots (differential features ranked by effect size) and plot cladograms (representation of relevant features on taxonomic or phylogenetic trees).

## Functional prediction

To predict metagenome functional content from the OTU table, we utilized the bioinformatic software package PICRUSt2: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (https://github.com/picrust/picrust2 /). Functional inference was additionally explored by Piphillin metagenomics inference tool available at http://piphillin.secondgenome.com/. The normalized sequence abundance table and the weighted nearest-sequenced taxon index values per-sample were used to predict pathways. The resulting output was a list of MetaCyc pathway abundances according to the MetaCyc database (https://metacyc.org/). The STAMP software package (version 2.1.3), which relies on the concept of biological relevance in the form of confidence intervals, was employed to determine differentially enriched metabolic pathways ($p < 0.05$) and their effect size ($\eta^2$) (http://kiwi.cs.dal.ca/Software/STAMP) [17].

**Statistical analysis for clinical data**

For demographic, clinical and histological features of case and control populations, we used the Mann-Whitney test to identify statistically significant differences in continuous variables, and the $\chi 2$ or Fisher's exact test to identify differences in categorical variables. Quantitative data are expressed as mean ± standard deviation (SD) unless otherwise indicated. Ordinal variables were dichotomized as ballooning degeneration, lobular inflammation, and liver fibrosis, whether they are present or not as yes (1) or no (0). The CSS/Statistica program package version 6.0 (StatSoft, Tulsa, OK, USA) was used in these analyses.

## SUPPLEMENTARY RESULTS

### *Liver microbial DNA diversity is related to the host phenome*

We further tested microbial alpha diversity (Shannon's diversity index) of taxonomic profiles for associations with collected metadata (sex: $p = 0.45$, type 2 diabetes: $p = 0.3135$, and arterial hypertension: $p = 0.46$), and no significant differences were found (nonparametric ANOVA one-way test statistics). Likewise, analysis focused on beta diversities within the collected metadata indicated that neither sex (PERMDIS2 $p = 0.59$), type 2 diabetes (PERMDIS2 $p = 0.94016$) nor arterial hypertension (PERMDIS2 $p = 0.76$) had any effect on the composition of the liver microbial DNA.

### Association between the liver tissue metataxonomic profiling and BMI as continuous variable

We further assessed the association between the liver tissue metataxonomic profiling and BMI, which was treated as continuous variable. As expected, similar results were found confirming that the abundance of bacterial DNA of Bacteroidetes and Firmicutes was significantly and positively correlated with BMI (**Supplementary Table 4**). Conversely, the abundance of Alphaproteobacteria and Gammaproteobacteria significantly decrease with increasing BMI (**Supplementary Table 4**). LDA (linear discriminant analysis) score on the x-axis representing log changes in relative bacterial DNA in patients having a BMI <40 kg/m$^2$ compared to patients having a BMI ≥40 kg/m$^2$ is shown in **Supplementary Figure 5.**

### Quantitative detection of total bacterial DNA

**Supplementary Figure 6 A** shows the *in silico* predicted coverage of the universal primers used to amplify the 16S rRNA gene, which is higher than 86% of the all taxonomic units in SILVA database. Then, quantitative PCR detection of total bacterial DNA in the whole sample of NAFLD patients (NAFLD-morbidly obese and NAFLD-non morbidly obese cohort) showed that the content of 16S rRNA gene in the liver was ~1.8 times the amount of albumin (ALB), a diploid single copy gene that is highly expressed in the liver. Relative content of liver 16S rRNA gene negatively correlated with BMI (R Spearman -0.25 $p = 0.01$) and positively correlated with alkaline phosphatase levels (R Spearman 0.30, $p = 0.004$). A significant increase in the content of total bacterial DNA (16S rRNA gene/ALB) was found in the liver of overweight or moderately obese patients with NAFLD compared with morbidly obese NAFLD patients (**Supplementary Figure 6 B**). Total 16S rRNA gene load did not differ between NAFL and NASH patients in none of the cohorts; however, the liver of NASH patients presented a modest but still not significant increase in the abundance of 16S rRNA gene (**Supplementary Figure 6 C, D**).

**Validation of main findings by an independent molecular approach**

Relative abundances of 16S rRNA genes of Proteobacteria (Alpha and Gamma), Actinobacteria, Peptostreptococcus, Verrucomicrobia, Veillonella, Lachnospiraceae, and Bacteroides/Prevotella were investigated, and results were normalized to the amount of the single copy gene encoding human albumin (ALB) exon 12. The specificities and coverage's of primer pairs for the target bacterial DNAs were checked by running an *in silico* PCR on the SILVA database (https://www.arb-silva.de/search/testprime). No pre-amplification was needed. However, the performance of RT-qPCR was not homogeneous; there were bacteria taxa that presented low

copy number, for example *Peptostreptococcus* and *Bacteroides/Alloprevotella*. Although we were able to validate a large proportion of results from NGS-metataxonomic profiling, design and validation of genus specific primers to use them as molecular probes for qPCR based detection have been particularly challenging. For example, primers set used for detecting DNA from *Bacteroides/Prevotella* presented high specificity (98.5%) but relatively low (57%) coverage for *Alloprevotella*; coverage and specificity for detecting *Prevotella 9* (89.6% and 98.8%, respectively) were much greater.

On balance, we found that liver 16S rRNA gene copy number of the eight selected taxa for replication showed significant differences between groups of patients with diverse histological traits, including the disease severity

***Liver bacterial DNA profiling shows distinctive predicted functional signatures***

Functional profiling prediction suggested that the two NAFLD cohorts, i.e., morbid obesity and non-morbid obesity, not only differed with respect to the microbial DNA composition but also functional richness, as shown in the heat map predicted by PICRUSt from the 16S rRNA gene amplicon sequencing (**Supplementary Figure 11**). Analysis of predicted functional modules indicated that microbial-derived pathways involved in the heme biosynthesis, oleate biosynthesis, ubiquinol-7 biosynthesis (characteristics of Gram-negative facultative anaerobes), fatty acid biosynthesis initiation and beta oxidation, stearate and mycolate biosynthesis, 5Z-dodec-5-enoate biosynthesis (a feature of Gammaproteobacteria for the production of unsaturated fatty acids anaerobically), L-leucin degradation (leucine may serve as a source of nitrogen for synthesis of L-glutamine and for gluconeogenesis in liver and kidneys), glyoxylate bypass (that allow bacteria to grow on acetate) and TCA cycle VII (acetate producers) were

over-enriched among patients with non-morbid obesity (**Supplementary Figure 12**). On the other hand, pathways with the highest statistical significance in the bariatric surgery cohort were related to glycogen degradation, *N*-acetylneuraminate catabolism (the major end-products of which are acetate, formate, and ethanol), L-isoleucine biosynthesis IV, pyruvate fermentation to acetone, pyruvate fermentation to acetate and lactate, pyruvate fermentation to butanoate, and incomplete reductive TCA cycle (**Supplementary Figure 12**). Consistently, we observed that several bacterial pathways that contribute to the acetate/acetone biosynthesis were significantly overrepresented in both NAFLD cohorts. However, specialized pathways were observed in each particular set of patients, which were modulated by the host phenome-related factors (severe obesity versus non-severe obesity).

