Optimizing Microbiome Research: A Comparative Study of Laboratory Techniques and Bioinformatics Pipelines

PhD Thesis

Gábor Gulyás

Supervisor
Dr. Dóra Tombácz, PhD

Doctoral School of Experimental and Preventive Medicine
Institute of Medical Biology
Albert Szent-Györgyi Medical School
University of Szeged

Szeged

- 2025 -

Tabel of contents

List of publications

Scientific papers included in the thesis

1. **Gábor Gulyás** ; Balázs Kakuk* ; Ákos Dörmő* ; Tamás Járay ; István Prazsák ; Zsolt Csabai ; Miksa Máté Henkrich ; Zsolt Boldogkői ; Dóra Tombácz
   Cross-comparison of gut metagenomic profiling strategies
   COMMUNICATIONS BIOLOGY 7 : 1 Paper: 1445 , 22 p. (2024)
   Folyóirat szakterülete: Scopus - Agricultural and Biological Sciences (miscellaneous)
   SJR indikátor: D1
   Nyilvános idéző összesen: 5, Független: 5, Függő: 0, Nem jelölt: 0

Publications not related to the thesis

1. Dóra Tombácz*; Zoltán Maróti* ; Péter Oláh* ; Ákos Dörmő ; **Gábor Gulyás** ; Tibor Kalmár ; Zsolt Csabai ; Zsolt Boldogkői
   Temporal transcriptional profiling of host cells infected by a veterinary alphaherpesvirus using nanopore sequencing
   SCIENTIFIC REPORTS 15 : 1 Paper: 3247 , 12 p. (2025)
   Scimago Journal Rank indicator: Q1
2. Dóra Tombácz ; Balázs Kakuk ; Gábor Torma ; Ádám Fülöp ; Ákos Dörmő ; **Gábor Gulyás** ; Zsolt Csabai ; Zsolt Boldogkői✉
   Mapping the temporal transcriptomic signature of a viral pathogen through CAGE and nanopore sequencing
   PLOS ONE 20 : 4 Paper: e0320439 , 28 p. (2025)
   Scimago Journal Rank indicator: Q1
3. István Prazsák ; Dóra Tombácz* ; Ádám Fülöp* ; Gábor Torma ; **Gábor Gulyás** ; Ákos Dörmő ; Balázs Kakuk ; Lauren Spires McKenzie ; Zsolt Toth ; Zsolt Boldogkői
   KSHV 3.0: A State-of-the-Art Annotation of the Kaposi's Sarcoma-Associated Herpesvirus Transcriptome Using Cross-Platform Sequencing
   MSYSTEMS 9 : 2 Paper: e01007-23 , 19 p. (2024)
   Scimago Journal Rank indicator: D1
   Nyilvános idéző összesen: 10, Független: 8, Függő: 2, Nem jelölt: 0
4. Gergely Ármin Nagy ; Dóra Tombácz ; István Prazsák ; Zsolt Csabai ; Ákos Dörmő ; **Gábor Gulyás** ; Gábor Kemenesi ; Gábor E. Tóth ; Jiří Holoubek ; Daniel Růžek ; Balázs Kakuk ; Zsolt Boldogkői
   Exploring the transcriptomic profile of human monkeypox virus via CAGE and native RNA sequencing approaches
   MSPHERE 9 : 9 Paper: e00356-24 , 22 p. (2024)
   Scimago Journal Rank indicator: Q1
5. Balázs Kakuk ; Ákos Dörmő ; Zsolt Csabai ; Gábor Kemenesi ; Jiří Holoubek ; Daniel Růžek ; István Prazsák ; Virág Éva Dani ; Béla Dénes ; Gábor Torma ; Ferenc Jakab ; Gábor E. Tóth ; Fanni V. Földes ; Brigitta Zana ; Zsófia Lanszki ; Ákos Harangozó ; Ádám Fülöp ; **Gábor Gulyás** ; Máté Mizik ; András Attila Kiss ; Dóra Tombácz ; Zsolt Boldogkői
   In-depth Temporal Transcriptome Profiling of Monkeypox and Host Cells using Nanopore Sequencing
   SCIENTIFIC DATA 10 : 1 Paper: 262 , 12 p. (2023)

Scimago Journal Rank indicator: D1
Nyilvános idéző összesen: 11, Független: 10, Függő: 1, Nem jelölt: 0

6. Dóra Tombácz * ; Gábor Torma* ; **Gábor Gulyás*** ; Ádám Fülöp ; Ákos Dörmő ; István Prazsák ; Zsolt Csabai ; Máté Mizik ; Ákos Hornyák ; Zoltán Zádori ; Balázs Kakuk ; Zsolt Boldogkői
Hybrid sequencing discloses unique aspects of the transcriptomic architecture in equid alphaherpesvirus 1
HELIYON 9 : 7 Paper: e17716 , 16 p. (2023)
Scimago Journal Rank indicator: Q1
Nyilvános idéző összesen: 4, Független: 1, Függő: 3, Nem jelölt: 0
7. Gábor Torma* ; Dóra Tombácz* ; Zsolt Csabai* ; Islam A. A. Almsarrhad ; Gergely Ármin Nagy ; Balázs Kakuk ; **Gábor Gulyás** ; Lauren Spires McKenzie ; Ishaan Gupta ; Ádám Fülöp ; Ákos Dörmő ; István Prazsák ; Máté Mizik ; Virág Éva Dani ; Viktor Csányi ; Ákos Harangozó ; Zoltán Zádori ; Zsolt Toth ; Zsolt Boldogkői
Identification of herpesvirus transcripts from genomic regions around the replication origins
SCIENTIFIC REPORTS 13 : 1 Paper: 16395 , 25 p. (2023)
Scimago Journal Rank indicator: D1
Nyilvános idéző összesen: 7, Független: 5, Függő: 2, Nem jelölt: 0
8. See-Chi Lee ; Nenavath Gopal Naik ; Dora Tombacz ; **Gabor Gulyas** ; Balazs Kakuk ; Zsolt Boldogkoi ; Kevin Hall ; Bernadett Papp ; Steeve Boulant ; Zsolt Toth
Hypoxia and HIF-1α promote lytic de novo KSHV infection
JOURNAL OF VIROLOGY 97 : 11 Paper: e0097223 , 18 p. (2023)
Scimago Journal Rank indicator: D1
Nyilvános idéző összesen: 4, Független: 3, Függő: 1, Nem jelölt: 0
9. Dóra Tombácz ; Balázs Kakuk* ; Gábor Torma ; Zsolt Csabai ; **Gábor Gulyás** ; Vivien Tamás ; Zoltán Zádori ; Victoria A. Jefferson ; Florencia Meyer ; Zsolt Boldogkői
In-Depth Temporal Transcriptome Profiling of an Alphaherpesvirus Using Nanopore Sequencing
VIRUSES 14 : 6 Paper: 1289 , 25 p. (2022)
Scimago Journal Rank indicator: Q1
Nyilvános idéző összesen: 7, Független: 3, Függő: 4, Nem jelölt: 0
10. István Prazsák ; Zsolt Csabai ; Gábor Torma ; Henrietta Papp ; Fanni Földes ; Gábor Kemenesi ; Ferenc Jakab ; **Gábor Gulyás** ; Ádám Fülöp ; Klára Megyeri ; Béla Dénes ; Zsolt Boldogkői ; Dóra Tombácz
Transcriptome dataset of six human pathogen RNA viruses generated by nanopore sequencing
DATA IN BRIEF 43 Paper: 108386 , 11 p. (2022)
Scimago Journal Rank indicator: Q2
Nyilvános idéző összesen: 2, Független: 2, Függő: 0, Nem jelölt: 0
11. Dóra Tombácz ; Ákos Dörmő ; **Gábor Gulyás** ; Zsolt Csabai ; István Prazsák ; Balázs Kakuk ; Ákos Harangozó ; István Jankovics ; Béla Dénes ; Zsolt Boldogkői
High temporal resolution Nanopore sequencing dataset of SARS-CoV-2 and host cell RNAs
GIGASCIENCE 11 Paper: giac094 , 11 p. (2022)

Scimago Journal Rank indicator: D1
Nyilvános idéző összesen: 3, Független: 2, Függő: 1, Nem jelölt: 0

12. Dóra Tombácz ; Norbert Moldován ; Gábor Torma ; Tibor Nagy ; Ákos Hornyák ; Zsolt Csabai ; **Gábor Gulyás** ; Miklós Boldogkői ; Victoria A. Jefferson ; Zoltán Zádori ; Florencia Meyer ; Zsolt Boldogkői
Dynamic Transcriptome Sequencing of Bovine Alphaherpesvirus Type 1 and Host Cells Carried Out by a Multi-Technique Approach
FRONTIERS IN GENETICS 12 Paper: 619056 , 8 p. (2021)
Scimago Journal Rank indicator: Q2
Nyilvános idéző összesen: 5, Független: 1, Függő: 4, Nem jelölt: 0
13. Zoltán Maróti ; Dóra Tombácz ; Norbert Moldován ; Gábor Torma ; Victoria A. Jefferson ; Zsolt Csabai ; **Gábor Gulyás** ; Ákos Dörmő ; Miklós Boldogkői ; Tibor Kalmár ; Florencia Meyer ; Zsolt Boldogkői
Time course profiling of host cell response to herpesvirus infection using nanopore and synthetic long-read transcriptome sequencing
SCIENTIFIC REPORTS 11 : 1 Paper: 14219 , 11 p. (2021)
Scimago Journal Rank indicator: D1
Nyilvános idéző összesen: 6, Független: 4, Függő: 2, Nem jelölt: 0
14. Dóra Tombácz ; Gábor Torma ; **Gábor Gulyás** ; Norbert Moldován ; Michael Snyder ; Zsolt Boldogkői
Meta-analytic approach for transcriptome profiling of herpes simplex virus type 1
SCIENTIFIC DATA 7 : 1 Paper: 223 , 11 p. (2020)
Scimago Journal Rank indicator: D1
Nyilvános idéző összesen: 7, Független: 0, Függő: 7, Nem jelölt: 0
15. Norbert Moldován ; Gábor Torma ; **Gábor Gulyás** ; Ákos Hornyák ; Zoltán Zádori ; Victoria A. Jefferson ; Zsolt Csabai ; Miklós Boldogkői ; Dóra Tombácz ; Florencia Meyer ; Zsolt Boldogkői
Time-course profiling of bovine alphaherpesvirus 1.1 transcriptome using multiplatform sequencing
SCIENTIFIC REPORTS 10 : 1 Paper: 20496 , 14 p. (2020)
Scimago Journal Rank indicator: D1
Nyilvános idéző összesen: 11, Független: 0, Függő: 11, Nem jelölt: 0
16. Dóra Tombácz ; Norbert Moldován ; Balázs Zsolt ; **Gábor Gulyás** ; Zsolt Csabai ; Miklós Boldogkői ; Michael Snyder ; Zsolt Boldogkői
Multiple Long-read Sequencing Survey of Herpes Simplex Virus Dynamic Transcriptome
FRONTIERS IN GENETICS 10 Paper: 834 , 20 p. (2019)
Scimago Journal Rank indicator: Q1
Nyilvános idéző összesen: 39, Független: 20, Függő: 19, Nem jelölt: 0

