DOI: 10.4081/ijfs.2025.13171

SUPPLEMENTARY MATERIAL

16S rRNA metabarcoding applied to the microbiome of insect products (novel food): a comparative analysis of three reference databases

Gabriele Spatola,[1] Alice Giusti,[1] Laura Gasperetti,[2] Roberta Nuvoloni,[1] Alessandra Dalmasso,[3] Francesco Chiesa,[3] Andrea Armani[1]

[1]Department of Veterinary Sciences, University of Pisa; [2]Experimental Zooprophylactic Institute of Lazio and Tuscany, Pisa; [3]Department of Veterinary Sciences, University of Turin, Grugliasco (TO), Italy

Correspondence: Gabriele Spatola, Department of Veterinary Sciences, University of Pisa, Viale delle Piagge, 2, 56124, Pisa (PI), Italy.
E-mail: g.spatola2@studenti.unipi.it

Key words: 16S rRNA metabarcoding, NGS, food microbiome, genomic reference database.

Supplementary Materials. Materials and Methods

Library preparation and Miseq (Illumina®) sequencing

The 460 bp V3-V4 hypervariable regions of the 16S rRNA gene was amplified using the Illumina adapter sequence and the primer overhang (forward and reverse) reported in the Library Preparation Guide (LPG) (https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html). Each DNA sample was amplified with the following PCR reaction: 2.5 µl of 5 ng/ µl DNA, 5 µl of primer forward overhang, 5 µl of primer reverse overhang, 12.5 µl of 2x KAPA HiFi HotStart ReadyMix (KAPA Biosystems). The cycling program reported in the LPG was used. PCR products were resolved by electrophoresis on a 2 % agarose gel (GellyPhor LE, Euroclone SPA, Milano, Italy) stained with GelRed™ Nucleid Acid Gel Stain (Biotium, Hayward, CA, USA). The presence of PCR products of the expected length was visualized by UV light transilluminator. Then, DNA amplicons were purified with the NucleoMag kit for clean up and size selection of NGS library prep reactions (Macherey-Nagel) and each sample was indexed using Illumina® DNA/RNA UD Indexes Tagmentation kit, following the manufacturer's instructions. The libraries were further purified before their validation and quantification. The libraries validation was performed on an Agilent 4150 TapeStation D1000 ScreenTape assay (Agilent Technologies Inc.) to verify the size, while quantification was performed using a Qubit 4 Fluorometer (Thermo Fisher Scientific, USA). The final DNA concentration was calculated in nM, based on the size of DNA amplicons, and applying the formula reported in the Illumina LPG. Finally, aliquots of 5 µl of diluted DNA from each library were mixed for pooling libraries with unique indexes. Pooled libraries were denatured and diluted according to Illumina LPG instructions before Miseq loading. Pooled libraries were loaded using MiSeq Reagent Micro Kit v2 (500-cycles), and the run included a 20% PhiX to serve as an internal control.

Bioinformatic analysis

Sequencing data, consisting in folder containing FASTQ files with raw reads (R1 files containing forward reads of each sample and R2 files containing reverse reads of each sample) were quality checked using FastQC (Babraham Institute, Cambridge, UK). Then, R1 and R2 files were processed to generate amplicon sequence variants (ASVs) using DADA2 R package (Callahan et al., 2016) (**Figure 1**). A final ASV table summarizing the number of different ASVs per sample was produced.

Supplementary Table 1. Selected reference formatted databases. The version used for each reference formatted database is provided, along with the total number of sequences contained in the database and their distribution across domains.

RFDs	Version	Archaea sequences	Bacteria sequencies	Eukaryota sequences	Total sequences
SILVA	Silva_nr99_version 138.1 (McLaren & Callahan, 2021),	20,389	431,575	100	452,064
REF	NCBI RefSeq 16S rRNA database supplemented by RDP v.16 (Alishum, 2019)	1055	22433	99	23,587
RDP	RDP trainset 18 (Callahan, 2020).	601	20,593	1	21,195

RFDs, reference formatted databases.

Supplementary Table 2. Amount amplicon sequence variants assigned to each taxonomic level using different reference formatted databases after rarefaction and filtration.

RFDs	ASVs assigned at Phylum level (%; n)	ASVs assigned at Class level (%; n)	ASVs assigned at Order level (%; n)	ASVs assigned at Family level (%; n)	ASVs assigned at Genus level (%; n)
SILVA	98.7% (n=3949/4002)	98.6% (n=3946/4002)	98.4% (n=3939/4002)	94.7% (n=3788/4002)	76.1% (n=3045/4002)
RDP	98.3% (n=3922/3991)	97.2% (n=3880/3991)	96.5% (n=3852/3991)	90.5% (n=3610/3991)	66.1% (n=2637/3991)
REF	99.0 % (n=3919/3958)	97.2% (n=3840/3958)	96.2% (n=3807/3958)	88.2% (n=3489/3958)	67.9% (n=2686/3958)

ASVs, amplicon sequence variants; RFDs, reference formatted databases.