These results were further validated by the Piphillin resource (**Supplementary Figure 13**).

By using KEGG (Kyoto Encyclopedia of Genes and Genomes) KO (KEGG Orthology) annotation (**Supplementary Figure 14**) and BioCyc reactions (**Supplementary Figure 15**) we confirmed differential predicted metabolic pathways in the liver bacterial DNA of the two cohorts. While fatty acid metabolism, beta alanine and porfirin metabolism, and microbial metabolism in diverse environment were over-enriched in NAFLD patients presenting non-morbid obesity (**Supplementary Figure 14**), biosynthesis of amino acids, pentose and glucoronate interconversions, pyrimidine metabolism, and amino sugar and nucleotide sugar metabolism were over-enriched in the NAFLD-morbid obese cohort (**Supplementary Figure 14).** Both cohorts showed predicted over-enrichment in the KEGG pathway of biosynthesis of secondary metabolites.

## SUPPLEMENTARY DISCUSSION

Assessment of total bacterial 16S rRNA gene copy number showed significant differences between morbidly obese and non-morbidly obese patients with NAFLD. Compared with patients of the morbid obesity-cohort, the liver of patients with non-morbid obesity had a modest ~1.47-fold increase in the relative 16S rRNA gene content, which could be explained by differences in bacterial DNA' composition between cohorts and/or pathophysiological changes in the host. Analysis of predicted functional modules also showed possible differences in predicted bacterial functions of morbidly obese relative to non-morbidly obese patients with NAFLD. Among NAFLD-non-morbidly obese patients, we found a greater proportion of reads with MetaCyc annotation in the initiation, beta oxidation, and biosynthesis of free fatty acid-related pathways—which can affect tissue physiology as well as a systemic inflammatory state [18;19]—and pathways involved in acetate production. Conversely, bacterial DNA profiles observed in the liver of morbidly obese patients with NAFLD presented an enrichment of pathways related with glycogen degradation and fermentation processes, including fermentation of pyruvate to acetone, lactate, butanoate or even ethanol. In any case, putative enrichment of pathways in both cohorts seems to be consistent with acetate production, in which Gram-negative bacteria are most likely involved. Increased acetate production was suggested as a potential driver of *de novo* hepatic lipogenesis [20]. The potential functions found in the liver likely reflect functions that possibly are selected elsewhere in the gastrointestinal tract.

It is known that metataxonomic profiles are modified by environmental factors, including dietary patterns and habits. While our results might not be necessarily comparable with data potentially emerging from other cohorts around the world, the interrelationship among disease features, tissue microbial DNAs and host factors—particularly fat mass, provide a plausible frame to understand NAFLD biology. In our study we recruited Argentinean subjects; thus, ethnicity

could be a covariate with dietary patterns. As individuals recruited in this study were from a relatively homogeneous ethnic group, the validity of our findings should be assessed in other cohorts and ethnic groups.

Finally, we provided evidence of potential mechanisms involved in the state of chronic inflammation associated with NAFLD in non-morbidly obese patients, for example chronic colonization of the liver by bacterial DNA originating in the mouth or the stomach and/ or the potential antigen generated by them. Notably, recent evidence links several chronic systemic diseases, including Alzheimer's disease [21] and cardiovascular disease [22], among many others [23], with the oral microbiota.

The hepatic microbial DNA profile reflects diverse bacterial DNA that can be found across the entire gastrointestinal tract, from the oral cavity to distal colon, which is aligned with the concept that the liver modulates the host-microbial interactions [24]. However, this profile does not necessarily match the bacterial DNA communities of a specific body organ / tissue. Accordingly, seminal studies have demonstrated that the biogeography of bacterial communities on the human body varies systematically across body habitats [25]. More importantly, the liver metataxonomic signature exhibits a unique bacterial DNA composition, including the Acidobacteria and Actinobacteria phylum (**Figure 1D**, **Supplementary Figure 8 and 9**). Members of these two bacterial Phyla have been previously reported in the liver of rodents [26].

## References

(1) Kleiner DE, Brunt EM, Van NM *et al*. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;**41**(6):1313-21.

(2) Brunt EM, Kleiner DE, Wilson LA *et al*. Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. *Hepatology* 2011;**53**(3):810-20.

(3) Kawasaki ES. Sample preparation from blood, cells, and other fluids. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR Protocols. A guide to Methods and Applications. San diego: Academic Press, INC., 1990: 146-52.

(4) Bokulich NA, Kaehler BD, Rideout JR *et al*. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 2018;**6**(1):90.

(5) Dhariwal A, Chong J, Habib S *et al*. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res* 2017;**45**(W1):W180-W188.

(6) Zakrzewski M, Proietti C, Ellis JJ *et al*. Calypso: a user-friendly web-server for mining and visualizing microbiome-environment interactions. *Bioinformatics* 2017;**33**(5):782-3.

(7) Kim BR, Shin J, Guevarra R *et al*. Deciphering Diversity Indices for a Better Understanding of Microbial Communities. *J Microbiol Biotechnol* 2017;**27**(12):2089-93.

(8) Fuller Z, Louis P, Mihajlovski A *et al*. Influence of cabbage processing methods and prebiotic manipulation of colonic microflora on glucosinolate breakdown in man. *Br J Nutr* 2007;**98**(2):364-72.

(9) Ruijter JM, Ramakers C, Hoogaars WM *et al*. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res* 2009;**37**(6):e45.

(10) Bacchetti De GT, Aldred N, Clare AS *et al*. Improvement of phylum- and class-specific primers for real-time PCR quantification of bacterial taxa. *J Microbiol Methods* 2011;**86**(3):351-6.

(11) Pfeiffer S, Pastar M, Mitter B *et al*. Improved group-specific primers based on the full SILVA 16S rRNA gene reference database. *Environ Microbiol* 2014;**16**(8):2389-407.

(12) Muhling M, Woolven-Allen J, Murrell JC *et al*. Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J* 2008;**2**(4):379-92.

(13) Furet JP, Firmesse O, Gourmelon M *et al*. Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. *FEMS Microbiol Ecol* 2009;**68**(3):351-62.

(14) Lyra A, Forssten S, Rolny P *et al*. Comparison of bacterial quantities in left and right colon biopsies and faeces. *World J Gastroenterol* 2012;**18**(32):4404-11.

(15) Ranjan K, Paula FS, Mueller RC *et al*. Forest-to-pasture conversion increases the diversity of the phylum Verrucomicrobia in Amazon rainforest soils. *Front Microbiol* 2015;**6**:779.

(16) Klindworth A, Pruesse E, Schweer T *et al*. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 2013;**41**(1):e1.