Abbreviations

I:	Invitrogen
MN:	Macherey-Nagel
Q:	Qiagen
PE:	PerkinElmer
Z:	Zymo Research
I WGS:	Illumina DNA Prep mWGS library
ONT V1-V9:	Oxford Nanopore Technologies V1-V9 amplicon library
PacBio V1-V9:	Pacific Biosciences V1-V9 amplicon library
V1-V3:	PerkinElmer V1-V3 amplicon library
V1-V2:	Zymo Research V1-V2 amplicon library
V3-V4:	Zymo Research V3-V4 amplicon library
LRS:	Long-Read Sequencing
mWGS:	Metagenomic Whole Genome Sequencing
ONT:	Oxford Nanopore Technologies
PacBio:	Pacific Biosciences
SRS:	Short-Read Sequencing
WGS:	Whole-Genome Sequencing
GMS:	ZymoBIOMICS Gut Microbiome Standard D6331
MCS:	ZymoBIOMICS Microbial Community Standard D6300
ANOVA:	Analysis of Variance
HSD:	Tukey's Honestly Significant Difference
MAPQ:	Mapping Quality
NMDS:	Non-Metric Multi-Dimensional Scaling
PCoA:	Principal Coordinate Analyses

PERMANOVA:	Permutational Analysis of Variance
PERMIDISP:	Permutational Analysis of Dispersion
db:	database
Gram-positive:	Gram+
Gram-negative:	Gram-
HMW:	High Molecular Weight
LCA:	Lowest Common Ancestor

1. Introduction

1.1. Microbiome

The roots of microbiome research can be traced back to the early 20th century, when scientists first recognized that large numbers of microorganisms, including bacteria, fungi, and viruses, inhabit different parts of the human body such as the gut, oral cavity, and skin.[1] Early culture-based methods revealed only a small fraction of microbial diversity, but these initial findings laid the foundation for modern microbiome research. Today it is known that the human microbiota contains approximately 150 times more genetic information than the entire human genome, and is therefore often referred to as the "second genome," contributing substantially to human physiology and health.[2]

Over the past two decades, explosive advances in metagenomics, as well as in molecular biology, genomics, and bioinformatics, have fundamentally transformed microbiome research. These new technologies have enabled comprehensive taxonomic and functional investigation of microbial communities without the need for cultivation.[3,4] High-throughput sequencing methods, such as Illumina, Pacific Biosciencees (PacBio), and Oxford Nanopore Technologies (ONT), have elevated the study of microbial ecosystems to a new level and provided deeper insights into their metabolic, immunological, and pathophysiological roles.[5,6]

Current research is increasingly shifting toward multi-omics approaches: the integration of metatranscriptomics, metaproteomics, and metabolomics allows for the exploration of dynamic interactions between microorganisms and their host. This paradigm shift is moving microbiome science from descriptive studies toward functional and mechanistic understanding.[7,8] The knowledge gained in this way is not only crucial for understanding human health but also opens broad applications in agriculture[9], biotechnology[10], and environmental sciences.[11]

1.2. Gut microbiome

The composition of the gut microbiome plays the most important role in maintaining health. The bacteria residing in the gut perform a wide range of functions, including participation in digestion, protection against pathogens, vitamin production, and stimulation of the immune system.[12] The gut microbiome is typically composed of six major phyla: *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria*, *Fusobacteria*, and *Verrucomicrobia*, among which *Firmicutes* and *Bacteroidetes* dominate.[13]

The human gastrointestinal tract is the largest microbial habitat in the body, hosting approximately 10^{18} microorganisms.[14] The balance of the gut microbiome is closely linked to health and disease. Several studies have demonstrated that the gut microbiota plays a key role in nutrient extraction and metabolic processes, as it possesses a wide repertoire of metabolic genes that provide enzymes and biochemical pathways for energy and nutrient harvest.[15] In addition, the microbiome is a crucial contributor to the biosynthesis of bioactive molecules such as vitamins, amino acids, and lipids.[16]

A healthy gut microbiome is stable, resilient, and in a symbiotic relationship with the host. A healthy community is characterized by diverse species composition, a rich gene pool, and a stable "core microbiome".[17] It is important to note, however, that the relative distribution of microorganisms is unique to each individual and can change over time within the same individual. The composition of the gut microbiota is influenced by age, environmental factors (e.g., medication use), and anatomical differences across different regions of the gastrointestinal tract.[17]

Previous studies have demonstrated a strong similarity between the human and canine gut microbiomes.[18] Dogs represent a genetically more homogeneous population, and their lifestyle and diet can be more easily controlled, making them excellent model organisms for microbiome research.[19]

1.1. DNA purification techniques

DNA isolation from fecal samples represents a critical initial step in microbiome studies. The aim is to extract large amounts of high-quality DNA that are suitable for subsequent sequencing and bioinformatic analyses. Over the past decades, numerous DNA isolation techniques have been developed, which differ substantially in their methodology and efficiency.

Previous studies have demonstrated that different commercially available DNA isolation kits yield variable results in terms of DNA quantity, purity, and the representation of microbial community composition.[20] One of the major sources of variation is whether the kit includes a mechanical cell lysis step (such as bead-beating). This step is particularly important for disrupting the cell walls of Gram-positive bacteria, thereby ensuring efficient DNA recovery from these taxa.[21,22]

These methodological differences greatly contribute to the low reproducibility of microbiome research and the limited comparability of results across studies. The low rate of reproducibility is one of the major challenges in biomedical research and is widely referred to as the

"reproducibility crisis".[23] Although many laboratory and bioinformatic approaches exist for processing metagenomic data, there is currently no universally accepted standard workflow that would ensure complete comparability of results. The lack of standardization remains one of the greatest obstacles to the consistent and reproducible evaluation of metagenomic studies.[20,24,25]

1.2. Sequencing platform

In metagenomic studies the selection of the appropriate sequencing platform is of fundamental importance. Short-read sequencing (SRS) has traditionally been the most frequently used method especially in the case of metagenomic whole genome sequencing (mWGS) which is also supported by numerous bioinformatic tools such as Kraken2 and sourmash.[26–30] In recent years however long-read sequencing (LRS) has been gaining an increasingly important role represented primarily by the platforms of Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio).[31–33]

1.2.1. Illumina sequencing

Illumina technology operates on the principle of sequencing by synthesis, the central element of which is the DNA-dependent DNA polymerase enzyme. [34,35] In this method, DNA fragments are first provided with adapter sequences, which make it possible to attach the fragments to the surface of a special glass slide (flow cell). The inner surface of the flow cell is densely coated with oligonucleotides that are complementary to the adapter sequences, so the DNA fragments can specifically bind to them.[36]

The fixed fragments are multiplied by bridge amplification. In this process the DNA molecules "bend over" to another oligonucleotide on the surface and serve as a template for the synthesis of a new strand. By repeating this cycle thousands of identical copies are created from each original fragment, and well separated DNA clusters are formed, which can also be detected visually. The cluster-based strategy has the advantage that individual sequencing errors can be filtered out by consensus analysis of the identical copies.[37,38]

During the actual sequencing the polymerase enzyme incorporates fluorescently labeled reversible terminator nucleotides into the growing DNA strand. Each of the four nucleotides carries a different color fluorescent label. Since the nucleotides have a terminating group at the 3' end only one nucleotide can be incorporated at a time. At the moment of incorporation the detection system senses the emitted light upon laser excitation and records which base has been added to the DNA strand. Thus in every cycle only a single base is read, in complete synchrony across all clusters.[34,39]

At the end of the sequencing cycle the terminating group and the fluorescent label are removed through a chemical reaction, allowing DNA synthesis to continue with the next nucleotide. This "reversible termination" ensures the high accuracy of the process because after the identification of each base the cycle restarts. The detected fluorescent signals are collected as high resolution images and are then converted into nucleotide sequences by software.[35,39]

One of the greatest advantages of Illumina technology is the extremely high level of parallelization: in a single run billions of short DNA fragments can be read simultaneously. This makes the method particularly effective in the analysis of highly complex samples such as whole genomes, metagenomes or transcriptomes.[38,40] The short read length (typically 50–300 base pairs) however limits the method in the investigation of genomic regions containing many repetitive sequences or complex structural variations.[40,41]

The reduction of errors is ensured by the redundancy of the clusters and precise chemical control, which is why Illumina sequencing is today the most widely used second generation sequencing technology, especially when high accuracy and enormous amounts of data are required in a relatively short time.[34,37]

1.2.2. Single-molecule real-time (SMRT) sequencing

The SMRT technology developed by Pacific Biosciences is based on the mechanism of DNA replication, whose central enzyme is the DNA-dependent DNA polymerase.[42,43] During the operation of this enzyme, the reading of the DNA segment to be sequenced takes place with the help of fluorescently labeled nucleotides. When a labeled nucleotide is incorporated into the forming DNA strand, the fluorescent label molecule is released from it because during synthesis a phosphate group is cleaved from the nucleotide which at the same time provides the energy requirement of the reaction. Upon release of the label molecule a short millisecond-long light flash is generated which is detected by the detector.[42,44]

The essence of the method is that the process takes place in tiny chambers called Zero Mode Waveguide (ZMW) cells.[43,45] In these chambers the incorporation of a single nucleotide can be observed in real time while the background signal can be suppressed since at any given moment only a single nucleotide is present. Thanks to this sequencing is faster and more accurate than in the case of second-generation technologies. In the ZMW chamber a modified DNA polymerase enzyme is immobilized which attaches to the fragment to be sequenced with the help of an adapter sequence ligated to the ends of the cDNA strands. If the adapter sequence is missing from the end of the cDNA the polymerase cannot bind therefore no information is

generated about the nucleotide sequence of the given fragment.[43,46] Each of the four types of nucleotides is labeled with a different fluorescent molecule which ensures their unambiguous distinction.[42,44]

For the rapid and accurate operation of the polymerase a large number of free nucleotides is required. However, this causes significant background fluorescence. Earlier technologies solved this problem by inserting a washing step after nucleotide incorporation but before detection with which the background noise could be reduced.[47] In the case of SMRT the extremely small size of the ZMW chambers allows the background signal to be separated directly during measurement. Therefore there is no need for washing steps which significantly accelerates sequencing.[43,47]

One of the greatest advantages of the PacBio SMRT technology is the long read length typically 10–20 kb but with newer developments reads longer than 30 kb can also be achieved.[42,46] At the same time these reads initially showed a relatively high error rate which mainly resulted from insertions and deletions as well as electrophysiological noise.[42,44] To reduce the error rate the Circular Consensus Sequencing (CCS) approach was developed in which a consensus sequence is created from multiple readings of the same molecule.[43,46] With this method the newer generation HiFi reads already provide Illumina-like accuracy while preserving the advantages arising from the long read length.[46,48]

1.2.3. Nanopore sequencing

The nanopore-based sequencing developed by Oxford Nanopore Technologies (ONT) is based on measuring the direct physical properties of DNA molecules.[49,50] The central element of the method is a biological nanopore (for example α-hemolysin or MspA porin) which is embedded into a synthetic membrane. When an electric potential is applied across the two sides of the membrane an ion current forms through the pore. As the DNA strand passes through the bases inhibit the ion current in different ways thus characteristic current fluctuations are generated from which the nucleotide sequence can be determined.[51,52]

To start sequencing special adapter sequences are ligated to the ends of the DNA fragments which also contain the motor protein. This motor protein regulates the speed at which the DNA passes through the pore so that the current signals corresponding to individual bases are sufficiently resolved.[53] One significant advantage of nanopore technology is that DNA can be read in real time and directly without the need for amplification or synthesis.[50,54]

The method is capable of producing extremely long reads typically 10–100 kb but under favorable conditions DNA molecules of several megabases in length can also be sequenced.[55] This makes it possible to accurately map complex genome regions repetitive sequences and structural variations. At the same time the error rate of the technology was initially relatively high mainly in the form of insertions and deletions.[51,52] Developments in recent years especially basecalling algorithms based on neural networks (for example Guppy) and consensus strategies have significantly improved accuracy bringing it increasingly closer to Illumina and PacBio HiFi level.[56]

A unique property of nanopore sequencing is that it can not only determine the DNA nucleotide sequence but also directly detect post-translational modifications such as DNA methylation since these modifications also cause characteristic changes in the ion current.[57] This ability makes nanopore sequencing particularly valuable in epigenetic studies.

