Funding
HAS (MTA)
Lendület Programme
LP-2020/8

Molecular Med TRI-CON 2023
March 6-8, 2023, San Diego, CA, USA

Significant Inconsistences among the Technical Approaches Applied for Gut Microbiome Analyses

Dóra Tombácz[1,2], Gábor Gulyás[1,2], Balázs Kakuk[1,2], Ákos Dörmő[1,2], Tamás Járay[1,2], István Prazsák[1,2], Zsolt Csabai[1,2], Miksa Henkrich[2], Zsolt Boldogkői[1,2]

[1]Department of Medical Biology, Albert Szent-Györgyi Medical School, University of Szeged
[2]MTA-SZTE Lendület GeMiNI Research Group, University of Szeged

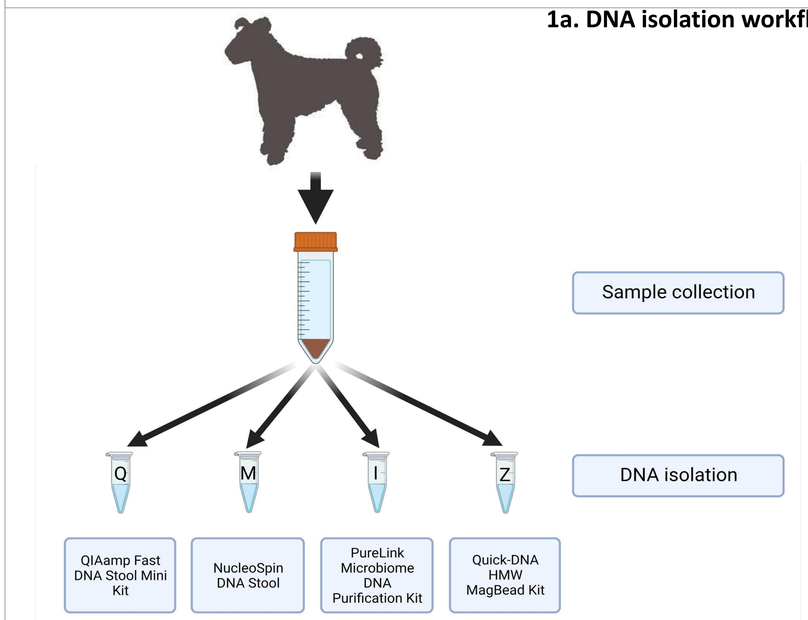
INTRODUCTION Gut microbiome plays an important role in affecting host physiology and in the etiology of several diseases. A number of protocols are available for isolation, sequencing and analysis of microbial nucleic acids. The process leading to the data however, is a complex series of steps that can each have different approaches, which may in turn lead to dissimilar results. Choosing the most reliable method is crucial at every step of the sample processing workflow, as the approach of these may differ significantly. We conducted a set of experiments comparing some of the most widely used DNA isolation, sequencing and bioinformatic approaches and a bioinformatic tool, developed by our group. We have also used a microbial community mix for the validation of our results. Alpha diversity analysis was used to describe the variety and richness of the microbiota within the sample. Our results provide an optimized experimental design, a start-to-end wet-lab and bioinformatic pipeline for gut microbiome analysis.