(17) Parks DH, Tyson GW, Hugenholtz P *et al*. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 2014;**30**(21):3123-4.

(18) Cani PD, Bibiloni R, Knauf C *et al*. Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes* 2008;**57**(6):1470-81.

(19) Cani PD, Possemiers S, van de Wiele T *et al*. Changes in gut microbiota control inflammation in obese mice through a mechanism involving GLP-2-driven improvement of gut permeability. *Gut* 2009;**58**(8):1091-103.

(20) Perry RJ, Peng L, Barry NA *et al*. Acetate mediates a microbiome-brain-beta-cell axis to promote metabolic syndrome. *Nature* 2016;**534**(7606):213-7.

(21) Dominy S, Casey L, Ermini, F, Benedyk, M, Marczyk A, Konradi, and A. Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors. Science Advances 5. 2019;5: eaau3333 DOI: 10.1126/sciadv.aau3333

(22) Leishman SJ, Do HL, Ford PJ. Cardiovascular disease and the role of oral bacteria. *J Oral Microbiol* 2010;**2**.

(23) Jia G, Zhi A, Lai PFH *et al*. The oral microbiota - a mechanistic role for systemic diseases. *Br Dent J* 2018;**224**(6):447-55.

(24) Macpherson AJ, Heikenwalder M, Ganal-Vonarburg SC. The Liver at the Nexus of Host-Microbial Interactions. *Cell Host Microbe* 2016;**20**(5):561-71.

(25) Costello EK, Lauber CL, Hamady M *et al*. Bacterial community variation in human body habitats across space and time. *Science* 2009;**326**(5960):1694-7.

(26) Lluch J, Servant F, Paisse S *et al*. The Characterization of Novel Tissue Microbiota Using an Optimized 16S Metagenomic Sequencing Pipeline. *PLoS One* 2015;**10**(11):e0142334.

**Supplementary Table 1: Identification of differentially abundant bacteria at the taxa order level in NAFLD-morbidly obese (MO) patients versus NAFLD-non-morbidly obese patients (non-MO)**

| Taxa: order | *P* (rank test) | Adjusted P (Bonferroni) | FDR | MO mean | Non-MO mean | Fold Change |
|---|---|---|---|---|---|---|
| Bacteroidales | 3.7E-15 | 7.3E-13 | 7.3E-13 | 5.48 | 2.31 | 2.375 |
| Clostridiales | 4E-13 | 7.8E-11 | 3.9E-11 | 4.17 | 2.11 | 1.97 |
| Azospirillales | 1.2E-08 | 0.0000024 | 0.00000078 | 0.042 | 0.3 | -7.017 |
| Sphingomonadales | 0.00000069 | 0.00014 | 0.000034 | 0.54 | 1.06 | -1.97 |
| Salinisphaerales | 0.0000012 | 0.00024 | 0.000047 | 0.2 | 0.61 | -3.068 |
| Xanthomonadales | 0.0000018 | 0.00035 | 0.000059 | 0.26 | 0.93 | -3.521 |
| Fimbriimonadales | 0.000013 | 0.0025 | 0.00036 | 0.0096 | 0.095 | -9.856 |
| Bacillales | 0.000068 | 0.013 | 0.0017 | 1.32 | 2.54 | -1.93 |
| Thermales | 0.00015 | 0.029 | 0.0033 | 0.53 | 1.47 | -2.77 |
| Reyranellales | 0.00019 | 0.037 | 0.0037 | 0.05 | 0.19 | -3.727 |
| Isosphaerales | 0.00034 | 0.067 | 0.0061 | 0.022 | 0.13 | -5.736 |
| Thiotrichales | 0.00053 | 0.1 | 0.0087 | 0.037 | 0.18 | -4.939 |
| Gemmatales | 0.00085 | 0.17 | 0.013 | 0.063 | 0.2 | -3.201 |

Statistical comparison of taxa abundances across sample groups; *p* values are adjusted for multiple testing by FDR, Benjamini-Hochberg or Bonferoni correction. Fold change stands for the MO mean/non-MO mean ratio, except where "-" indicates the inverse relation.

**Supplementary Table 2: Identification of differentially abundant bacteria at the taxa family level in NAFLD-morbidly obese (MO) patients versus NAFLD-non-morbidly obese patients (non-MO).**

| Taxa: family | P (rank test) | Adjusted P (Bonferroni) | FDR | MO mean | Non-MO mean | Fold Change |
|---|---|---|---|---|---|---|
| Muribaculaceae | 2.3E-15 | 6.9E-13 | 6.9E-13 | 3.58 | 1.04 | 3.451 |
| Lachnospiraceae | 2.4E-14 | 7.2E-12 | 3.6E-12 | 3.4 | 1.29 | 2.643 |
| Bacteroidaceae | 9.1E-14 | 2.7E-11 | 9.1E-12 | 3.45 | 1.57 | 2.197 |
| Prevotellaceae | 3E-12 | 9E-10 | 2.2E-10 | 1.85 | 0.79 | 2.341 |
| Rikenellaceae | 7.3E-10 | 0.00000022 | 4.4E-08 | 1.12 | 0.51 | 2.183 |
| Ruminococcaceae | 8.8E-10 | 0.00000026 | 4.4E-08 | 2.19 | 1.32 | 1.661 |
| Xanthomonadaceae | 7.1E-09 | 0.0000021 | 0.0000003 | 0.17 | 0.91 | -5.345 |
| Azospirillaceae | 0.000000012 | 0.0000036 | 0.00000045 | 0.042 | 0.3 | -7.017 |
| Solimonadaceae | 0.0000006 | 0.00018 | 0.00002 | 0.19 | 0.61 | -3.214 |
| Sphingomonadaceae | 0.00000069 | 0.00021 | 0.000021 | 0.54 | 1.06 | -1.97 |
| Hydrogenophilaceae | 0.0000049 | 0.0015 | 0.00013 | 0.063 | 0.31 | -4.856 |
| Bacillaceae | 0.000011 | 0.0033 | 0.00028 | 0.83 | 2.1 | -2.534 |
| Fimbriimonadaceae | 0.000013 | 0.0039 | 0.0003 | 0.0096 | 0.095 | -9.856 |
| Thermaceae | 0.00015 | 0.045 | 0.0032 | 0.53 | 1.47 | -2.77 |
| Reyranellaceae | 0.00019 | 0.057 | 0.0038 | 0.05 | 0.19 | -3.727 |
| Isosphaeraceae | 0.00034 | 0.1 | 0.0064 | 0.022 | 0.13 | -5.736 |
| Thiotrichaceae | 0.00053 | 0.16 | 0.0094 | 0.037 | 0.18 | -4.939 |
| Family_XVII | 0.00064 | 0.19 | 0.011 | 0 | 0.11 | -15590 |
| Chitinophagaceae | 0.00082 | 0.25 | 0.013 | 0.28 | 0.53 | -1.934 |
| Gemmataceae | 0.00085 | 0.26 | 0.013 | 0.063 | 0.2 | -3.201 |
| Caulobacteraceae | 0.0015 | 0.45 | 0.02 | 0.26 | 0.49 | -1.871 |
| Leuconostocaceae | 0.0015 | 0.45 | 0.02 | 0.28 | 0.13 | 2.104 |
| Akkermansiaceae | 0.0019 | 0.57 | 0.025 | 0.48 | 0.34 | 1.384 |
| Beijerinckiaceae | 0.0025 | 0.75 | 0.031 | 0.37 | 0.77 | -2.103 |
| Paenibacillaceae | 0.0032 | 0.96 | 0.038 | 0.17 | 0.59 | -3.43 |
| SAR116_clade | 0.0036 | 1 | 0.04 | 0.025 | 0.1 | -4.101 |
| uncultured_Chloroflexi_bacterium | 0.0036 | 1 | 0.04 | 0.01 | 0.068 | -6.826 |
| Solibacteraceae_Subgroup_3 | 0.0047 | 1 | 0.05 | 0.065 | 0.13 | -2.062 |
| Bacteroidetes_vadinHA17 | 0.0051 | 1 | 0.053 | 0.074 | 0.0053 | 13.86 |
| Streptococcaceae | 0.0059 | 1 | 0.059 | 0.63 | 0.89 | -1.416 |