Another special advantage is that the platform is capable of sequencing both DNA and RNA natively. In the latter case the RNA molecule itself passes directly through the nanopore so no reverse transcription or PCR is required which provides a unique opportunity to study RNA structure and modifications (for example methylations). This function is currently not offered by any other sequencing platform.[58]

Another advantage of the Oxford Nanopore platform is portability the MinION device is palm-sized and can be operated from a laptop via USB port thus enabling genomic studies even in field conditions.[54,55]

1.3. Comparison of Sequencing Platforms for Microbiome Research

The Illumina technology is the most widely used second-generation platform which provides short but extremely accurate reads. Due to its high throughput and low cost it is ideal for studies with a large number of samples as well as for metagenomic and 16S amplicon-based microbiome profiling.[26,40] Its drawback however is that the read length typically ranges between 50–300 base pairs which makes it difficult to study complex genome regions repetitive sequences and structural variants.[40]

The PacBio SMRT technology belongs to the third-generation methods and provides long reads (10–20 kb and recently >30 kb). Although the raw error rate is higher the circular consensus sequencing (HiFi reads) allows a significant increase in sequence accuracy thus the method can achieve Illumina-level accuracy.[46,48] The advantage of long reads is particularly evident in the

study of full-length 16S rRNA genes which enables taxonomic identification even at the species level in contrast to Illumina amplicon sequencing which is limited to short regions.[59]

The Oxford Nanopore Technologies (ONT) platform is also a third-generation method which stands out with the production of ultra-long reads even in the megabase range.[55] A particular feature of ONT is that it is capable of native DNA and RNA sequencing without the need for PCR amplification or reverse transcription.[58] This provides a unique opportunity for the direct investigation of RNA modifications and DNA methylation.[57] Although the error rate was previously a significant limiting factor the continuously improving basecalling algorithms and consensus-based analyses have significantly increased accuracy.[56] In addition ONT devices such as the MinION Mk1B and Mk1D provide portability thus the technology can be applied in field and clinical settings as well.[55]

The differences between the three platforms are particularly emphasized in microbiome research. With Illumina sequencing genus-level information can be obtained with high accuracy and cost efficiency however species-level resolution is limited due to the short read length.[26,40] With PacBio HiFi reads full-length 16S rRNA genes can be sequenced which allows species-level identification although the method is more expensive and has lower throughput.[59] The ONT technology provides the possibility to sequence full 16S genes and even whole genomes natively as well as to perform epigenetic and transcriptomic studies which is a unique advantage in microbiome research.[57,58] Comparative studies have shown that Illumina sequencing provides accurate but less detailed taxonomic profiles while PacBio and ONT technologies yield significantly better species-level resolution and detection of taxonomic diversity.[59–62]

1.4. Databases

The choice of an appropriate reference database is one of the most critical factors in metagenomic analyses, as it fundamentally determines the quality of taxonomic and functional annotation. For short-read sequencing data, widely used resources include NCBI RefSeq and GenBank, which serve as general genomic repositories.[63] For microbiome-focused studies, dedicated reference databases such as SILVA, Greengenes, and RDP provide curated collections of 16S rRNA gene sequences, and are therefore essential for amplicon-based community profiling.[64–66]

One of the most recent and widely adopted resources for taxonomic classification is the Genome Taxonomy Database (GTDB), which offers a standardized phylogeny-based framework for

bacteria and archaea, enabling more accurate and consistent species-level assignments compared with traditional NCBI taxonomy.[67]

For the annotation of shotgun metagenomic data, commonly used resources include the extensive pathogen and microbiome databases developed for Kraken2[28], as well as reference sets compatible with sourmash.[30] In addition, the MetaPhlAn and HUMAnN databases are widely applied, supporting species-level taxonomic profiling as well as functional annotation of metagenomes and metatranscriptomes.[68,69]

From a functional annotation perspective, KEGG[70], eggNOG[71], and UniProtKB[72] are of particular importance, as they allow the investigation of the metabolic potential and gene functions of microbial communities.

In recent years, databases focusing specifically on the human microbiome have gained increasing significance. For examples, the Human Microbiome Project (HMP)[4], Integrative Human Microbiome Project[73] and the Unified Human Gastrointestinal Genome (UHGG) collection[6] provide high-resolution reference genomes for studying the gut microbiome. These resources enable more accurate identification of species- and strain-level differences, which are critical in both clinical and ecological research.

2. Aims

Since the microbiota plays a crucial role in human health and actively contributes to a wide range of biological processes and disease development, [74] Such methods should enable the acquisition of data that most accurately reflect biological reality. In this context, the present study aims to address the following objectives:

1. Establishing a wet lab protocol that minimizes methodological bias in gut microbiome profiling
2. Assessing the efficiency and reproducibility of DNA extraction methods
3. Evaluating the performance of library preparation protocols
4. Investigating the accuracy and potential biases of data generated by different sequencing platforms (short read and long read)
5. Designing a bioinformatics pipeline that delivers the highest possible accuracy in microbiome analysis
6. Conducting a comparative evaluation of outputs from different annotation tools
7. Assessing the effectiveness of various reference databases in microbiome research
8. Quantifying the extent to which currently accepted standard methods deviate from biological reality
9. Developing a versatile annotation tool capable of processing both short read and long read sequencing datasets

3. Methods

We applied three different sample types to evaluate the performance of four different DNA isolation kits, using six library preparation methods across three sequencing platforms. The samples originated from dog feces and from two distinct microbial community mixtures, ZymoBIOMICS Microbial Community Standard (MCS) and ZymoBIOMICS Gut Microbiome Standard (GMS), which comprised eight and eighteen bacterial strains respectively.

3.1. Sample collection

A fecal sample was collected from a 13.5-year-old healthy male Pumi (a Hungarian purebred dog) within one minute after defecation, immediately frozen, and stored at –80 °C. As controls, fecal samples were obtained from six other healthy Pumis, including four puppies, a 7 year old female, and a 6.5 year old neutered male (Table 3). For each DNA isolation, a single 2.5 gram fecal sample was used, which was divided according to the requirements of each isolation kit, resulting in a total of 16 isolations with four technical replicates per kit. To avoid potential bias, the sample was not partitioned based on its internal distribution but rather randomly, and the allocation to the kits was also carried out randomly. Due to the small sample size, homogenization was not performed, which has been considered justified by Liang et al. (2020) [75], since although homogenization is essential for metabolomic analyses, it is not required in microbiome studies. For control purposes, random 200 mg subsamples were taken from six different dogs. In addition, to validate the results, we used the ZymoBIOMICS Microbial Community Standard (MCS, Zymo Research, D6300) and the ZymoBIOMICS Gut Microbiome Standard (GMS, Zymo Research, D6331) mixtures.

3.2. DNA purification and library preparation

In our study we tested the following commercially available DNA isolation kits, each with four technical replicates: QIAGEN QIAamp Fast DNA Stool Mini Kit (Ref. #51604, Lot. #169025369), Invitrogen PureLink™ Microbiome DNA Purification Kit (Cat. #423482), Macherey Nagel NucleoSpin DNA Stool Mini Kit (Ref. #740472.50, Lot. #2302 001), and Zymo Research Quick DNA™ HMW MagBead Kit (Cat. #D6060). For library preparation we applied the following kits: Illumina DNA Prep (Doc. #1000000025416 v09), Oxford Nanopore Technologies Rapid Sequencing 16S Barcoding Kit (SQK RAB204), Pacific Biosciences Full Length 16S Library Preparation Using SMRTbell Express Template Prep Kit 2.0 (PN 101 916 900), PerkinElmer NEXTFLEX® 16S V1 V3 Amplicon Seq Kit for Illumina (Cat. #4202 02), and Zymo Research Quick 16S™ NGS Library Prep Kit (Cat. #D6400). DNA isolation and library preparation were carried out according to the manufacturer protocols, which are

described in detail in the Supplementary Methods to ensure that the exact versions applied in this study remain available even if future protocol modifications are introduced.

3.3. DNA quantification and quality assessment

DNA yield was quantified using the Qubit 4.0 Fluorometer, while DNA quality including assessment of fragment length distribution was evaluated with the TapeStation 4150 system.

3.4. Statistical analysis

The data were analyzed using one way analysis of variance (ANOVA) to compare the DNA yields obtained from the four different kits. ANOVA tests were performed separately for each sample type (dog feces, MCS, and GMS). To evaluate the significance of differences between groups we applied the Tukey Honest Significant Difference (HSD) [76,77] post hoc test which allowed us to identify which pairs of kits differed significantly from each other.

3.5. Relative standard deviation (RSD)

The RSD was calculated to assess the internal variability of kit performance using the following formula RSD = (standard deviation / mean) × 100 %.

3.6. Sequencing

For SRS a total of nine MiSeq Reagent Kit v2 and four MiSeq Reagent Kit v2 Nano kits were used (Table 4). Sequencing of the V1–V9 region was carried out on ONT MinION and PacBio Sequel IIe platforms. ONT V1–V9 barcoded libraries prepared from dog samples were loaded onto three MinION flow cells while an additional three flow cells were used for sequencing the MCS and GMS samples. PacBio V1–V9 libraries were also barcoded and sequenced on a single Sequel SMRT Cell 8M run.

3.7. Bioinformatic analysis

3.7.1. Brief description of the minitax software

We developed *minitax*, a versatile taxonomic assignment tool designed to address the challenges posed by different types of sequencing data. *Minitax* enables taxonomic profiling across multiple sequencing platforms (ONT, PacBio, Illumina) and library types, including metagenomic whole genome sequencing (mWGS) and 16S rRNA gene sequencing. The tool employs minimap2[78] with platform specific parameter settings for initial read alignment against the reference database. Alignment results are imported into the R environment via Rsamtools[79] and integrated with database information. The workflow relies on the data.table[80] package for efficient handling of large datasets and subsequently performs post alignment processing steps to determine the best alignment for each read.

Post alignment processing involves several key steps. First, *minitax* applies a general MAPQ based filtering (for example for MAPQ values between 1 and 59) and then selects the alignment with the highest MAPQ score for each read. Next, it retains the alignment with the highest CIGAR score based on platform specific scoring matrices (Supplementary Data 8). After filtering, *minitax* determines the lowest common taxonomic level ("tax.identity") and its corresponding rank ("tax.identity.level") for each read. Additional refinement options are provided including the *BestAln* method which evaluates the proportion of alignments supporting a given taxonomic level. If a taxon is supported by the majority of alignments (or above a predefined default threshold of 60 percent) and is more specific than the current assignment the identification is updated accordingly. A simpler approach is offered by the *RandAln* method which randomly selects an alignment from the filtered set whereas the *SpeciesEstimate* method uses all valid alignments normalizing read counts by the number of alignments to estimate species level abundance. The *LCA* (Lowest Common Ancestor) method applies a conservative approach assigning each read to the lowest common ancestor of all matched taxa thus ensuring that assignments remain at the most specific taxonomic level supported by all alignments.

