Supplementary Material and Methods - Bioinformatics and statistical analysis

Raw sequences, obtained by 16S rRNA amplicon sequencing on an Illumina MiSeq platform (NCBI SRA BioProject ID: PRJNA1216941), were processed using a pipeline combining PANDASeq (1) and QIIME 2 (2). After filtering for length and quality, reads were grouped into amplicon sequence variants (ASVs) using DADA2 (3). Taxonomic assignment was performed using the VSEARCH algorithm (4) against the SILVA database (August 2020 release) (5), with chimeras systematically discarded during analysis. Alpha diversity was assessed using several metrics, such as the Shannon index, the number of observed ASVs and Faith's phylogenetic diversity. Beta diversity was assessed using weighted UniFrac distances, which were then used for Principal Coordinates Analysis (PCoA) plots.

All statistical analyses were performed using R software. PCoA plots were generated using the "vegan" (https://cran.r-project.org/package=vegan) and "Made4" (6) packages, and data separation was tested using PERMANOVA (function "Adonis" in "vegan"). Group differences in alpha diversity and relative taxon abundance were assessed using the Kruskal-Wallis test followed by post-hoc Wilcoxon tests. P-values were adjusted using the Benjamini-Hochberg method. A false discovery rate (FDR) ≤ 0.05 was considered statistically significant, and FDR ≤ 0.1 was considered a trend.

REFERENCES

1. Masella AP, Bartram AK, Truszkowski JM, et al. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 2012;13:31.

2. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37(8):852-7.

3. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13(7):581-3.

4. Rognes T, Flouri T, Nichols B, et al. VSEARCH: a versatile open source tool for metagenomics. Peer J 2016;4:e2584.

5. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013;41(Database issue):D590-6.

6. Culhane AC, Thioulouse J, Perrière G, et al. MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics 2005;21(11):2789-90.