Statistical comparison of taxa abundances across sample groups; *p* values are adjusted for multiple testing by FDR, Benjamini-Hochberg or Bonferoni correction. Fold change stands for the MO mean/non-MO mean ratio, except where "-" indicates the inverse relation.

**Supplementary Table 3: Identification of differentially abundant bacteria at the taxa genus level in NAFLD- morbidly obese (MO) patients versus NAFLD-non-morbidly obese patients (non-MO)**

| Taxa.__genus___ | P (rank test) | Adjusted P (Bonferroni) | FDR | MO mean | Non-MO mean | Fold Change |
|---|---|---|---|---|---|---|
| Roseburia | 1.5E-15 | 4.5E-13 | 2.7E-13 | 1.28 | 0.22 | 5.908 |
| Prevotellaceae_UCG001 | 2.2E-15 | 6.6E-13 | 2.7E-13 | 1.02 | 0.14 | 7.163 |
| uncultured_bacterium | 2.7E-15 | 8.1E-13 | 2.7E-13 | 3.79 | 1.65 | 2.291 |
| Alloprevotella | 3.6E-15 | 1.1E-12 | 2.7E-13 | 1.13 | 0.2 | 5.741 |
| Lachnospiraceae_NK4A136_group | 3.3E-14 | 9.9E-12 | 2E-12 | 2.27 | 0.5 | 4.515 |
| Bacteroides | 9.1E-14 | 2.7E-11 | 4.6E-12 | 3.45 | 1.57 | 2.197 |
| Oscillibacter | 1.9E-13 | 5.7E-11 | 8.1E-12 | 0.97 | 0.2 | 4.872 |
| Ruminiclostridium | 3.2E-13 | 9.6E-11 | 1.2E-11 | 0.77 | 0.14 | 5.39 |
| Ruminiclostridium_9 | 1.6E-11 | 4.8E-09 | 5.3E-10 | 0.72 | 0.19 | 3.753 |
| Azospirillum | 1.2E-10 | 3.6E-08 | 3.6E-09 | 0.005 | 0.23 | -46.11 |
| Lachnospiraceae_UCG006 | 1.9E-10 | 5.7E-08 | 5.2E-09 | 0.37 | 0.027 | 13.55 |
| Alistipes | 3.6E-10 | 0.00000011 | 9E-09 | 1.06 | 0.45 | 2.375 |
| Nevskia | 5.8E-09 | 0.0000017 | 0.00000013 | 0.1 | 0.56 | -5.584 |
| Sphingomonas | 0.00000011 | 0.000033 | 0.0000024 | 0.25 | 0.89 | -3.592 |
| Hydrogenophilus | 0.00000027 | 0.000081 | 0.0000054 | 0.018 | 0.2 | -11.15 |
| Cupriavidus | 0.00000084 | 0.00025 | 0.000016 | 0.014 | 0.21 | -14.6 |
| Anoxybacillus | 0.0000012 | 0.00036 | 0.000021 | 0.25 | 1.06 | -4.242 |
| Ruminococcus_torques_group | 0.0000013 | 0.00039 | 0.000021 | 0.52 | 0.24 | 2.183 |
| Vulcaniibacterium | 0.0000013 | 0.00039 | 0.000021 | 0.037 | 0.37 | -10.11 |
| Ruminiclostridium_5 | 0.000016 | 0.0048 | 0.00024 | 0.45 | 0.21 | 2.152 |
| Aquabacterium | 0.000094 | 0.028 | 0.0013 | 0.046 | 0.15 | -3.219 |
| Asinibacterium | 0.00016 | 0.048 | 0.0022 | 0 | 0.068 | -9667 |
| Reyranella | 0.00019 | 0.057 | 0.0025 | 0.05 | 0.19 | -3.727 |
| Leuconostoc | 0.00021 | 0.063 | 0.0026 | 0.25 | 0.085 | 2.933 |
| Meiothermus | 0.00027 | 0.081 | 0.0032 | 0.52 | 1.45 | -2.803 |
| Singulisphaera | 0.00028 | 0.084 | 0.0032 | 0.017 | 0.13 | -7.383 |
| Sediminibacterium | 0.0003 | 0.09 | 0.0033 | 0.15 | 0.39 | -2.522 |
| Thermomonas | 0.00032 | 0.096 | 0.0034 | 0 | 0.11 | -16350 |
| Thermobacillus | 0.00039 | 0.12 | 0.004 | 0.032 | 0.29 | -8.965 |
| Bryobacter | 0.0004 | 0.12 | 0.004 | 0.016 | 0.087 | -5.357 |
| Tyzzerella | 0.00055 | 0.16 | 0.0053 | 0.23 | 0.089 | 2.58 |
| Prevotella_1 | 0.00062 | 0.19 | 0.0056 | 0.26 | 0.1 | 2.486 |
| Thermaerobacter | 0.00064 | 0.19 | 0.0056 | 0 | 0.11 | -15590 |
| Methylocella | 0.00064 | 0.19 | 0.0056 | 0 | 0.067 | -9606 |
| uncultured | 0.00074 | 0.22 | 0.0063 | 1.96 | 1.61 | 1.218 |
| Lachnospiraceae_FCS020_group | 0.001 | 0.3 | 0.0083 | 0.12 | 0.022 | 5.62 |
| Akkermansia | 0.0019 | 0.57 | 0.015 | 0.48 | 0.34 | 1.384 |
| Maribacter | 0.002 | 0.6 | 0.016 | 0.37 | 0.22 | 1.715 |
| Methylobacterium | 0.0026 | 0.78 | 0.02 | 0.32 | 0.68 | -2.117 |
| Ruminococcaceae_UCG005 | 0.0031 | 0.93 | 0.023 | 0.27 | 0.13 | 2.121 |
| Caldibacillus | 0.0031 | 0.93 | 0.023 | 0.015 | 0.13 | -8.823 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Bacillus | 0.0036 | 1 | 0.025 | 0.57 | 1.16 | -2.048 |
| uncultured_Chloroflexi_bacterium | 0.0036 | 1 | 0.025 | 0.01 | 0.068 | -6.826 |
| Ruminococcus_gnavus_group | 0.0044 | 1 | 0.03 | 0.14 | 0.036 | 3.77 |
| Muribaculum | 0.0048 | 1 | 0.032 | 0.22 | 0.11 | 2.058 |
| Eubacterium_brachy_group | 0.0052 | 1 | 0.034 | 0.079 | 0.0081 | 9.788 |
| Chryseobacterium | 0.0061 | 1 | 0.039 | 0.079 | 0.2 | -2.481 |
| Unclassified | 0.0092 | 1 | 0.057 | 4.18 | 4.53 | -1.084 |
| Polaribacter_1 | 0.0094 | 1 | 0.058 | 0.024 | 0.075 | -3.111 |
| NS2b_marine_group | 0.0098 | 1 | 0.058 | 0.04 | 0.12 | -3.137 |
| Stenotrophomonas | 0.01 | 1 | 0.058 | 0.086 | 0.2 | -2.391 |
| Pedomicrobium | 0.01 | 1 | 0.058 | 0.047 | 0.12 | -2.582 |
| Succiniclasticum | 0.013 | 1 | 0.074 | 0.078 | 0.011 | 7.046 |
| Lachnoclostridium | 0.014 | 1 | 0.078 | 0.16 | 0.061 | 2.658 |
| Brevibacillus | 0.016 | 1 | 0.084 | 0.046 | 0.27 | -5.841 |
| Rhodococcus | 0.016 | 1 | 0.084 | 0.022 | 0.14 | -6.332 |
| Limnohabitans | 0.016 | 1 | 0.084 | 0.064 | 0.01 | 6.097 |
| Candidatus_Nitrosopumilus | 0.017 | 1 | 0.086 | 0.044 | 0.13 | -3.003 |
| FS14016B02_marine_group | 0.017 | 1 | 0.086 | 0.031 | 0.075 | -2.416 |
| Streptococcus | 0.02 | 1 | 0.1 | 0.42 | 0.63 | -1.508 |
| Ralstonia | 0.021 | 1 | 0.1 | 1.83 | 1.5 | 1.221 |
| SAR92_clade | 0.021 | 1 | 0.1 | 0.15 | 0.32 | -2.193 |
| Bifidobacterium | 0.021 | 1 | 0.1 | 0.13 | 0.27 | -2.119 |
| Ruegeria | 0.022 | 1 | 0.1 | 0.16 | 0.056 | 2.908 |
| Eubacterium_nodatum_group | 0.029 | 1 | 0.13 | 0.11 | 0.038 | 2.894 |
| Nocardioides | 0.032 | 1 | 0.15 | 0.018 | 0.073 | -3.985 |
| Geobacillus | 0.038 | 1 | 0.17 | 0.28 | 0.74 | -2.606 |
| Candidatus_Nitrosopelagicus | 0.04 | 1 | 0.18 | 0.028 | 0.071 | -2.527 |
| Staphylococcus | 0.044 | 1 | 0.19 | 0.82 | 0.68 | 1.205 |
| Diaphorobacter | 0.044 | 1 | 0.19 | 0.015 | 0.041 | -2.745 |
| Acidibacter | 0.045 | 1 | 0.19 | 0.31 | 0.22 | 1.411 |
| Aeromonas | 0.045 | 1 | 0.19 | 0.094 | 0.02 | 4.699 |
| Blautia | 0.048 | 1 | 0.19 | 0.25 | 0.17 | 1.415 |
| Legionella | 0.048 | 1 | 0.19 | 0.079 | 0.11 | -1.423 |
| Kocuria | 0.048 | 1 | 0.19 | 0.063 | 0.11 | -1.788 |
| Vulgatibacter | 0.048 | 1 | 0.19 | 0.015 | 0.11 | -7.004 |
| Massilia | 0.05 | 1 | 0.19 | 0.15 | 0.25 | -1.637 |