Finally *minitax* aggregates read counts at the chosen taxonomic rank and exports the results of each processing step in .tsv format while providing the final output as a *phyloseq*[81] object for downstream analysis. The tool is available on GitHub at https://github.com/Balays/minitax. CIGAR scoring schemes and other parameters can be user defined. In this study we applied the *RandAln* method although in many cases the *BestAln* method may yield more accurate results.

3.7.2. Analysis of the raw data

For Illumina V1–V2, V3–V, and V1–V3 regions the DADA2 pipeline[82] was used for quality control, filtering, trimming of Illumina amplicon reads, followed by dereplication, chimera detection and removal, and generation of ASVs (amplicon sequence variants), concluding with taxonomic assignment. Taxonomic classification was performed using either the SILVA 16S database[64] (version 138.1) or the Emu database (version 3.4.4). Exact parameters are provided in SUPPTAB_X (dada2_config.tsv) and in the full workflow available on GitHub (dada2.WF.R). Illumina reads were additionally processed with Emu [83](v3.4.4) and the in house developed *minitax* tool (v1.0) using parameters "--type sr" and "--N 10" with default settings otherwise. In both cases the default Emu database was applied, while *minitax* also included an NCBI genome collection. Both programs employ minimap2[78] for initial read mapping.

For ONT V1–V9 regions raw signal data from the MinION platform were basecalled with Guppy[84] version 6.1.5 (MinKNOW 20.05.8) using the high accuracy model. Reads were demultiplexed based on SQK RAB204 barcodes. During basecalling a minimum quality threshold of 8 was applied, separating pass and fail reads, and only pass reads were retained for downstream analysis. These pass reads were processed with Emu[83] (v3.4.4) and *minitax* (v1.0) using parameters "--type map-ont" and "--N 10" with default settings. Both workflows used the default Emu database, while *minitax* additionally incorporated an NCBI genome collection. The ONT EPI2ME pipeline[85] (v3.6.1) was also applied.

For PacBio V1–V9 regions basecalling, demultiplexing, and HiFi read generation were performed with PacBio SMRT Link[86] version 10.2.0.133434. High quality CCS reads were filtered with "--min-qv 20" and the resulting reads were analyzed with Emu[83] (v3.4.4) and *minitax* (v1.0) using parameters "--type map-pb" and "--N 10" under default settings. Both used the default Emu database, while *minitax* also employed an additional NCBI genome collection.

For Illumina WGS data raw reads were trimmed with Trim Galore[87] and host derived reads were removed with BMTagger[88] using the Canis lupus familiaris reference genome (GCF_014441545.1). Quality filtered reads were then processed either with the fast and sensitive taxonomic classifier *sourmash*[30] (v4.8.2) using the GenBank genome database (March 2022), or with *minitax* (v1.0) using default parameters and an NCBI genome collection.

3.7.3. Benchmarking datasets

In addition to the Zymo D6300 Microbial Community Mixture sequenced with the previously described ONT V1–V9 and Illumina V1–V2 methods we evaluated the performance of *minitax* on two further publicly available datasets. The Zymo MCM D6331 Microbial Community PacBio HiFi WGS dataset was used by Portik et al.[89] for benchmarking long read mWGS software. Fastq files were downloaded from NCBI (accession SRX9569057) and used as input for the *sourmash* and *minitax* programs. For both the D6300 and D6331 workflows all reads mapping to any *Veillonella* genus members were reassigned to *Veillonella rogosae* and all reads classified as belonging to the *Roseburia* genus were assigned to *Roseburia hominis*. Furthermore ten samples each from the CAMISIM simulated mouse gut dataset[90] were randomly selected for PacBio and Illumina platforms and used as input for the programs.

3.7.4. Databases

We utilized approximately 18 000 genomes representing around 13 000 bacterial archaeal and eukaryotic species. These genomes were downloaded from NCBI in February 2022 and were

used as reference databases for both WGS and 16S gene sequencing data. For amplicon sequencing datasets the choice of reference database depended on the software applied with either the Emu or the SILVA 16S database (version 138)[64] being used.

3.7.5. Taxonomy version

For the comparison of DNA isolation and library preparation techniques we used the version of NCBI that corresponded to the genomes included in the collection. In sections 2 and 3 of the results however we switched to the most recent NCBI release from July 2024. This update was not applied in other sections because the databases used for comparison with *minitax* results were based on an earlier version of the taxonomy.

3.7.6. CAMISIM

For the comparison of different DNA isolation and library preparation techniques we used the version of NCBI that corresponded to the genomes included in the collection. In sections 2 and 3 of the results however we switched to the most recent NCBI release from July 2024. This update was not applied in other sections because the databases used for comparison with *minitax* results were based on an earlier version of the taxonomy.

3.7.7. Statistics and downstream data analysis

The outputs from the different programs were organized into *phyloseq* objects which were then merged into a single comprehensive *phyloseq* object containing all metagenomic read count data. Subsequent analyses were carried out in the R environment using this *phyloseq* object together with the *tidyverse FactoMineR* and *vegan* packages. All scripts used for downstream data processing and figure generation are available in the GitHub repository https://github.com/Balays/Microbiome-Method-Comparison.

PERMANOVA (Permutational Multivariate Analysis of Variance): Significant differences in microbial composition between groups were tested with a full model PERMANOVA using the formula adonis2(otutab_t ~ DNA_isolation_method + library, data=sampdat, method="bray"). This model evaluated the combined effects of DNA isolation method and library preparation on microbial community structure with 999 permutations.

PERMDISP2 (Permutational Analysis of Multivariate Dispersion): Differences in within group variability were assessed with the betadisper function applied to the Bray–Curtis dissimilarity matrix using betadisper(distance_matrix, groups).

PCoA (Principal Coordinates Analysis): PCoA was carried out on the Bray–Curtis dissimilarity matrix separately for each DNA isolation method using the output of betadisper to evaluate the effect of library preparation within each method. Distances to centroids were calculated and the results were combined across isolation methods for visualization (Fig. 5b).

NMDS (Non-metric Multidimensional Scaling): To further explore relationships between samples NMDS was performed on normalized abundance data using Bray–Curtis dissimilarity with the command ordinate(ps.prop, method="NMDS", distance="bray"). NMDS was used to visualize similarities between samples with the relative closeness of points reflecting the similarity of microbial community compositions (Fig. 5a).

3.7.8. Performance metrics and statistics

To comprehensively assess the performance of *minitax* we calculated precision recall F1 and F0.5 scores at species detection thresholds of 1 percent 0.1 percent and 0.01 percent following the methodology of Portik et al.[89] In addition chi square tests were performed to determine whether significant differences existed between theoretical and observed distributions and Pearson correlations were computed between theoretical and observed community compositions with corresponding r^2 values at all taxonomic levels.

We also computed the Chi-squared statistic for each taxon using the formula:

$$\chi^2 = \frac{(Observed - Expected)^2}{Expected}$$

which was then summed across all taxa to assess the overall goodness-of-fit between the theoretical and observed distributions. The Chi-squared test was used to determine whether the observed distribution significantly deviated from the theoretical distribution. A Bonferroni correction was applied to account for multiple comparisons, and the null hypothesis was rejected if the corrected *p*-value was below the significance threshold (0.05).

In addition, we calculated Pearson's correlation coefficients (r^2 values) to evaluate the linear relationship between the theoretical and observed microbial abundances.

We also identified significantly different taxa (between the expected and observed compositions) using DESeq2[91] with the following commands:

```
de.seq <- phyloseq_to_deseq2(ps, ~DNA_isolation_method)
de.seq <- DESeq(de.seq, test = "Wald", fitType = "parametric", sfType = "poscounts")
```

Taxa with a *p*-value < 0.05 and a log2Fold difference of ≥ 2 or ≤ −2 were considered significantly different.

3.7.9. Downsampling

Raw Illumina mWGS reads were randomly downsampled to the following read counts 2 500 000, 2 000 000, 1 500 000, 1 000 000, 750 000, 500 000, 250 000, 225 000, 200 000, 175 000, 150 000, 125 000, 100 000, 75 000, 50 000, 20 000, 10 000, and 5 000. These subsets were subsequently analyzed with Kaiju[92] (version 1.9.0) using the progenomes (v2) database for taxonomic classification. Similarly, raw reads from non-mWGS datasets were downsampled to the same set of read counts, and processed with Emu[83] (version 3.4.4) under default settings.

Shannon index values were calculated from the resulting outputs using the *phyloseq*[81] R package by analyzing each taxonomic abundance profile within *phyloseq*[81], and the results were visualized with *ggplot2*[93] (https://github.com/gabor-gulyas/Technical-article-downsample).

4. Result

4.1. Study design

The goal of this study was to examine how different protocols influence the outcomes related to gut microbiome composition. These protocols include DNA extraction, library preparation, sequencing, and bioinformatics methods (Figs. 1, 2). We analyzed the amount, quality, and reproducibility of DNA extraction using four different isolation kits from Qiagen (Q), Macherey-Nagel (MN), Invitrogen (I), and Zymo Research (Z) (Fig. 3 and Supplementary Data 2a). We further investigated the microbial community by assessing dominant taxa and overall diversity. For mWGS, libraries were prepared with the Illumina DNA Prep Kit (I WGS), while amplicon libraries targeting the V1-V3 regions [PerkinElmer (PE V1-V3)], V1-V2, and V3-V4 [both from Zymo Research (Z V1-V2 and Z V3-V4, respectively)] of the 16S rRNA gene were also created (Supplementary Data 2b). In addition, we generated V1-V9 libraries covering the entire 16S rRNA gene for sequencing on two long-read sequencing (LRS) platforms: ONT MinION (ONT V1-V9) and PacBio Sequel IIe (PacBio V1-V9). All libraries were assessed for quality, yield, and reproducibility.

