Data in Brief 57 (2024) 110902

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Intestinal microbial profiles of wild Alaskan rainbow trout (*Oncorhynchus mykiss*) characterized by 16S rRNA amplicon data

Aikaterini Katirtzoglou[a,*], Jacob A. Rasmussen[a], Daniel E. Schindler[b], Morten T. Limborg[a,*]

[a] *Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, Denmark*
[b] *School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195, USA*

ARTICLE INFO

Article history:
Received 15 May 2024
Revised 24 August 2024
Accepted 27 August 2024
Available online 5 September 2024

Dataset link: Microbial profiles of Alaskan rainbow trout from 16S rRNA amplicon data (Original data)

Keywords:
Animal microbiome
16S rRNA
Amplicon-sequencing
Rainbow trout
Wild populations

ABSTRACT

Rainbow trout (*Oncorhynchus mykiss*) is a dominant aquaculture species of the Salmonidae family, native only to the North Pacific. Recently, the gut microbiome has been shown to reflect the health status and responses to environmental changes in farmed fish. In this analysis we investigated the microbiome composition of the intestinal tract in 20 wild-caught rainbow trout specimens sampled in Alaska, USA. The targeted 16S rRNA gene (V3-V4 region) was sequenced on the Illumina NovaSeq 6000 platform. After quality control, demultiplexing and adapter trimming reads were analyzed using the DADA2 pipeline to obtain Amplicon Sequencing Variants (ASVs) which were subsequently taxonomically assigned. We found two phyla dominating the gut ecosystem present in every sample, Firmicutes and Fusobacteria, followed by lower abundances of Cyanobacteria, Proteobacteria and Bacteroidetes. At the genus level, we found high relative abundances of Cetobacterium and Clostridium sensu stricto 1. Interestingly, we did not identify often dominant genera *Mycoplasma*, *Pseudomonas* or *Weisella* which were prevalent in numerous studies previously, in cultured rainbow trout. Wild fish are exposed to a plethora of unpredictable environmental challenges, ranging from fluctuating water tempera-

* Corresponding authors.
E-mail addresses: akatirtzogloy@gmail.com (A. Katirtzoglou), morten.limborg@sund.ku.dk (M.T. Limborg).
Social media: @kid4_j (J.A. Rasmussen), @MortenLimborg (M.T. Limborg)

https://doi.org/10.1016/j.dib.2024.110902

tures to variable food availability, as opposed to controlled conditions in production facilities. Examining and comparing the gut ecosystem of wild and reared individuals holds great potential in optimizing management practices for commercially important species. Microbiome studies can provide novel ways to enhance the overall welfare of fish, strengthen disease prevention and increase sustainability in aquaculture production.

Specifications Table

Subject	Microbiology:Microbiome
Specific subject area	Molecular biology, Bioinformatics
Type of data	Raw (https://www.ebi.ac.uk/ena/browser/view/PRJEB75234) Analyzed, Filtered, Processed Table, Figure, Chart Scripts (https://github.com/katkat-ri/Microbiome-Analysis-RT)
Data collection	A total of 20 wild rainbow trout were collected using a non-invasive approach to squeeze out intestinal content, and the total bacterial DNA was extracted from these samples. The V3-V4 region of the 16S rRNA gene was amplified and sequenced on the NovaSeq Illumina sequencing platform. After filtering the raw reads, unique amplicon sequence variants (ASVs) were obtained using the DADA2 pipeline. The taxonomy assignment was implemented with the use of DADA2. The package phyloseq was used to investigate microbial community profiles and diversity estimators.
Data source location	Specimens were collected from the Wood River system of southwest Alaska, USA
Data accessibility	Repository name: ENA Data identification number: PRJEB75234 Direct URL to data: ?https://www.ebi.ac.uk/ena/browser/view/PRJEB75234 Scripts used for data processing available at GitHub (https://github.com/katkat-ri/Microbiome-Analysis-RT) Supplementary material available at: 10.6084/m9.figshare.26302240.v8

1. Value of the Data

1. The 16S rRNA amplicon data provide insight into the taxonomic composition and diversity of the gut microbiota in wild rainbow trout in Alaska.
2. The data are valuable for comparative analysis with other available studies of intestinal microbiome profiles in rainbow trout, with the potential to highlight differences observed between wild-caught and farmed individuals.
3. Access to this dataset can provide a baseline for the exploration of factors shaping the microbial composition in the gut of rainbow trout in the natural environment.
4. These data can help identify bacteria that have evolved adaptively relevant symbiotic relationships with rainbow trout to improve practices to modify the microbiome of farmed fish.