Statistical comparison of taxa abundances across sample groups; *p* values are adjusted for multiple testing by FDR, Benjamini-Hochberg or Bonferoni correction. Fold change stands for the MO mean/non-MO mean ratio, except where "-" indicates the inverse relation.

**Supplementary Table 4: Correlation analysis of BMI as continuous variable and liver tissue microbiome profiling**

| Feature | Spearman R | P value |
|---|---|---|
| ***Positive correlation with BMI*** | | |
| Bacteroidetes. D_2__Bacteroidia.D_3__Bacteroidales | 0.64 | $1.0 \times 10^{-8}$ |
| Firmicutes. D_2__Clostridia.D_3__Clostridiales | 0.60 | $1.0 \times 10^{-8}$ |
| Proteobacteria. D_2__Gammaproteobacteria.D_3__Cellvibrionales | 0.26 | 0.007055 |
| Proteobacteria. D_2__Gammaproteobacteria.D_3__Methylococcales | 0.27 | 0.005045 |
| ***Negative correlation with BMI*** | | |
| Firmicutes. D_2__Bacilli.D_3__Bacillales | -0.27 | 0.004633 |
| Proteobacteria. D_2__Alphaproteobacteria.D_3__Azospirillales | -0.43 | 0.000004 |
| Proteobacteria. D_2__Alphaproteobacteria.D_3__Reyranellales | -0.45 | 0.000001 |
| Proteobacteria. D_2__Alphaproteobacteria.D_3__Sphingomonadales | -0.37 | 0.000086 |
| Proteobacteria. D_2__Gammaproteobacteria.D_3__Salinisphaerales | -0.35 | 0.000172 |
| Proteobacteria. D_2__Gammaproteobacteria.D_3__Thiotrichales | -0.26 | 0.005681 |
| Proteobacteria. D_2__Gammaproteobacteria.D_3__Xanthomonadales | -0.40 | 0.000014 |