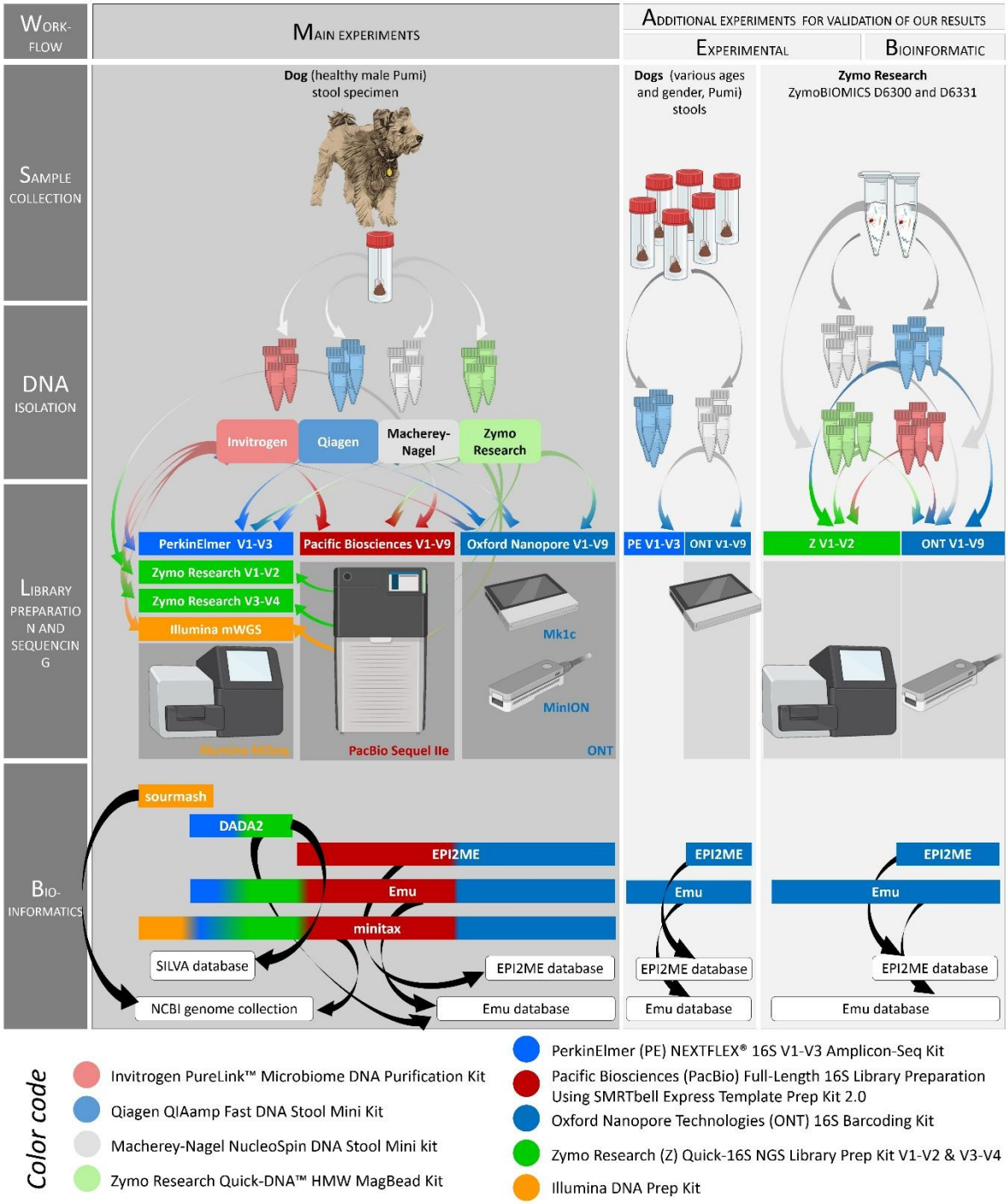


Figure 1: Workflow of the experimental part of the study.

The figure provides a detailed representation of the workflows conducted in this study. The figure illustrates the workflows conducted in this study, which included: 1) evaluating the efficacy of various DNA isolation kits in terms of quality, quantity, and microbial representation from canine stool samples; 2) comparing library preparation techniques on SRS and LRS platforms for reproducibility; 3) introducing "minitax", a tool designed to ensure consistent analysis across multiple sequencing platforms; 4) assessing the influence of different

databases and tools on microbial profiling; and 5) comparing 16S V1-V9 sequencing on ONT and PacBio platforms to address literature gaps and emphasize bioinformatics workflows. Our goal was to identify reliable procedures for robust and reproducible gut microbiome profiling across both wet-lab and dry-lab methodologies. We performed additional experiments to validate the most extreme experimental and bioinformatics results, particularly focusing on methods that yielded the most inconsistent outcomes in comparison to other techniques. For this purpose, we utilized samples from six additional dogs of various genders and ages. The workflow involved: 1) DNA isolation using the Q kit, 2) Library preparation with the PerkinElmer V1-V3 kit, and 3) Analysis of V1-V9 libraries using the EPI2ME software. Furthermore, we carried out experiments employing a Microbial Community Standard (MCS; Zymo Research D6300) as well as a Gut Microbiome Standard (GMS; Zymo Research D6331) to validate the effectiveness of the four DNA isolation kits used. This included: 1) DNA isolation using the kits applied for the dog samples, 2) Preparation of V1-V2 and V1-V9 libraries, and 3) Sequencing on the corresponding Illumina and ONT platforms based on the library. Created in BioRender. BioRender.com/k32q619.

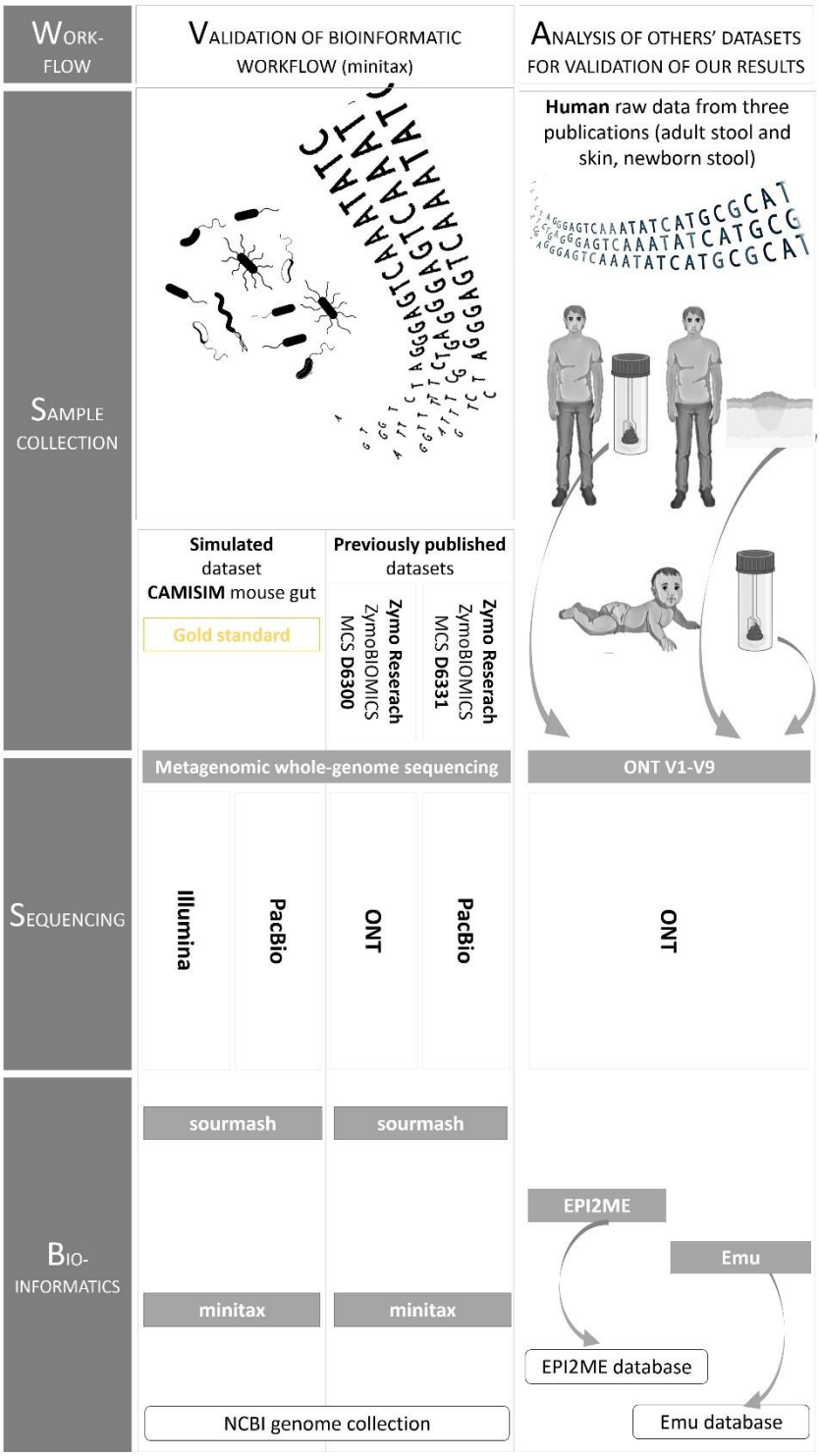


Figure 2: Workflow of the in silico analyses. Additionally, we conducted an in silico analysis to validate our bioinformatic tool, minitax. We compared the performance of minitax and sourmash by utilizing them for the analysis of: 1) simulated PacBio and Illumina data[90], and 2) previously published datasets[89]. For further validation of our results, we also used previously published metagenomics data from human sources encompassing skin[94] as well as fecal samples from newborns[95] and adults[96]. Created in BioRender. BioRender.com/e57v119.

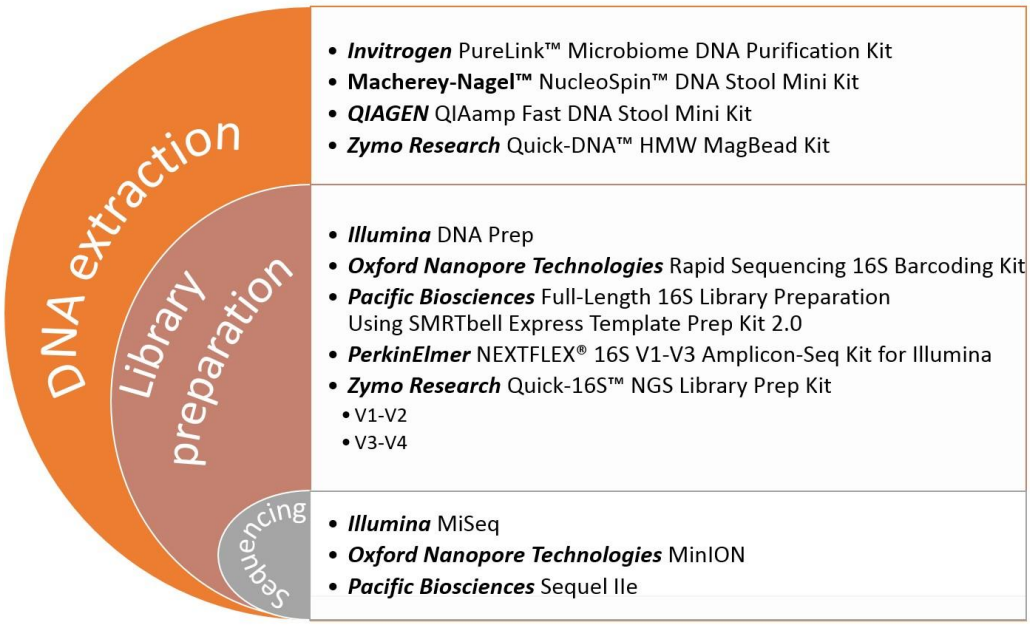


Figure 3: The kits and sequencers utilized in this project. This figure lists the DNA extraction and library preparation kits, along with sequencing devices which were used in this experiment.

The bioinformatics workflows included DADA244 for amplicon-based short-read sequencing (SRS) datasets, sourmash[30] for WGS samples, and Emu[83], a recently developed high-accuracy tool optimized for LRS 16S rRNA-Seq46. For nanopore sequencing data, we also used the vendor-specific EPI2ME[85,97] pipeline. Furthermore, we created a flexible and broadly applicable program named "minitax," designed to process multiple data types and provide reliable taxonomic assignment across metagenomic datasets.

After sequencing reads were aligned to either a reference genome database (default option) or a 16S database using minimap2[78], minitax selected the best alignment and assigned the most likely taxonomy for each read based on mapping quality (MAPQ values) and CIGAR strings (Supplementary Data 2c).

To examine how different library preparation methods affect the accuracy of microbial composition, we standardized DNA extraction using the same isolation kit to minimize inconsistencies. We compared and validated the results by assessing various stages of the workflow with multiple sample types and datasets (Figs. 1, 2).

4.2. Comparison of DNA preparation techniques

Most DNA isolation kits rely on affinity-based purification, inhibitor removal solutions or columns, and lysis buffers, enzymes, or bead-beating for breaking cell walls. While bead-beating is frequently recommended in previous studies[98,99], many commercial kits do not include this step. We evaluated four commercial kits that differed in these features (Supplementary Data 2a). Among the four methods tested (Fig. 3), the Z approach required the most hands-on effort, setting it apart as the most labor-intensive option (Supplementary Fig. 1). The other three methods were more similar in this aspect. The kits used in this work showed notable differences in both yield and quality of extracted DNA (Supplementary Data 3). These differences were visible in the DNA amount (Fig. 4a, b, Supplementary Fig. 2), the ratio of microbial to host DNA (Fig. 4c, d), and reproducibility (Fig. 4e, f).