2. Background

There is growing evidence that our gut harbours a diverse community of microorganisms that constitute an integral part of our health [1]. Hence studying the role of intestinal microflora has been a major research avenue in recent years towards improving the welfare, sustainability

Table 1
Sequencing results, ASV read counts and alpha-diversity indices for 20 rainbow trout individuals. In the first column: Observed alpha-diversity (species richness), in the second and third Shannon and Simpson indices respectively.

Sample	Raw reads (forward/reverse)	ASV reads	Observed	Shannon Index	Simpson Index
RT1	234702	126414	19	0.600	0.378
RT2	1482	626	2	0.596	0.406
RT3	222992	95056	52	0.729	0.359
RT4	168672	64726	34	0.429	0.154
RT5	69879	19822	15	0.183	0.056
RT6	161386	56365	22	0.130	0.039
RT7	39787	18480	7	0.058	0.017
RT8	165775	88349	34	0.355	0.136
RT9	212481	110770	20	0.591	0.371
RT10	30596	13686	7	0.630	0.398
RT11	84884	31357	18	0.657	0.332
RT12	234359	83756	36	0.382	0.138
RT13	160716	43882	18	0.199	0.063
RT14	221381	75777	31	0.188	0.057
RT15	95135	41062	10	0.064	0.018
RT16	175506	87375	40	0.444	0.174
RT17	194562	101083	19	0.474	0.266
RT18	17434	9269	4	0.616	0.382
RT19	229605	92249	47	0.700	0.355
RT20	216378	82810	40	0.419	0.152
Ext_blk1	136684	45444	23	0.221	0.070
Ext_blk2	239346	82139	35	0.178	0.052
PCR_Neg1	38768	18371	6	0.103	0.034
PCR_Neg2	164424	79116	39	0.408	0.154
Mean	146538.9	61166			
SD	78849.73	35908.11			
Total	3516934	1467984			

and successful production of cultured animals. Rainbow trout is an economically valuable fish being farmed across all six inhabited continents. Several studies have characterized the microbiome communities residing in the intestinal tract of the species. 16S gut microbiome profiling of cultured rainbow trout from various geographical locations has shown that bacteria belonging to two phyla, Firmicutes and Proteobacteria, are commonly present in farmed cohorts of rainbow trout [2]. Other dominant phyla in the rainbow trout intestine include Actinobacteria, Bacteroides, and Tenericutes. Mycoplasma spp. has also been shown to be the most prevalent microorganism in many studies [3,4]. However, the majority of the studies focuses on the gut microbiome composition of rainbow trout reared under captivity. Our motivation for this data description stems from the lack of data on the microbiome composition of wild rainbow trout in the literature. In this article, we report a fundamental analysis of the gut microbiome diversity and community composition of rainbow trout from Alaskan wild populations.

3. Data Description

This dataset describes the gut microbiome of 20 wild rainbow trout samples collected in the Wood River system of southwest Alaska (Table S1). Microbiota composition was revealed by high-throughput sequencing of amplicons of the V3-V4 region of the 16S rRNA gene. Raw data were deposited in ENA on the PRJEB75234 repository. The sequencing yielded 7,033,868 (3,516,934 read pairs) in total (Table 1). After the filtering pipeline, we obtained a count table of 178 unique sequences representing Amplicon Sequencing Variants (ASVs) (Table S2). The number of ASV counts varied between samples, ranging from 626 to 126,414. The total number of ASV reads was 1,467,984 and the average number of ASV reads per sample was 61,166 (SD 35,908.11) (Table 1). Aside from the read count summary, Table 1 describes three alpha diversity estimators

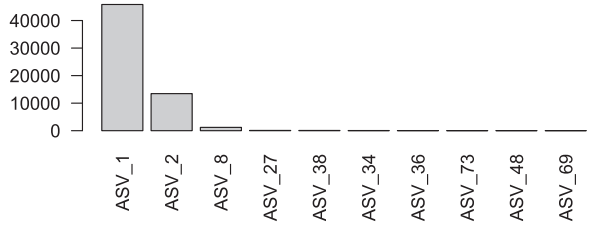

Fig. 1. Rank abundance bar plot of the most dominant ASVs in the dataset. The y-axis portrays the total number of counts for the 10 most abundant ASVs divided by the number of samples. The abundance data is not control-filtered.