R stands for Spearman's rank correlation coefficient

**Supplementary Table 5: Enrichment of bacterial DNA according to histological traits in the NAFLD-bariatric surgery cohort (morbidly obese patients)**

| Disease severity (NASH vs. NAFL ) | | | | | |
|---|---|---|---|---|---|
| **Taxa (genus level)** | ***P* value** | **Mean NAFL** | **Mean NASH** | **Fold change** | **NCBI Taxonomy** |
| *Bifidobacterium* * | 0.0059 | 0.038 | 0.21 | 5.7 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Bifidobacteriales; f__Bifidobacteriaceae; g__Bifidobacterium. |
| *Ulvibacter* | 0.0081 | 0.19 | 0.42 | 2.2 | d__Bacteria; p__Bacteroidetes; c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__Ulvibacter |
| *Brevibacillus* * | 0.0087 | 0.095 | 0 | - | d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Paenibacillaceae; g__Brevibacillus |
| *Peptostreptococcus* * | 0.011 | 0.015 | 0.15 | 10 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptostreptococcaceae; g__Peptostreptococcus |
| *Lachnoclostridium* * | 0.018 | 0.24 | 0.09 | 0.4 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Lachnoclostridium |
| *Reyranella* | 0.015 | 0 | 0.097 | - | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodospirillales; f__; g__Reyranella |
| *Roseibacillus* | 0.016 | 0.01 | 0.11 | 11 | d__Bacteria; p__Verrucomicrobia; c__Verrucomicrobiae; o__Verrucomicrobiales; f__Verrucomicrobiaceae; g__Roseibacillus |
| *Stenotrophomonas* * | 0.019 | 0.041 | 0.13 | 3.1 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__Stenotrophomonas |
| *Massilia* * | 0.029 | 0.093 | 0.21 | 2.3 | d__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Oxalobacteraceae |
| *Streptomyces* * | 0.042 | 0.027 | 0.11 | 4.1 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Streptomycetales; f__Streptomycetaceae; g__Streptomyces |
| *Odoribacter* * | 0.043 | 0.051 | 0.16 | 3.1 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Odoribacteraceae; g__Odoribacter |
| **Taxa (family level)** | | | | | |
| *Bifidobacteriaceae* * | 0.0059 | 0.038 | 0.21 | 5.7 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Bifidobacteriales; f__Bifidobacteriaceae |
| *Rubritaleaceae* | 0.011 | 0.25 | 0.48 | 1.9 | d__Bacteria; p__Verrucomicrobia; c__Verrucomicrobiae; o__Verrucomicrobiales; f__Rubritaleaceae |
| *Peptostreptococcaceae* * | 0.013 | 0.041 | 0.22 | 5.3 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptostreptococcaceae |
| *Chthoniobacteraceae* | 0.024 | 0.14 | 0.031 | 0.2 | d__Bacteria; p__Verrucomicrobia; c__Spartobacteria; o__Chthoniobacterales; f__Chthoniobacteraceae |

| | | | | | |
|---|---|---|---|---|---|
| *Xanthomonadaceae* * | 0.039 | 0.11 | 0.23 | 2.1 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae |
| **Balloononig degeneration ( yes vs. no)** | | | | | |
| **Taxa (genus level)** | ***P* value** | **Mean no** | **Mean yes** | **Fold change** | **NCBI Taxonomy** |
| *Lachnoclostridium* * | 0.014 | 0.24 | 0.09 | 0.4 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Lachnoclostridium |
| *Turicibacter* * | 0.039 | 0.016 | 0.12 | 7.7 | d__Bacteria; p__Firmicutes; c__Erysipelotrichia; o__Erysipelotrichales; f__Erysipelotrichaceae; g__Turicibacter |
| **Taxa (family level)** | | | | | |
| *Nisaeaceae* | 0.029 | 0.023 | 0.12 | 5.3 | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodospirillales; f__Rhodospirillaceae |
| **Lobular inflammation ( yes vs. no)** | | | | | |
| **Taxa (genus level)** | ***P* value** | **Mean no** | **Mean yes** | **Fold change** | **NCBI Taxonomy** |
| *Succinivibrionaceae_UCG001** | 0.0055 | 0.12 | 0.012 | 0.1 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Aeromonadales; f__Succinivibrionaceae |
| *Anaeroplasma* * | 0.02 | 0.019 | 0.17 | 8.6 | d_Bacteria; p_Tenericutes; c_Mollicutes; o_Anaeroplasmatales; f_Anaeroplasmataceae; g_Anaeroplasma |
| *Parabacteroides* * | 0.023 | 0.067 | 0.25 | 3.8 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Tannerellaceae; g__Parabacteroides |
| *Roseomonas* * | 0.026 | 0.085 | 0.012 | 0.2 | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodospirillales; f__Rhodospirillaceae; g__Azospirillum |
| *Ruminococcaceae*_UCG005* | 0.036 | 0.37 | 0.21 | 0.6 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminococcus |
| **Taxa (family level)** | | | | | |
| *Tannerellaceae* | 0.023 | 0.067 | 0.25 | 3.8 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Tannerellaceae |
| **Portal inflammation ( yes vs. no)** | | | | | |
| **Taxa (genus level)** | ***P* value** | **Mean no** | **Mean yes** | **Fold change** | **NCBI Taxonomy** |
| *Lachnospiraceae*_NK4A136_group* | 0.0041 | 2.55 | 1.99 | 0.8 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__ |
| *Rubritalea* | 0.0043 | 0.039 | 0.2 | 5.1 | d__Bacteria; p__Verrucomicrobia; c__Verrucomicrobiae; o__Verrucomicrobiales; f__Rubritaleaceae; g__Rubritalea |
| *Bacteroides* * | 0.0064 | 3.7 | 3.19 | 0.9 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides |
| *Peptostreptococcus* * | 0.0073 | 0.014 | 0.16 | 11.4 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptostreptococcaceae; |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | g__Peptostreptococcus |
| *Reyranella* | 0.011 | 0 | 0.1 | - | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodospirillales; f__; g__Reyranella |
| *Roseburia ** | 0.012 | 1.42 | 1.15 | 0.8 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Roseburia |
| *Cloacibacillus ** | 0.018 | 0.01 | 0.14 | 14 | d__Bacteria; p__Synergistetes; c__Synergistia; o__Synergistales; f__Synergistaceae; g__Cloacibacillus |
| *Prevotella_9 ** | 0.021 | 0.24 | 0.51 | 2.17 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella |
| *Gemella ** | 0.022 | 0 | 0.13 | - | d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__; g__Gemella |
| *Prevotellaceae_UCG001* | 0.023 | 1.15 | 0.89 | 0.8 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__ |
| *Odoribacter ** | 0.027 | 0.049 | 0.16 | 3.3 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Odoribacteraceae; g__Odoribacter |
| *Ruminococcaceae*_UCG014* | 0.03 | 0.14 | 0.36 | 2.6 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminococcus |
| *Escherichia/Shigella sp ** | 0.038 | 0.17 | 0.41 | 2.4 | D_Bacteria; p_Proteobacteria; c_Gammaproteobacteria; o_Enterobacterales; f_Enterobacteriaceae; g_ unclassified Enterobacteriaceae |
| *Fusobacterium ** | 0.038 | 0.064 | 0.23 | 3.6 | d__Bacteria; p__Fusobacteria; c__Fusobacteriia; o__Fusobacteriales; f__Fusobacteriaceae; g__Fusobacterium |
| *Oscillibacter ** | 0.044 | 1.08 | 0.86 | 0.8 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Oscillospiraceae; g__Oscillibacter |
| **Taxa (family level)** | | | | | |
| *Lachnospiraceae ** | 0.0051 | 3.74 | 3.07 | 0.8 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae |
| *Bacteroidaceae ** | 0.0064 | 3.7 | 3.19 | 0.9 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f_ |
| *Peptostreptococcaceae ** | 0.0081 | 0.04 | 0.23 | 5.8 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptostreptococcaceae |
| *Enterobacteriaceae ** | 0.012 | 0.36 | 0.64 | 1.8 | d_Bacteria; p_Proteobacteria; c_Gammaproteobacteria; o_Enterobacterales; f_ |
| *Solimonadaceae* | 0.017 | 0.11 | 0.26 | 2.4 | d_Bacteria; p_Proteobacteria; c_Gammaproteobacteria; o_Nevskiales; f_ |
| *Synergistaceae* | 0.018 | 0.01 | 0.14 | 14 | d__Bacteria; p__Synergistetes; c__Synergistia; o__Synergistales; f__ |
| *Bacteriovoracaceae* | 0.028 | 0.097 | 0.24 | 2.5 | d__Bacteria; p__Proteobacteria; c__Oligoflexia; o__Bdellovibrionales; f__ |
| *Xanthomonadaceae ** | 0.04 | 0.11 | 0.23 | 2.04 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae |