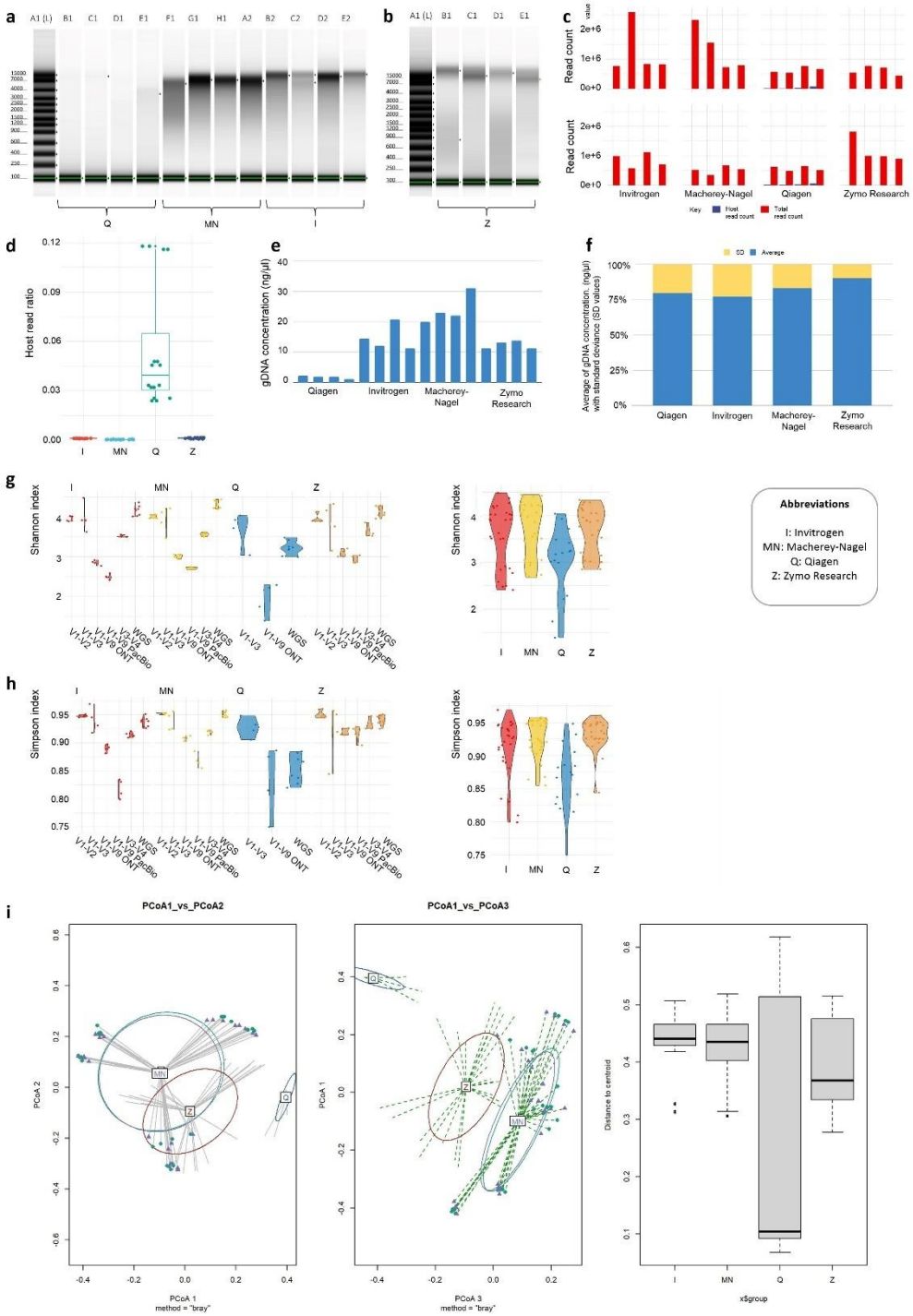


Figure 4: Impact of DNA isolation kit on yield, quality, and bacterial composition analysis. Quality of the Isolated DNA; molecular weight of DNA **a**: lanes B1, E1, F1, A2 and B2, D2: four replicates of DNA samples were isolated using Q, MN and I kit, respectively. **b** lanes B1, E1: DNA samples extracted by Z kit. **c, d:** ratio of host DNA. **e, f:** Yield and reproducibility of the isolated DNA. **g–i:** Microbial composition, diversity and dispersion; Alpha-diversities in the different library preparation methods, according to each DNA isolation method: **g:** Shannon index; **h:** Simpson index; Beta-diversity analysis of each DNA isolation method: **i:** PCoA plots and distances to group centroids.

We found marked differences in DNA degradation across the kits, with the Q kit producing the most degraded samples (Fig. 4a, lanes B1, E1). To obtain DNA of sufficient integrity for full-length 16S rRNA sequencing with ONT, we applied the Quick-DNA High Molecular Weight (HMW) MagBead Kit from Zymo Research. Although longer DNA fragments indicate lower degradation, our findings suggest that HMW DNA is not strictly required for sequencing the complete V1-V9 region. DNA quality and yield from the MN (Fig. 4a, lanes F1, A2) and I (Fig. 4a, lanes B2, E2) kits were also appropriate for LRS, as the differences in average fragment length across these three kits had little impact on V1-V9 sequencing (Fig. 4a, lanes F1, E2 and Fig. 4b). However, for LRS-based WGS, maintaining HMW DNA was critical. Based on fragment length, the Z kit proved most suitable for this purpose (Supplementary Data 3). We further assessed each kit's ability to selectively extract bacterial DNA. Even when using the Q kit with its "Isolation of DNA from Stool for Pathogen Detection" protocol, host DNA contamination remained significantly higher than with the other kits (Fig. 4c, d).

For canine stool samples, all kits except the Q kit provided DNA of sufficient quality and quantity for sequencing (Fig. 4e, f and Supplementary Data 3a). Analysis of the Q kit on six different dog samples consistently produced low yields and poor quality (Supplementary Data 3b), whereas the other three kits showed reproducible performance across replicates. Among them, the Z kit displayed the greatest consistency, while the I kit had the highest variability (Fig. 4e, f). Substantial yield differences were observed across kits (Supplementary Data 3). For canine feces, the Q kit gave the lowest DNA yield, the I kit produced moderate amounts, and the MN kit delivered the highest yield. The Z kit achieved relatively high yields despite using only half the starting sample size (Table 1 and Fig. 4e, f). An ANOVA revealed strong differences among the kits with an F-value of 511.63 and a p-value below 0.0001. Pairwise comparisons indicated that the I kit yielded more DNA than Q, the MN kit outperformed both, and the Z kit also gave higher yields than Q, although MN still provided the best overall yield. Adjusting for input volume, the Z kit would likely match MN, suggesting the two are comparable when equal sample sizes are used.

The average of DNA concentration isolated from dog stool	Average gDNA cc. (ng/µl)
Qiagen	1.815 ± 0.464
Invitrogen	14.6 ± 4.35
Macherey-Nagel	24.0 ± 4.83
Zymo Research	12.4 ± 1.36 ng/µL
The average of DNA concentration isolated from the MCS sample	**Average gDNA cc. (ng/µl)**
Qiagen	0.364 ± 0.12
Invitrogen	0.287 ± 0.048
Macherey-Nagel	10.564 ± 0.91
Zymo Research	38.72 ± 3.33
The average of DNA concentration isolated from the GMS sample	**Average gDNA cc. (ng/µl)**
Qiagen	1.146 ± 0.157
Invitrogen	0.0683 ± 0.0076
Macherey-Nagel	12.25 ± 0.636
Zymo Research	4.087 ± 0.31
The average of the longest DNA fragments isolated from dog stool	**Average length of DNA (bp)**
Qiagen	9361
Invitrogen	26,341
Macherey-Nagel	21,325
Zymo Research	33,867
The average of the longest DNA fragments isolated from the MCS sample	**Average length of DNA (bp)**
Qiagen	>60,000
Invitrogen	>60,000
Macherey-Nagel	54,74
Zymo Research	>60000
The average of the longest DNA fragments isolated from the GMS sample	**Average length of DNA (bp)**
Qiagen	>60,000
Invitrogen	58,897
Macherey-Nagel	55,538
Zymo Research	>60,000

Table 1: Average concentrations and lengths of obtained genomic DNA samples The first sections present the average concentrations of genomic DNA (gDNA) obtained from dog stool, MCS, and GMS samples, measured in ng/µL, respectively. Values are reported as means with standard deviations. The last blocks of tables report the average length of the longest DNA fragments (in base pairs, bp) obtained from dog stool, MCS, and GMS samples, respectively.

In canine fecal samples, the Z method generated the longest average fragment peaks, followed by I with moderately long fragments, MN with slightly shorter peaks, and Q with the shortest fragments (Supplementary Data 3 and Table 1). A Tukey HSD test confirmed significant differences in fragment length among several kit pairs (Supplementary Data 3c). Supplementary Fig. 3a illustrates these differences, while Supplementary Fig. 3b shows violin plots of fragment length distributions across kits with colors reflecting concentration.

To determine how much the DNA extraction kit influenced microbial composition, we analyzed data using the same database and bioinformatics workflow. Because commonly used pipelines are optimized for specific sequencing methods, we applied the minitax tool with an NCBI genome collection as the reference.

Earlier studies suggested that α-diversity indices can reflect DNA extraction efficiency[20]. In our study, the Q kit consistently led to a strong reduction in both richness (Shannon index) and evenness (Simpson index) compared to the other kits (ANOVA, p-values < 0.000 for all comparisons) (Fig. 4g, h). Other pairwise richness differences were not significant, although the I kit showed reduced evenness relative to Z (Supplementary Data 4). We also identified read number thresholds for each sequencing approach at which α-diversity values remained stable without significant drop (Supplementary Fig. 4).

For β-diversity, we calculated a Bray-Curtis distance matrix at the sample level. From these distances, NMDS, PERMDISP, and PERMANOVA analyses were performed. These revealed clear compositional differences depending on the DNA extraction method. The full-model PERMANOVA indicated that DNA extraction accounted for 28.2% of the variance in microbial community composition across sequencing methods and platforms ($p < 0.001$).

We further examined the dispersion of samples for each extraction method, calculated as the average distance to the centroid in multivariate space (PERMDISP2). These values showed notable variability that can strongly influence conclusions about microbial composition. The results are visualized in PCoA plots for each extraction method (Fig. 4i). These plots revealed strong overlap between I and MN, partial overlap of Z with these two, and Q clustering apart from all others.

When analyzed separately for each library, the I kit consistently produced the lowest dispersion values with small standard deviations, showing a mean distance of 0.0626 and a standard deviation of 0.0586 (Supplementary Fig. 5a). The MN kit also gave low dispersion across nearly all libraries except PE V1-V3. The Z kit produced higher dispersions in LRS datasets but results

comparable to I and MN in SRS libraries. The Q kit showed variability comparable to the others in the two libraries suitable for deeper analysis (PE V1-V3 and I WGS). PERMANOVA tests for each library type confirmed that DNA extraction methods influenced microbial composition across both V-regions and WGS. The MN and I kits consistently yielded similar community profiles across protocols, while Z and Q diverged significantly from both MN and I, and often from each other. These trends are represented in the library-specific PCoA plots based on PERMDISP values (Supplementary Fig. 5b). Pairwise tests confirmed these differences across most V-regions, with the I/M pair being the only exception.