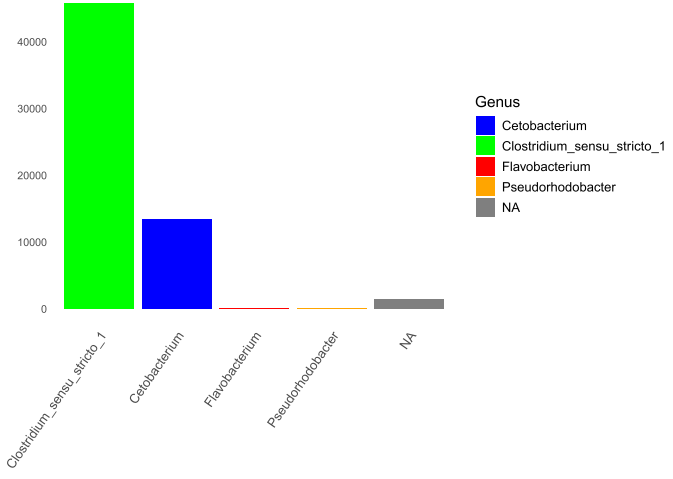

Fig. 2. Rank abundance bar plot of the most dominant ASVs in the dataset at the genus level. The y-axis portrays the total number of counts for the 10 most abundant genera divided by the number of samples. The rank abundance data of the most dominant genera in the dataset is not controlled filtered.

(Observed, Shannon and Simpson) based on the ASV counts. The dataset contains four negative control samples, two for microbial DNA extraction (Ext_blk1, Ext_blk2) and two for PCR amplification (PCR_Neg1, PCR_Neg2).

Fig. 1 represents a rank abundance bar plot of the top 10 most dominant ASVs. The counts are dramatically decreased after the three most abundant ASVs (ASV_1, ASV_2, ASV_8), which comprise 98.87 % of the total reads. Fig. 2 represents a rank abundance bar plot of the top 10 most dominant ASVs on the Genus level. Six out of ten ASVs (ASV_8, ASV_34, ASV_36, ASV_73, ASV_48, ASV_69) were not assigned to a known Genus.

A total of 16 different phyla in total were identified in the gut contents of our 20 rainbow trout samples (Table S3). Bacterial communities observed were dominated by the phyla: Firmicutes followed by Fusobacteria, Cyanobacteria, Proteobacteria and Bacteroidetes. The remaining phyla present in our dataset presented very low relative abundances and were thus aggregated in the “Other” category (Fig. 3a). Fig. 3b also shows the relative bacterial composition at the Genus level. We found 57 genera present in total and the most dominant were: Clostridium *sensu stricto* 1, Cetobacterium, Flavobacterium and Pseudorhodobacter. The “Unknown” category represented taxa for which we did not have available taxonomical data at the genus level, and “Other” represents different taxa with lower relative abundances which are not displayed here.

Due to elevated alpha diversity estimates in the control samples (excluding PCR_Neg1), we investigated the ASV composition between the control group and the biological samples. Out of 178 ASVs that this analysis was based on, 27.5 % (n=49) were shared between the control group

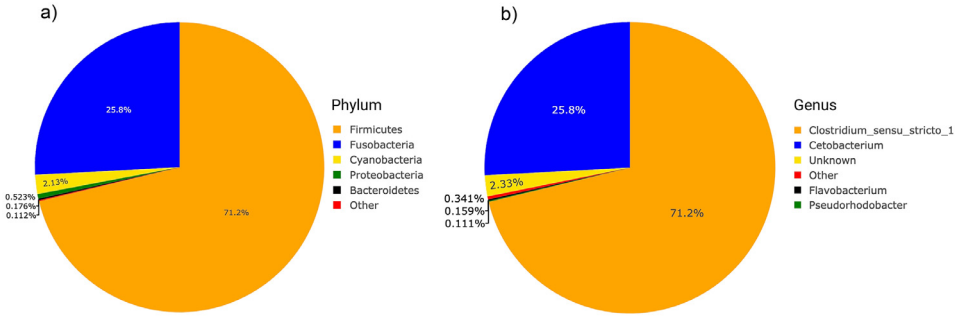

Fig. 3. Pie diagrams of bacterial community composition based on the relative abundance of microbiota in the 20 rainbow trout samples at : (a) Phylum level. Five most abundant phyla are displayed and the remaining are concatenated in the "Other" taxon category. (b) Genus level. The four most abundant genera are displayed while the "Unknown" category represents different taxa with no available information for the Genus level. The "Other" category represents the remaining genera present in the dataset with lower relative abundances. The microbiome composition presented in this figure is based on all the taxa identified in the dataset, thus it is not filtered for control-specific taxa (see text for further details).