| | | | | | |
|---|---|---|---|---|---|
| *Enterococcaceae* * | 0.04 | 0.036 | 0.19 | 5.3 | d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Enterococcaceae |
| **Fibrosis ( yes vs. no)** | | | | | |
| **Taxa (genus level)** | ***P* value** | **Mean no** | **Mean yes** | **Fold change** | **NCBI Taxonomy** |
| *Arenibacter* | 0.0034 | 0.19 | 0.032 | 0.16 | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Flavobacteriales; f__Flavobacteriaceae; g__Arenibacter |
| *Lactobacillus* * | 0.0094 | 0.87 | 0.62 | 0.7 | d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus |
| *Ruminococcaceae *_UCG009* | 0.013 | 0.042 | 0.17 | 4.1 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminococcus |
| *Halothiobacillus* | 0.029 | 0.015 | 0.096 | 6.5 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Chromatiales; f__Halothiobacillaceae; g__Halothiobacillus |
| *Nitrospira* * | 0.04 | 0.06 | 0.14 | 2.4 | d__Bacteria; p__Nitrospirae; c__Nitrospira; o__Nitrospirales; f__Nitrospiraceae; g__Nitrospira |
| *Blastococcus* * | 0.042 | 0.0044 | 0.096 | 22 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Geodermatophilales; f__Geodermatophilaceae; g__Blastococcus |
| *Roseibacillus* | 0.045 | 0.031 | 0.12 | 3.8 | d__Bacteria; p__Verrucomicrobia; c__Verrucomicrobiae; o__Verrucomicrobiales; f__Verrucomicrobiaceae; g__Roseibacillus. |
| **Taxa (family level)** | | | | | |
| *Crocinitomicaceae* | 0.0086 | 0.29 | 0.079 | 0.27 | d__Bacteria; p__Bacteroidetes; c__Flavobacteriia; o__Flavobacteriales; f__Crocinitomicaceae |
| *Lactobacillaceae* * | 0.0093 | 0.87 | 0.62 | 0.7 | d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae |
| *Chthoniobacteraceae* | 0.026 | 0.13 | 0.011 | 0.08 | d__Bacteria; p__Verrucomicrobia; c__Spartobacteria; o__Chthoniobacterales; f__Chthoniobacteraceae |
| *Halothiobacillaceae* | 0.029 | 0.015 | 0.096 | 6.5 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Chromatiales; f__Halothiobacillaceae |

* Denotes microorganisms that inhabit the human gastrointestinal tract (37). NCBI Taxonomy: stands for d: domain; p: phyla; c: class; o: order; f: family; g: genus (https://gtdb.ecogenomic.org/). *P* value: nonparametric one-way ANOVA (Wilcoxon rank test) statistics.

**Supplementary Table 6: Enrichment of bacterial DNA according to histological traits in NAFLD-overweight or moderately obese patients (non-morbidly obese patients)**

| **Disease severity (NASH vs. NAFL)** | | | | | |
|---|---|---|---|---|---|
| **Taxa (genus level)** | ***P* value** | **Mean NAFL** | **Mean NASH** | **Fold change** | **NCBI Taxonomy** |
| *Ruminiclostridium_9* | 0.0026 | 0.33 | 0.079 | 0.3 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminiclostridium |
| *Alloprevotella** | 0.0092 | 0.32 | 0.098 | 0.3 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Alloprevotella |
| *Ruminiclostridium* | 0.013 | 0.25 | 0.061 | 0.2 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminiclostridium |
| *Reyranella* | 0.015 | 0 | 0.097 | - | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodospirillales; f__; g__Reyranella |
| *Oscillibacter** | 0.026 | 0.33 | 0.098 | 0.3 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Oscillospiraceae; g__Oscillibacter |
| *Arenibacter* | 0.033 | 0.19 | 0.08 | 0.4 | d__Bacteria; p__Bacteroidetes; c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__Arenibacter |
| *Lachnospiraceae *_FCS020_group* | 0.039 | 0.049 | 0 | - | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__ |
| **Taxa (family level)** | | | | | |
| *Xanthomonadaceae** | 0.05 | 0.57 | 1.18 | 2 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae |
| **Balloononig degeneration (yes vs. no)** | | | | | |
| **Taxa (genus level)** | ***P* value** | **Mean no** | **Mean yes** | **Fold change** | **NCBI Taxonomy** |
| *Ruminiclostridium_9* | 0.0086 | 0.28 | 0.051 | 0.2 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminiclostridium |
| *Sphingobacterium* | 0.011 | 0.12 | 0.01 | 0.08 | d__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__Sphingobacteriaceae; g__Sphingobacterium |
| *Alloprevotella** | 0.017 | 0.28 | 0.068 | 0.2 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Alloprevotella |
| *Paracoccus ** | 0.017 | 0.28 | 0.49 | 1.75 | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__Paracoccu |
| *Ruminiclostridium* | 0.019 | 0.21 | 0.033 | 0.2 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminiclostridium |