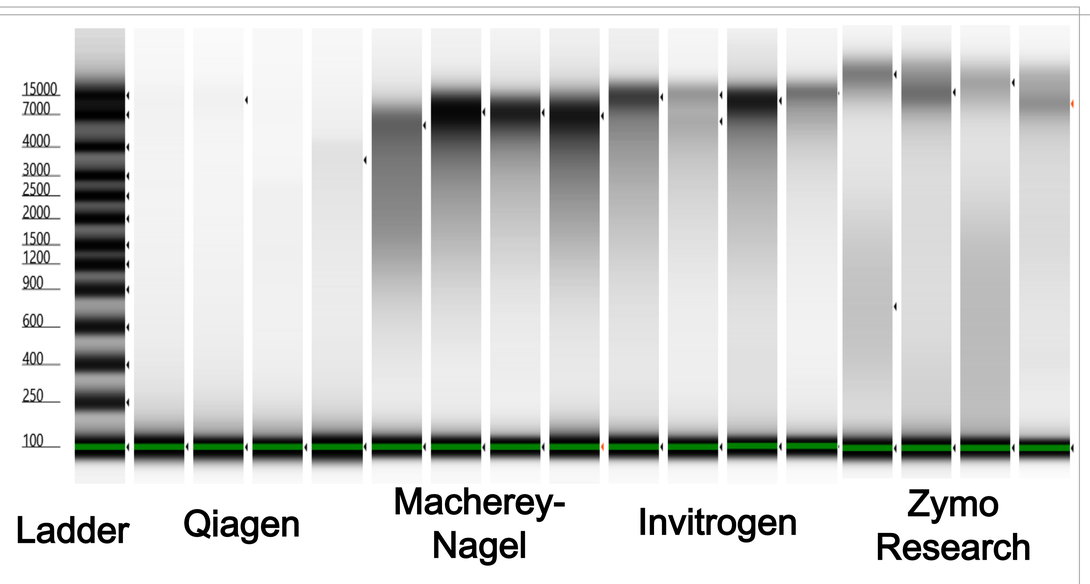
METHODS In this pilot project, the fecal microbiome of a 13 years old healthy dog (a Hungarian purebred dog: Pumi) was analyzed. The DNA isolated from the stool sample using four different purification kits (**Qiagen** (Q) *QIAamp Fast DNA Stool Mini Kit*, **Invitrogen** (I) *PureLink™ Microbiome DNA Purification Kit*, **Macherey-Nagel** (MN) *NucleoSpin DNA Stool Mini kit*, **ZymoResearch** (ZR) *ZymoBIOMICS DNA Miniprep Kit*). DNA samples were then used to generate libraries for short- and long-read sequencing. The 16S-based amplicon sequencing (V1-V2, V1-V3, V3-V4) was carried out on **Illumina MiSeq**, and (V1-V9) on **Oxford Nanopore Technologies MinION**, as well as on **Pacific Biosciences Sequel IIe** platforms. Shotgun sequencing approach was also applied. The V1-V3 libraries were prepared with the **PerkinElmer** ***NEXTFLEX®*** *16S* ***V1-V3*** *Amplicon-Seq Kit*. The *Quick-16S NGS Library Prep Kit* from **ZymoResearch** was applied to generate V1-V2 and V3-V4 libraries. The V1-V9 libraries for full-length 16S sequencing on long read sequencing platforms were produced by using the *16S Barcoding Kit* from **Oxford Nanopore Technologies** (ONT) and by using the *SMRTbell Express Template Prep Kit 2.0* from **Pacific Biosciences** (PacBio). The *DNA Prep Kit* (**Illumina**) was used for the generation of shotgun libraries. The following software packages and tools were used for the analysis: *Kaiju* (https://github.com/bioinformatics-centre/kaiju), *Emu* (https://gitlab.com/treangenlab/emu), *EPI2ME* (ONT) and *DADA2 v.1.18* (https://benjjneb.github.io/dada2/) and **MiniTax** (developed by our group).
Analysis of DNA and libraries: 1) quantity - *Qubit 4.0* and *Qubit HS DNA Assay Kit* (**Invitrogen/Thermo Scientific**); 2) quality: **Agilent** *TapeStation 4150, D1000 ScreenTape* and *Genomic DNA ScreenTape*

RESULTS DNA purification kits produce different total DNA concentrations. We also found significant effect of DNA isolation kits on DNA purity and integrity. Three from the four DNA isolation kits provided good quality DNA for high throughput sequencing with non-significant differences in the number of reads recovered, alpha, or beta diversity. One of the tested DNA purification method resulted in significantly different DNA yield in the case of the technical parallels. Our data shown that both the choice of isolation and library preparation (sequencing approaches), as well as the bioinformatics significantly affected the *Firmicutes* to *Bacteroidetes* ratio. All taxonomic levels are affected by the applied methods.

CONCLUSION Our results are in a concordance with other microbiome studies suggesting that the applied DNA extraction techniques could bias results. The choice of DNA isolation kits, library preparation and sequencing approaches, as well as bioinformatics pipelines (databases, analysis techniques) has an important effect on the results that has to be taken into account mainly in meta-analyses.