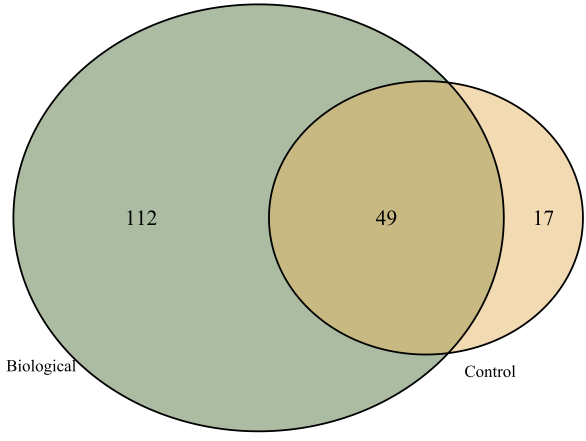

Fig. 4. Venn diagram of shared ASVs between biological samples and control group.

and biological samples. Approximately 63 % of the ASVs (n=112) were only detected in biological samples and 9.5 % of ASVs (n=17) were only found in controls (Fig. 4).

In Table S4 we show taxonomic assignment of the 49 ASVs that were identified both in the control group and biological samples. The most dominant genera represented in the dataset (Clostridium *sensu stricto* 1, Cetobacterium, Flavobacterium and Pseudorhodobacter) were identified both in the control group and the biological samples. Clostridium *sensu stricto* 1 and Cetobacterium were present in 100 % of the 20 biological samples, whereas Flavobacterium and Pseudorhodobacter were present in 75 % and 55 % of the samples, respectively (Table S4). This observation highlights the importance of accounting for controls in microbiome studies, especially in fish which have been often associated with low biomass and diversity of intestinal microbiota [5]. However, while these 49 ASVs occuring in both the control and biological samples cannot be confidently assigned to the wild rainbow trout guts, we argue these ASVs remain relevant as candidate microbes putatively residing in wild rainbow trout microbiomes.

The taxonomic assignment of the 17 ASVs which were not present in biological samples is shown in Table S5. Out of the 17 control-specific ASVs, four were assigned to the Bacteroidetes phylum and two ASVs to the Actinobacteria phylum (Fig. 5). The genera identified in the control-

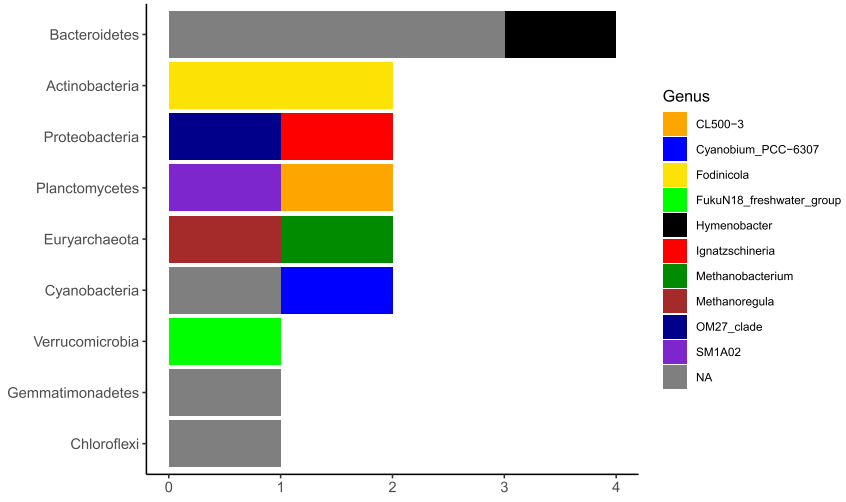

Fig. 5. Barplot of taxonomic assignment of control-specific ASVs. The y-axis shows the different phyla exclusively present in the four controls; the x-axis shows the number of ASVs. Each bar consists of colored segments, where each one represents a genus.

specific samples did not overlap with any of the most abundant taxa in the gut microbiome composition of trout (Fig. 2). It is crucial to distinguish the taxa that are not identified in genuine gut samples as they are likely a result of contamination, and do not represent the true microbiome composition.