| | | | | | |
|---|---|---|---|---|---|
| *Ruminococcaceae *_UCG009* | 0.036 | 0.053 | 0 | - | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminococcus |
| **Taxa (family level)** | | | | | |
| *Thiotrichaceae* | 0.017 | 0.11 | 0.3 | 2.8 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Thiotrichales; f__Thiotrichaceae; g__ |
| *Sphingobacteriaceae** | 0.024 | 0.17 | 0.053 | | d__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__Sphingobacteriaceae; g__Sphingobacterium |
| *Nocardiaceae** | 0.05 | 0.074 | 0.24 | 3.3 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Corynebacteriales; f__Nocardiaceae; g__Nocardia |
| **Lobular inflammation (yes vs. no)** | | | | | |
| **Taxa (genus level)** | ***P* value** | **Mean no** | **Mean yes** | **Fold change** | **NCBI Taxonomy** |
| *Lawsonella* | 0.00047 | 0.73 | 0.25 | 0.34 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Corynebacteriales; f__; g__Lawsonella |
| *Alloprevotella** | 0.0026 | 0.39 | 0.12 | 0.3 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Alloprevotella |
| *Prevotellaceae *_UCG001* | 0.0027 | 0.32 | 0.069 | 0.2 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella |
| *Bradyrhizobium* | 0.0028 | 0.096 | 0.47 | 5 | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Bradyrhizobiaceae; g__Bradyrhizobium |
| *Gemella** | 0.008 | 0.12 | 0.021 | 0.2 | d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__; g__Gemella |
| *Streptomyces** | 0.01 | 0.19 | 0.025 | 0.2 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Streptomycetales; f__Streptomycetaceae; g__Streptomyces |
| *Cutibacterium* | 0.014 | 1.02 | 0.6 | 0.6 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Propionibacteriales; f__Propionibacteriaceae; g__Cutibacterium |
| *Lachnospiraceae*_NK 4A136_group* | 0.016 | 0.84 | 0.36 | 0.4 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__ |
| *Haliangium* | 0.026 | 0.35 | 0.15 | 0.4 | d__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Myxococcales; f__Kofleriaceae; g__Haliangium |
| *Alistipes** | 0.031 | 0.66 | 0.36 | 0.5 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__Alistipes |
| *Acidovorax** | 0.033 | 0.58 | 0.32 | 0.6 | d__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Acidovorax |
| **Taxa (family level)** | | | | | |
| *Streptomycetaceae** | 0.0042 | 0.31 | 0.066 | 0.2 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Streptomycetales; f__Streptomycetaceae |
| *Xanthobacteraceae** | 0.0054 | 0.17 | 0.51 | 3 | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | f__Xanthobacteraceae |
| *Corynebacteriaceae** | 0.011 | 0.9 | 0.53 | 0.6 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Corynebacteriales; f__Corynebacteriaceae |
| *Propionibacteriaceae* | 0.017 | 1.03 | 0.61 | 0.6 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Propionibacteriales; f__Propionibacteriaceae |
| *Prevotellaceae** | 0.018 | 1.14 | 0.64 | 0.6 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae |
| *Nitrosomonadaceae* | 0.031 | 0.19 | 0.073 | 0.4 | d__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Nitrosomonadales; f__Nitrosomonadaceae |
| **Portal inflammation ( yes vs. no)** | | | | | |
| **Taxa (genus level)** | ***P* value** | **Mean no** | **Mean yes** | **Fold change** | **NCBI Taxonomy** |
| *Massilia** | 0.017 | 0.16 | 0.34 | 2.1 | d__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Oxalobacteraceae; g__Massilia |
| *Paenibacillus** | 0.031 | 0.17 | 0.045 | 0.3 | d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Paenibacillaceae; g__Paenibacillus |
| *Gemella** | 0.039 | 0.087 | 0.015 | 0.2 | d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__; g__Gemella |
| *Cycloclasticus* | 0.046 | 1.43 | 2.01 | 1.4 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Thiotrichales; f__Piscirickettsiaceae; g__Cycloclasticus |
| **Taxa (family level)** | | | | | |
| *Pirellulaceae* | 0.015 | 0.35 | 0.11 | 0.3 | d__Bacteria; p__Planctomycetes; c__Planctomycetia; o__Planctomycetales; f__Planctomycetaceae; g__Pirellula |
| *Planococcaceae** | 0.017 | 0.027 | 0.16 | 5.9 | d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Planococcaceae; g__ |
| *Hyphomonadaceae* | 0.024 | 0.78 | 1.14 | 1.5 | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodobacterales; f__Hyphomonadaceae; g__ |
| *Cycloclasticaceae* | 0.046 | 1.43 | 2.01 | 1.40 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Thiotrichales; f__Piscirickettsiaceae; g__Cycloclasticus |
| *Vibrionaceae** | 0.048 | 0.5 | 0.77 | 1.5 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Vibrionales; f__Vibrionaceae; g__Vibrio |
| **Fibrosis (yes vs. no)** | | | | | |
| **Taxa (genus level)** | ***P* value** | **Mean no** | **Mean yes** | **Fold change** | **NCBI Taxonomy** |
| *Veillonella** | 0.00076 | 0.06 | 0.26 | 4.4 | d__Bacteria; p__Firmicutes; c__Negativicutes; o__Veillonellales; f__Veillonellaceae; g__Veillonella |
| *Bosea* | 0.0089 | 0.0047 | 0.087 | 18.5 | d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Bradyrhizobiaceae; g__Bosea |
| *Alloprevotella** | 0.027 | 0.26 | 0.045 | 0.2 | d__Bacteria; p__Bacteroidetes; c__Bacteroidia; |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | o__Bacteroidales; f__Prevotellaceae; g__Alloprevotella |
| *Helicobacter** | 0.034 | 0.063 | 0.22 | 3.5 | d__Bacteria; p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Helicobacteraceae; g__Helicobacter |
| *Kocuria** | 0.035 | 0.15 | 0.018 | 0.1 | d__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Micrococcales; f__Micrococcaceae; g__Kocuria |
| *Christensenellaceae*_R7_group* | 0.037 | 0.044 | 0.17 | 3.9 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Christensenellaceae; g__ |
| *Ruminiclostridium_9* | 0.041 | 0.25 | 0.05 | 0.2 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminiclostridium |
| *Ruminiclostridium* | 0.05 | 0.19 | 0.025 | 0.2 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminiclostridium |
| **Taxa (family level)** | | | | | |
| *Thiotrichaceae* | 0.014 | 0.12 | 0.34 | 2.8 | d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Thiotrichales; f__Thiotrichaceae; g__ |
| *Hydrogenophilaceae* | 0.029 | 0.22 | 0.54 | 2.5 | d__Bacteria; p__Proteobacteria; c__Hydrogenophilalia; o__Hydrogenophilales; f__; g__; |
| *Helicobacteraceae** | 0.034 | 0.063 | 0.22 | 3.5 | d__Bacteria; p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Helicobacteraceae |
| *Christensenellaceae** | 0.037 | 0.044 | 0.17 | 3.9 | d__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Christensenellaceae |
| *Veillonellaceae** | 0.05 | 0.4 | 0.62 | 1.6 | d__Bacteria; p__Firmicutes; c__Negativicutes; o__Veillonellales; f__Veillonellaceae |

* Denotes microorganisms that inhabit the human gastrointestinal tract (37). NCBI Taxonomy: stands for d: domain; p: phyla; c: class; o: order; f: family; g: genus (https://gtdb.ecogenomic.org/). *P* value: nonparametric one-way ANOVA (Wilcoxon rank test) statistics