4.3. Assessment of library construction and sequencing methods

In this part of the study, our aim was to assess the consistency of variability among six different library preparation strategies. To do this, we compared the outcomes of each DNA isolation method (I, MN, Q, and Z) across the available sequencing libraries. Because the Q kit presented quality and quantity issues, only the WGS, PerkinElmer (PE) V1-V3, and ONT V1-V9 libraries were prepared from those samples. As most commonly applied programs are designed specifically for either amplicon sequencing or WGS, we used the same tool (minitax) across all sequencing methods, with an NCBI genome collection serving as the reference database.

4.3.1. Quality, yield, and reproducibility

Among the tested methods, the ONT 16S rRNA amplicon library required the least experimental input, while the Illumina WGS and PacBio 16S rRNA amplicon workflows required the most manual effort (Supplementary Fig. 6). With the exception of the PE method, which frequently produced two non-specific DNA fragments (400 and 800 bp), the other approaches showed excellent performance in quality, yield, and reproducibility for canine samples (Supplementary Fig. 7). However, read quality with Nanopore sequencing was somewhat lower (Supplementary Data 5).

4.3.2. Microbial composition, diversity and dispersion

Analysis of α-diversity revealed that in the short-read sequencing (SRS) libraries, richness and evenness values were similar across the three kits, except for Q. In the long-read sequencing (LRS) libraries, the I method produced somewhat lower diversity, the Z method the highest, and MN also showed comparably high values. For β-diversity, we calculated Bray-Curtis distances at the sample level. The NMDS plot in Fig. 5a shows that samples tended to cluster according to library preparation rather than DNA extraction method. Samples prepared with the I and MN kits displayed broader spread, suggesting that community composition varied with library preparation. The V3-V4 group formed a cluster clearly separated from V1-V3, which

was more closely related to V1-V2. The Z method produced somewhat tighter groupings, while the Q kit's WGS and V1-V3 libraries clustered closely together. Taken together, both extraction and sequencing approaches influenced community composition, with sequencing protocols being the stronger factor. A full-model PERMANOVA confirmed this, showing that library type was the dominant variable, explaining 58.8% of the total variance. We also performed PERMDISP for each group (Fig. 5b). PCoA visualization showed that none of the extraction methods produced identical microbial profiles across all libraries. Noticeable overlaps were only seen in V1–V2 and V1–V3 libraries prepared using I and MN (Fig. 5b–d). LRS and WGS libraries created from Z DNA yielded results that were broadly similar to those from I and MN. Variability within each library type differed, with the tightest clustering observed in the V3–V4 libraries of I samples, pointing to higher consistency.

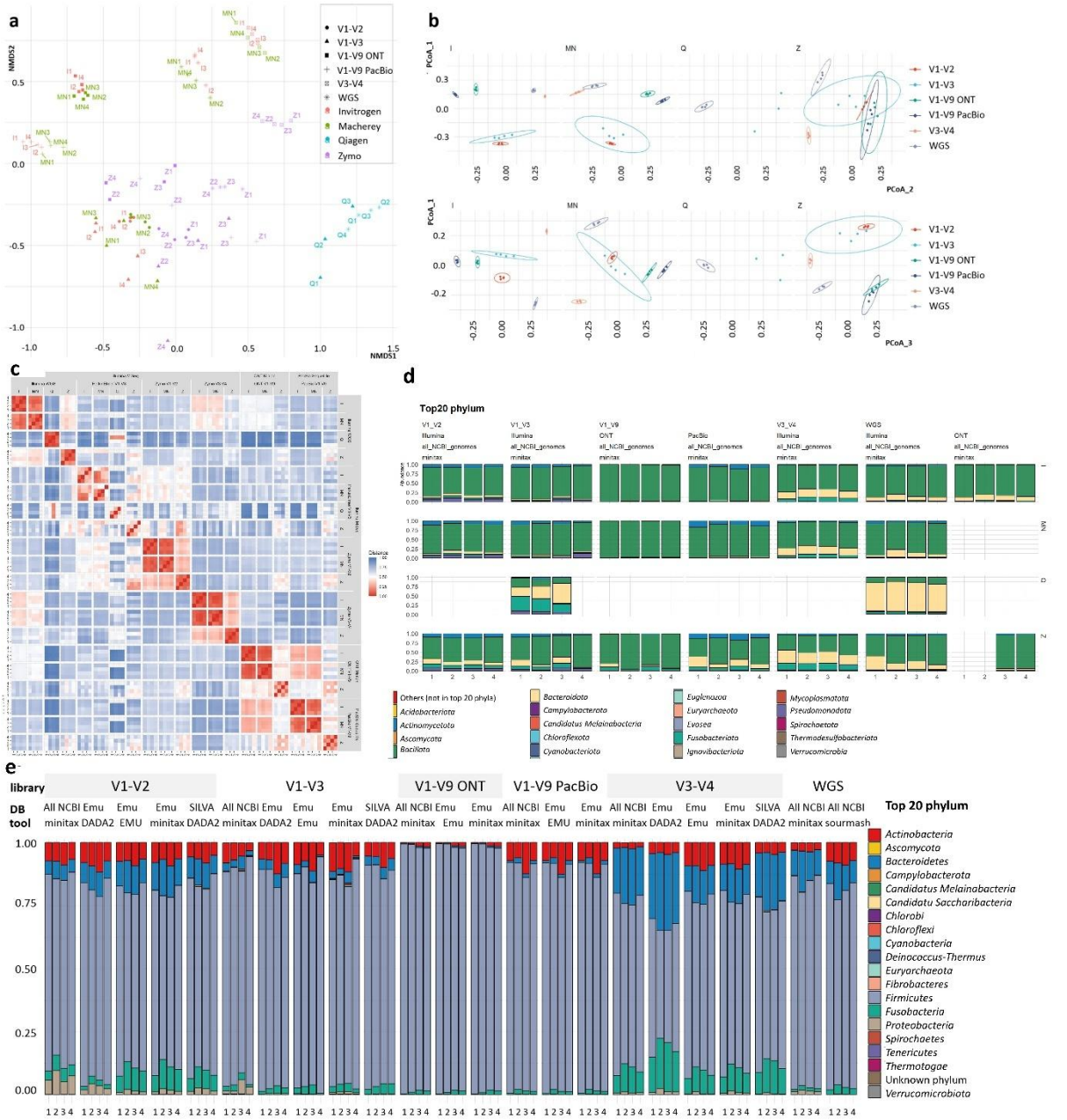


Figure 5: Comparison of library preparation protocols. a: β-diversity analysis – NMDS. The sample-wise Bray-Curtis distances between the samples were calculated and plotted using NMDS. Colors show the DNA isolation methods, while shapes indicate the library preparation protocols. **b:** β-diversity analysis—PERMDISP. PCoA visualization of the dispersion values from the PERMIDISP results. The plot is faceted according to the DNA isolation methods and colors indicate the library preparation protocols. **c:** β-diversity analysis—Heatmap. The sample-wise Bray-Curtis distances between the samples were calculated and visualized on a heatmap. The WGS data composition for the I and MN isolation kits is most similar to the bacterial composition from the Z V3-V4 library, followed by the ONT V1-V9 library. Notably, for both I and MN DNA isolation kits, the WGS and ONT V1-V9 libraries exhibit fairly similar

compositional profiles. The Z DNA isolates produce relatively consistent microbial community profiles across different library preparation techniques. The PE V1-V3 library for Q DNA shows some similarity to WGS. The I and MN DNA samples exhibit significant overlap with the V1-V2 libraries prepared from I, MN, and Z DNA, with Z DNA showing a particularly high degree of overlap with itself. Additionally, Z DNA shows good overlap with the V1-V3 library compared to other libraries. The V3-V4 library consistently shows the least similarity to other libraries across all DNA isolation methods. The PacBio results are most similar to ONT, regardless of the DNA isolation kit used. When using Z DNA isolation method, PacBio shows significant similarity in microbial composition with other SRS amplicon-based sequencing results, particularly with V1-V2, V1-V3, and then V3-V4. WGS is significantly different. The Z-isolated DNA ONT V1-V9 libraries also show similar compositional similarities with these, with PacBio possibly showing a closer match. **d:** Relative abundance of the top 20 phyla in each sample with. Rows indicate the DNA isolation methods, while columns indicate the library preparation protocols. **e:** Barplot showing the top 20 phyla in the MCS samples, sequenced using Invitrogen DNA isolation kit, according to library preparation protocols, analyzed with minitax, DADA2, and Emu programs using several different databases.

The β-diversity heatmap in Fig. 5c illustrates pairwise comparisons across DNA extraction and library preparation methods. Regardless of the isolation kit, PacBio results showed similarity between V1–V2 and V3–V4, with the latter two being most alike. In these same regions, I and MN results were closer to each other than to Z. When libraries were compared internally, V1–V3 consistently performed the weakest.

Because overlaps in Fig. 5b were incomplete and variability across libraries was evident, we measured the distance of each sample to the centroid of its group. Considerable variability was observed in V1–V3 libraries regardless of extraction method. With the exception of V1–V3 and LRS libraries made from Z DNA, most library types showed relatively low variability (Supplementary Fig. 5c).

Pairwise PERMANOVA with library preparation as the only factor confirmed that all libraries differed significantly from each other, even after multiple-testing correction. The partial overlaps visible in the PCoA plots (Supplementary Fig. 5b) were not significant by PERMANOVA, which tests centroid separation. In V1-V3 libraries, however, differences were influenced not only by centroid shifts but also by differences in dispersions.

Community composition analysis for I samples revealed that the V1–V3 libraries contained the highest proportion of *Bacillota* (formerly *Firmicutes*), while *Bacteroidota* (formerly *Bacteroidetes*) and *Fusobacteriota* (formerly *Fusobacteria*) were detected in much smaller amounts. In contrast, the V1–V2 and V3–V4 libraries showed a more balanced representation of *Bacteroidota* and *Fusobacteriota* (Fig. 5d). We compared these results with published datasets[18,100–104] (Supplementary Data 6 and Fig. 6), all of which focused on specific aspects of the canine gut microbiome. A common feature across these studies, as well as our own, was the dominance of five phyla. Nonetheless, their relative abundances varied markedly across methods (Fig. 6), highlighting the strong influence of methodological choices on outcomes.

a

Invitrogen

	V1-V2	V1-V3	V3-V4	WGS	V1-V9 (ONT)	V1-V9 (PacBio)
Firmicutes	60,147%	67,040%	63,333%	78,122%	69,710%	48,198%
Bacteroidetes	9,474%	2,741%	8,237%	1,156%	0,234%	0,099%
Fusobacteria	9,868%	2,080%	6,555%	2,082%	0,625%	1,100%
Proteobacteria	0,595%	0,849%	2,151%	1,029%	0,120%	0,156%
Actinobacteria	9,621%	4,539%	4,753%	17,149%	0,415%	0,908%

Zymo Research

	V1-V2	V1-V3	V3-V4	WGS	V1-V9 (ONT)	V1-V9 (PacBio)
Firmicutes	43,191%	50,512%	44,391%	78,736%	75,603%	63,706%
Bacteroidetes	15,003%	11,389%	14,198%	1,762%	2,402%	0,533%
Fusobacteria	17,569%	8,854%	12,108%	1,683%	6,342%	10,450%
Proteobacteria	0,971%	1,676%	4,658%	1,461%	1,320%	1,701%
Actinobacteria	11,834%	3,413%	5,576%	15,979%	1,178%	2,350%