4. Experimental Design, Materials and Methods

4.1. Data Collection

Rainbow trout were sampled with stick seines in Hidden Creek and Lynx Creek, which are third-order tributaries flowing into Lake Nerka within the Wood River system [6]. Fish were transferred to aerated buckets before anaesthetising them with Eugenol for gut sampling. Alaskan wild rainbow trout populations are protected so we were restricted to a non- invasive sampling protocol. Briefly, all fish were trapped using a gill net and gently placed on a wet surface whereafter intestinal content (digesta) were retrieved with a gentle squeeze along the abdomen. Therefore, our samples likely represent mainly the distal intestinal content with potential inclusion of content from the midgut, but exclusive of content from the stomach and pyloric caeca regions. We also assume limited inclusion of gut epithelium as the fish were sampled non-invasively. The gut content was sampled in sterile white weight trays or Petri dishes. Approximately 200 mg of gut content was sampled from a Petri dish using forceps. Forceps were disinfected with 5 % bleach and 80 % ethanol between individuals to prevent cross-contamination among samples. All samples were preserved in the SHIELD^{TM}buffer, from Zymo Research, following the Zymo Research standard procedure.

4.2. Microbial DNA Extraction and Amplicon Sequencing

All DNA extractions were performed in a dedicated clean laboratory that is isolated from post-PCR laboratories to limit the chance of contamination from previously amplified DNA. DNA extractions for 16S rRNA gene profiling were carried out using Zymo Research Quick-DNA/RNA

(Cat. D2131) following the suppliers' recommendation. We used qPCR on extracts to quality-control for inhibitors and optimal PCR settings before metabarcoding. Two extraction negatives, two library negatives, and two PCR negatives were included for each plate. Metabarcoding was carried out by amplifying the V3-V4 region of the bacterial 16S rRNA gene, using the primers 341F (5′-CCTAYGGGRBGCASCAG-3′) and 806R (5′-GGACTACNNGGGTATCTAAT-3) combined with unique forward and reverse 8-bp tags as previously done in Rasmussen et al. 2022 [4]. All amplifications were carried out in triplicates to lower the number of cycles needed for PCR and to minimise procedural false positives [7]. PCR amplification was carried out, using 35 cycles. PCR products were visualized using gel electrophoresis (GE). PCR replicates were pooled into a single pool based on gel intensity, and then converted to Illumina sequencing libraries using the BEST protocol [8]. After BEST preparation libraries were indexed for Illumina sequencing using 10 cycles. Amplicons were sequenced on an Illumina NovaSeq 6000 PE250bp to obtain 250bp paired-end reads aiming for a minimum of 50,000 reads per PCR replica.

4.3. Bioinformatic Analysis

Raw sequence data were quality controlled, using FastQC/v0.11.8 to assess low-quality reads. Demultiplexing and removal of adaptors and low-quality reads were done with AdapterRemoval/v2.2.4, with a base quality threshold of 30 and a minimum read length of 50bp.

This analysis began with 20 Illumina-sequenced paired-end, demultiplexed fastq files from which the adapters had already been removed. Trimmed reads were dereplicated and chimaeras were removed using the R package DADA2 [9]. The output was an amplicon sequence variant (ASV) table, which records the counts of ASVs in each sample. Taxonomy was assigned through DADA2 using Silva/v138 database [10]. In addition, a post-clustering algorithm was applied to remove false positives from the dataset using LULU [11]. Composition plots of the relative abundance of taxa were carried out using phyloseq [12] and microbiome [13] R packages. Estimators for alpha diversity were also calculated using the phyloseq package. The four negative controls (Ext_blk1, Ext_blk2, PCR_Neg1, PCR_Neg2) were not filtered out prior to the microbiome composition analysis. We investigated the distribution of ASVs specific to the controls and to the biological samples. Finally, we separately reported the taxa found exclusively in the negative controls.

Limitations

One potential limitation is the small sample size which could affect the precision of characterization of microbial communities and the diversity estimators.

Ethics Statement

The current work does not involve human subjects, animal experiments, or any data collected from social media platforms. Sampling of trout in the field was allowed under permits from the State of Alaska, and University of Washington IACUC.