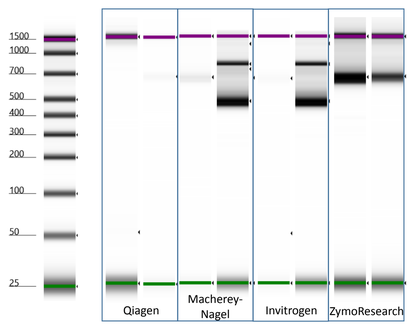
1a. DNA isolation workflow



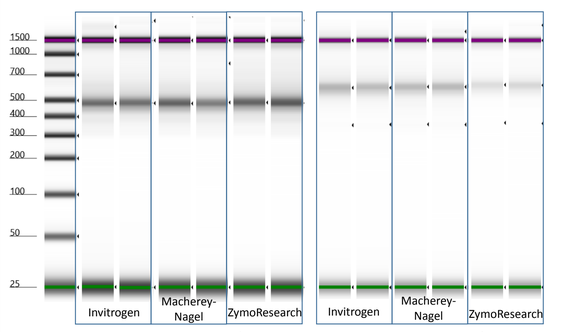


This figure shows the sizing analysis of DNA samples using the TapeStation 4150 system and the Genomic DNA ScreenTape assay. The longest DNA samples were generated by the ZymoReserach kit, while the Qiagen produced the shortest and worst quality samples.

Analysis of amplicon libraries using D1000 ScreenTape assay. The ZymoResearch kit resulted in high-quality libraries from each of the applied DNA purification methods (except Qiagen). In contrast, the PerkinElmer kit resulted in false product in many cases (Qiagen DNA: low library amount, MN and Invitrogen: aspecific products)