Data Availability

Microbial profiles of Alaskan rainbow trout from 16S rRNA amplicon data (Original data) (European Nucleotide Archive (ENA))

CRediT Author Statement

Aikaterini Katirtzoglou: Conceptualization, Investigation, Visualization, Software; **Jacob A. Rasmussen:** Conceptualization, Investigation, Data curation, Software; **Daniel E. Schindler:** Data curation, Methodology, Writing – review & editing; **Morten T. Limborg:** Conceptualization, Investigation, Funding acquisition, Supervision, Writing – review & editing.

Acknowledgements

This research received funding from a Danish National Research Foundation grant to M.T.L. (DNRF143). The study did not receive any funding from commercial sectors.

Declaration of Competing Interest

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

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110902.

References

[1] S. Consuegra, T.U. Webster, I. Anka, Microbiome, epigenetics and fish health interactions in aquaculture, in: Epigenetics in Aquaculture, 1st ed., Wiley, 2023, pp. 245–262, doi:10.1002/9781119821946.ch11.
[2] I.S. Hines, M.A. Marshall, S.A. Smith, D.D. Kuhn, A.M. Stevens, Systematic literature review identifying bacterial constituents in the core intestinal microbiome of rainbow trout (Oncorhynchus mykiss), Aquac. Fish Fish. 3 (5) (2023) 393–406, doi:10.1002/aff2.127.
[3] S. Cao, J. Dicksved, T. Lundh, A. Vidakovic, P. Norouzitallab, D. Huyben, A meta-analysis revealing the technical, environmental, and host-associated factors that shape the gut microbiota of Atlantic salmon and rainbow trout, Rev. Aquac. (2024), doi:10.1111/raq.12913.
[4] J.A. Rasmussen, K.R. Villumsen, M. Ernst, M. Hansen, T. Forberg, S. Gopalakrishnan, M.T.P. Gilbert, A.M. Bojesen, K. Kristiansen, M.T. Limborg, A multi-omics approach unravels metagenomic and metabolic alterations of a probiotic and synbiotic additive in rainbow trout (Oncorhynchus mykiss), Microbiome 10 (1) (2022) 21, doi:10.1186/s40168-021-01221-8.
[5] M.T. Limborg, P.Y.S. Chua, J.A. Rasmussen, Unexpected fishy microbiomes, Nat. Rev. Microbiol. 21 (6) (2023) 346–346, doi:10.1038/s41579-023-00879-1.
[6] K.T. Bentley, D.E. Schindler, J.B. Armstrong, T.J. Cline, G.T. Brooks, Inter-tributary movements by resident salmonids across a boreal riverscape, Plos one 10 (9) (2015) e0136985, doi:10.1371/journal.pone.0136985.
[7] A. Alberdi, O. Aizpurua, M.T.P. Gilbert, K. Bohmann, Scrutinizing key steps for reliable metabarcoding of environmental samples, Methods Ecol. Evol. 9 (1) (2018) 134–147, doi:10.1111/2041-210X.12849.
[8] C. Carøe, S. Gopalakrishnan, L. Vinner, S.S.T. Mak, M.H.S. Sinding, J.A. Samaniego, N. Wales, T. Sicheritz-Pontén, M.T.P. Gilbert, Single-tube library preparation for degraded DNA, Methods Ecol. Evol. 9 (2) (2018) 410–419, doi:10.1111/2041-210X.12871.
[9] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, S.P. Holmes, DADA2: High resolution sample inference from Illumina amplicon data, Nat. Methods 13 (7) (2016) 581–583, doi:10.1038/nmeth.3869.
[10] B. Callahan, (2018). Silva taxonomic training data formatted for DADA2 (Silva version 132). Zenodo. 10.5281/zenodo.1172783
[11] T.G. Frøslev, R. Kjøller, H.H. Bruun, R. Ejrnæs, A.K. Brunbjerg, C. Pietroni, A.J. Hansen, Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates, Nat. Commun. 8 (2017) 1188, doi:10.1038/s41467-017-01312-x.
[12] P.J. McMurdie, S. Holmes, phyloseq: An R Package for reproducible interactive analysis and graphics of microbiome census data, PLOS ONE 8 (4) (2013) e61217, doi:10.1371/journal.pone.0061217.
[13] L. Leo et al. Microbiome R package. URL: http://microbiome.github.io. (2024). [Computer software].