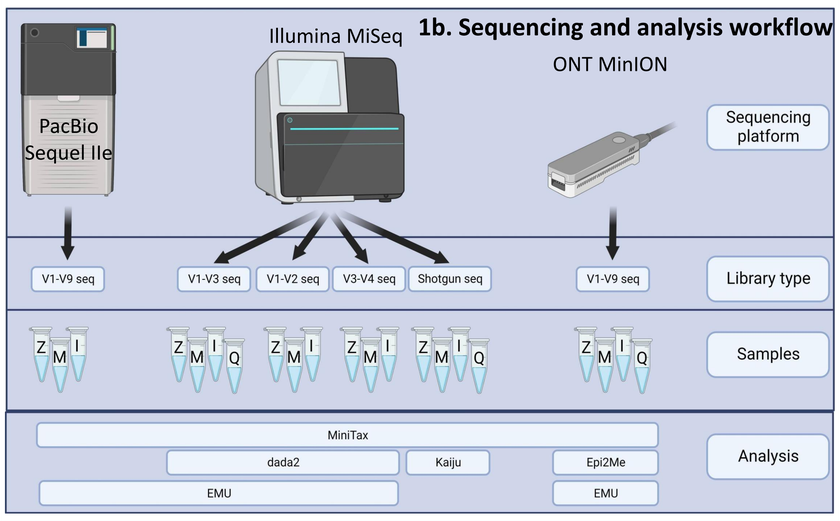
PerkinElmer
NEXTFLEX V3-V4

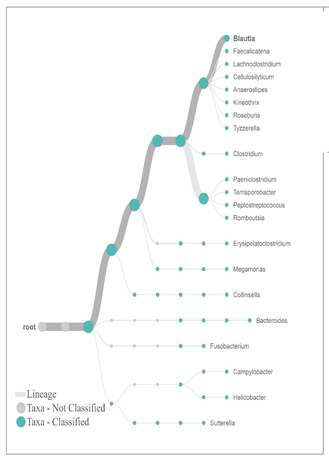


ZymoResearch Quick-16S NGS V1-V2

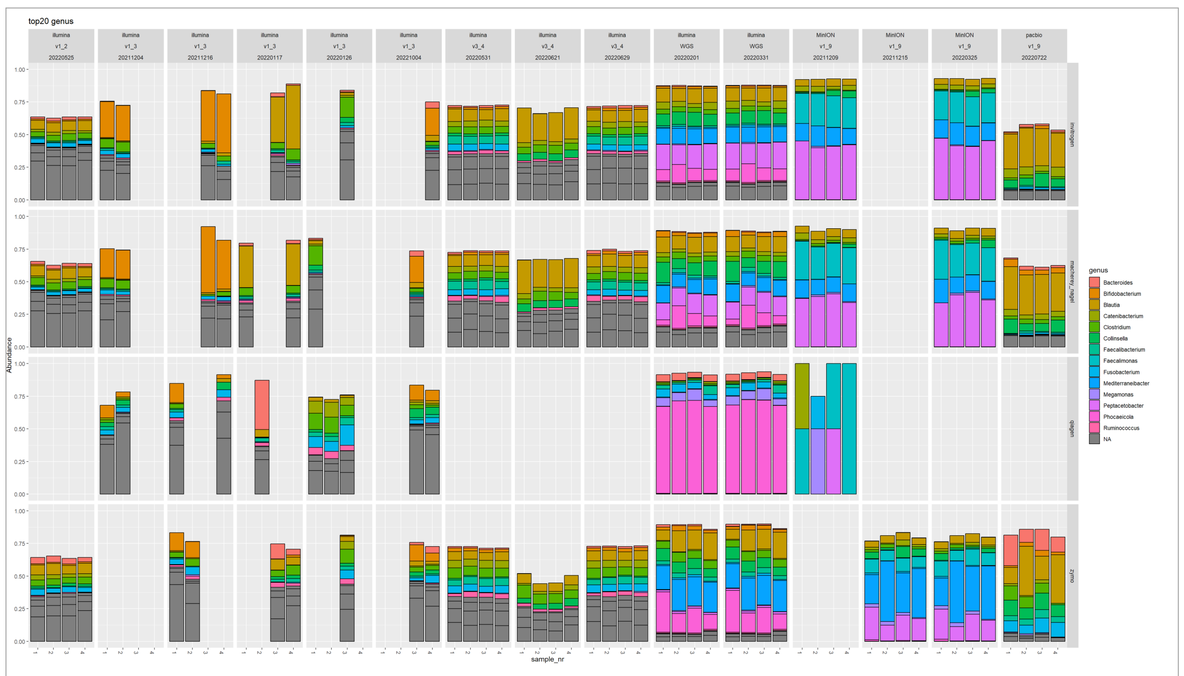
ZymoResearch Quick-16S NGS V3-V4

1b. Sequencing and analysis workflow

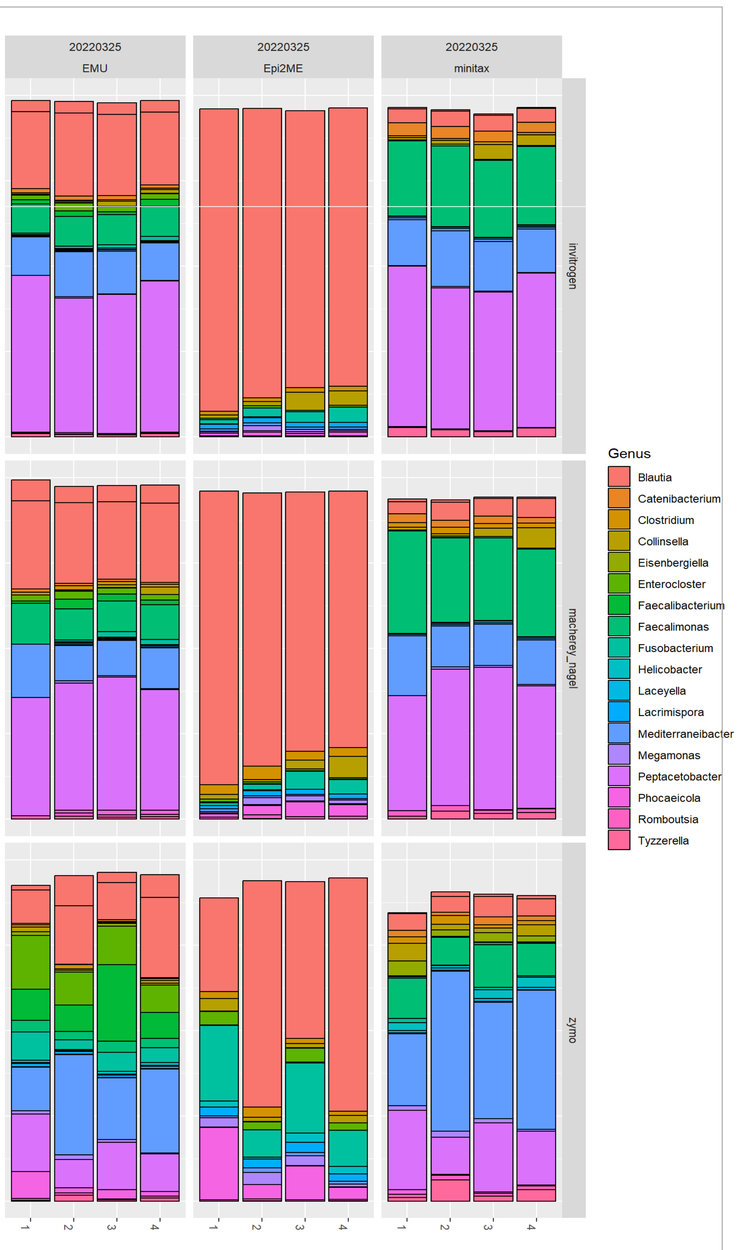




The most abundant 20 genus from ONT dataset by using EPI2ME 16S workflow. This analysis shows Blautia dominance within the sample which is inconsistent with the other approaches. Comparison of the results with other bioinformatics is shown below.

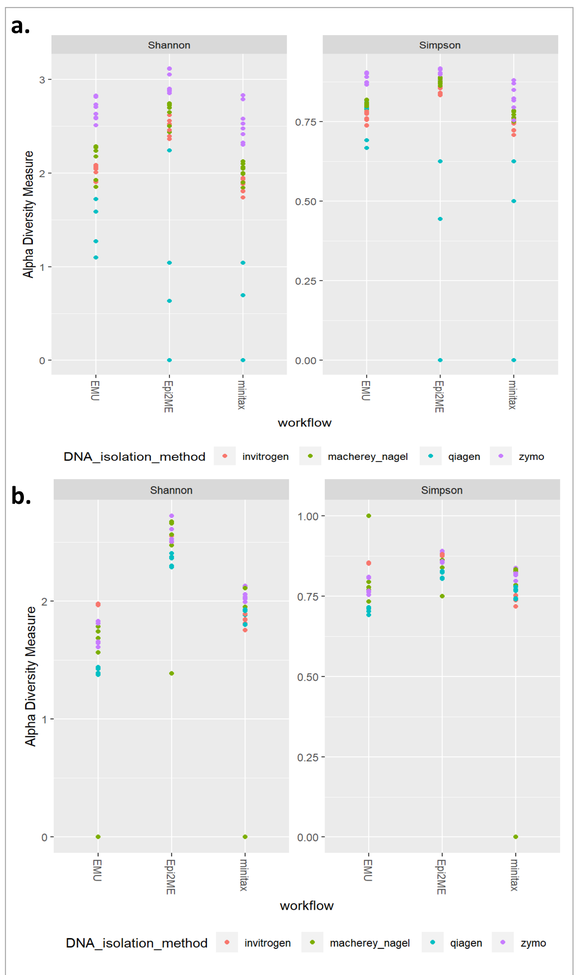


In this project the following three Variable regions: V1-V3, V1-V2 and V3-V4; along with V1-V9 via ONT and PacBIO; and moreover Illumina WGS sequencing methods prepared with four different DNA isolation methods: Invitrogen (I), Macherey-Nagel (MN), Qiagen (Q) and Zymo (Z) were compared. The bioinformatic workflow was carried out using **MiniTax**, which uses minimap2 to map the reads to the proGenomes database dada2* software's pipeline to: filter, correct errors, dereplicate and merge the .fastq files and subsequently to taxonomically annotate the resulting unique sequences (ASVs) using the SILVA 16S database
In this analysis the ASVs were combined at the species level after the diversity estimations. The figure shows the results at genus level (top 20). Results, obtained by the various sequencing strategies showed quantitative and qualitative differences at genus level.



Comparison of EMU, Epi2Me and MiniTax tools for the Nanopore analysis of V1–V9 regions of the 16S rRNA gene.

a.



Figures illustrating alpha diversity indices (Shannon diversity and Simpson index) in bacterial microbiomes of canine gut (a) and Microbial community mix (b) samples analyzed by three different bioinformatics tool.