Macherey-Nagel

	V1-V2	V1-V3	V3-V4	WGS	V1-V9 (ONT)	V1-V9 (PacBio)
Firmicutes	59,972%	66,778%	62,846%	74,882%	40,797%	29,404%
Bacteroidetes	8,561%	2,955%	7,517%	0,834%	0,287%	0,098%
Fusobacteria	9,428%	1,453%	6,409%	1,313%	0,390%	0,494%
Proteobacteria	0,528%	0,774%	2,154%	0,703%	0,083%	0,082%
Actinobacteria	11,527%	3,880%	5,752%	21,857%	0,217%	0,654%

Qiagen

	V1-V2	V1-V3	V3-V4	WGS	V1-V9 (ONT)	V1-V9 (PacBio)
Firmicutes		17,261%		51,174%		
Bacteroidetes		26,861%		18,961%		
Fusobacteria		27,488%		20,869%		
Proteobacteria		5,033%		6,620%		
Actinobacteria		1,584%		1,342%		

b

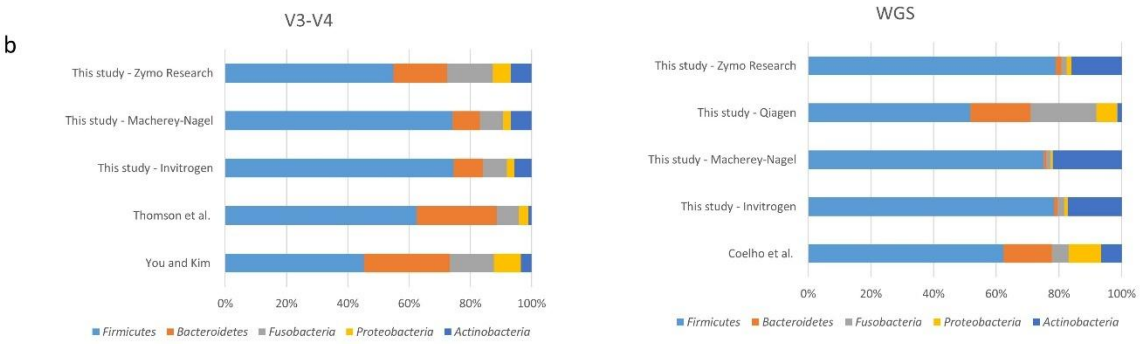


Figure 6: Comparison of the bacterial composition of various samples at the phylum level. This figure presents a comparison of bacterial composition at the Phylum level in various samples. **a:** Tables show the proportions of the most abundant Phyla in samples obtained using different DNA extraction and library preparation methods. **b:** Bar charts highlight the discrepancies between the proportions of the most abundant bacterial Phyla from other groups' data and our datasets. Taxa not included in the defined composition are grouped together under the label "other".

4.3.3. Difference in ratio of Gram-positive and Gram-negative bacteria

To investigate whether certain DNA extraction methods preferentially capture Gram-negative (Gram−) bacteria compared to Gram-positive (Gram+) ones, we grouped species abundances according to their cell-wall staining properties and performed a similar analysis. We observed that the differences between extraction methods could largely be explained by the varying resistance of bacterial cell walls to lysis. Only the I and MN kits produced microbial community profiles with nearly identical Gram+ to Gram− ratios. By contrast, all other pairwise comparisons revealed clear discrepancies in these ratios. Regarding library-level comparisons, we found several significant differences, with the V3-V4 region being particularly prominent in this respect (Supplementary Fig. 8).

4.4. Comparison of DNA isolation methods and sequencing libraries on synthetic microbial community standards

To assess how accurate our DNA extraction and library preparation approaches were, we used two synthetic microbial standards ZymoBIOMICS Microbial Community Standard D6300 (MCS) and ZymoBIOMICS Gut Microbiome Standard D6331 (GMS), both from Zymo Research] that have defined compositions (Fig. 7, Supplementary Fig. 9a, and Supplementary Data 7). From the MCS community, we prepared Illumina V1-2 and ONT V1-9 libraries, and from the GMS community, ONT V1-9 libraries, using DNA extracted by each of the four kits. The MCS contains eight bacterial species, of which five are Firmicutes and three are Proteobacteria. The GMS includes 18 bacterial species and one archaeal species, representing a more diverse mixture than MCS and offering a closer model of the human gut microbiome. Although Matsuo et al. reported that the V1-V9 primers are biased against Bifidobacterium[105], we chose to calculate our statistics without excluding this taxon, as neither the standard's manufacturer (Zymo) nor the primer manufacturer (ONT) recommends its removal. Unlike the fecal DNA analyses, where we focused on the consistency of methods within groups and the degree of similarity between them, our goal here was to determine how well each method reproduced the expected composition of the synthetic communities. We carried out this comparison using both the minitax and Emu software with the NCBI genome collection and the Emu database as references.

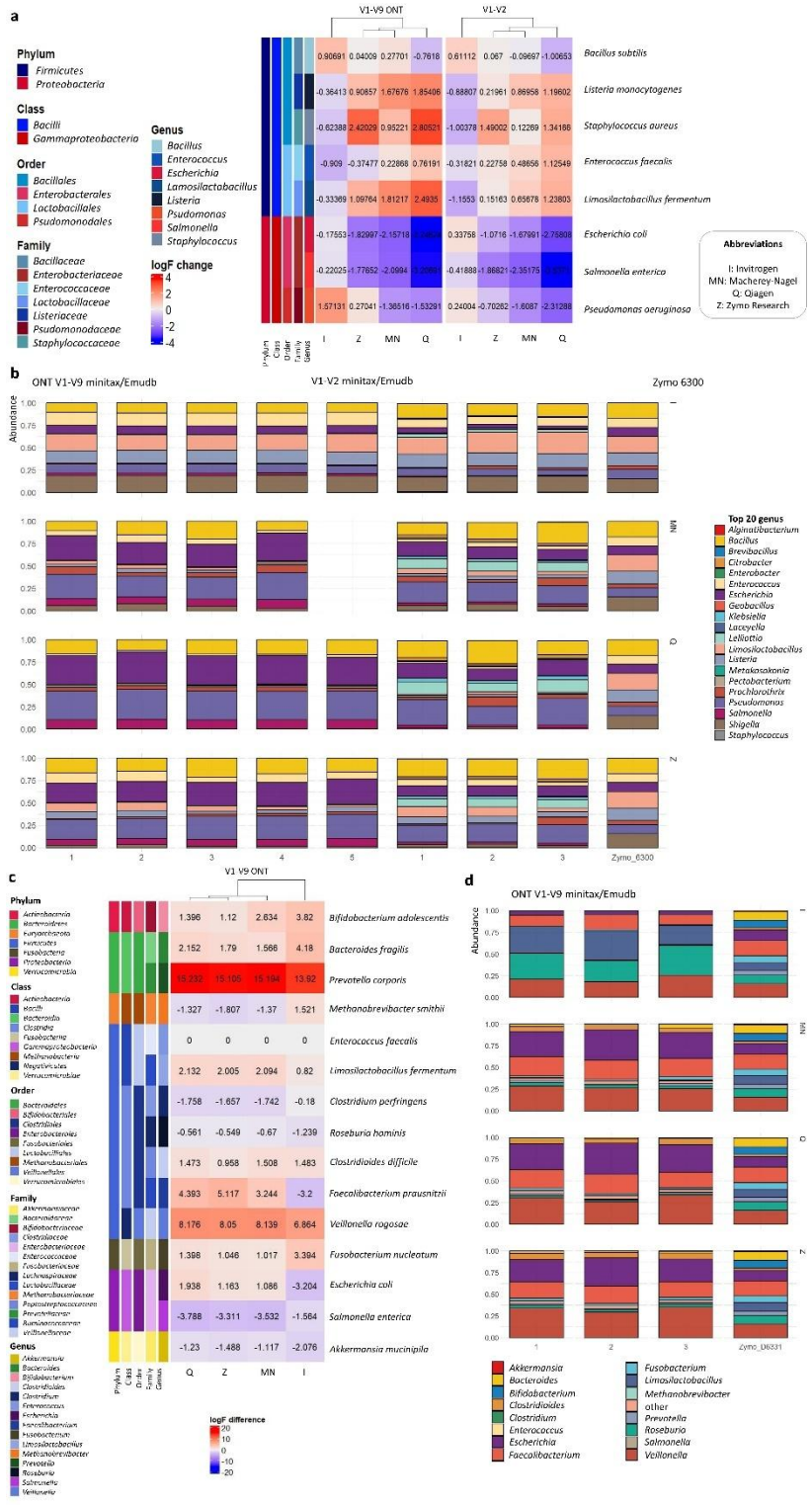


Figure 7: Comparative analysis of DNA isolation and library preparation protocols in MCS and GMS. a: Heatmap of sample-wise differences in the MCS samples. The abundance values identified by the Emu application were compared to the theoretical values provided by Zymo and log2 fold changes were estimated and are shown within the boxes. Deeper blue colors indicate lower experimental values compared to the theoretical, while more red colors indicate

higher experimental values. **b:** Barplots showing the top 20 phyla in MCS samples using the Illumina V1–V2 and ONT V1–V9 methods, according to the DNA isolation methods, analyzed with minitax using the Emu genome collection database. **c:** Heatmap of sample-wise differences in the GMS. The abundance values identified by the Emu tool were compared to the theoretical values provided by Zymo Research, with log2 fold changes calculated and displayed within the boxes. Darker blue colors represent lower experimental values compared to the theoretical, while dark red colors indicate higher experimental values. **d:** Barplots show the top 20 phyla in GMS samples using the ONT V1-V9 methods, according to the DNA isolation methods, analyzed with minitax using the Emu database.

4.4.1. Quality and yield of DNA

In our experiments, DNA yield differed significantly across methods for both MCS and GMS samples (Table 1). For MCS, the Z kit achieved the highest average yield, exceeding that of all others, including MN, which itself performed better than both Q and I. ANOVA confirmed the differences, with an F-value of 124.75 and a p-value below 0.0001. Post-hoc analysis showed that Z generated significantly more DNA than Q and I, and also outperformed MN. For GMS samples, the MN kit provided the highest yield, while Z produced a slightly lower but still substantial amount. Among the remaining kits, Q yielded more DNA than I, which produced the least. ANOVA for GMS also showed strong differences, with an F-value of 147.92 and a p-value below 0.0001. The HSD analysis indicated that MN generated significantly more DNA than I and Q, while Z produced nearly the same amount as MN. Overall, Z and MN performed best in both sample types, clearly surpassing Q and I (Table 1).

Our analysis of fragment lengths revealed further differences in degradation patterns across methods. In MCS samples, Z, I, and Q produced the longest average fragment sizes, all above 60,000 bp, whereas MN produced a shorter average peak length of 54,740 bp (Table 1). In GMS samples, Z and Q again achieved the longest fragments, both above 60,000 bp. I followed with an average peak length of 58,897 bp, and MN gave 55,538 bp (Table 1). These results indicate that both the extraction protocol and the sample type strongly affect DNA quality (Supplementary Data 7, Supplementary Fig. 9b).

The Tukey HSD test confirmed significant differences in fragment length between several kit pairs, specifically MN/I, MN/Q, and MN/Z. These findings underline the influence of extraction technique on DNA integrity in MCS samples, as presented in Supplementary Fig. 9b.

4.4.2. Microbial composition, diversity, and dispersion

4.4.2.1. ZymoBIOMICS Microbial Community Standard (MCS)

The dispersion analysis showed that in the SRS library, consistent with our fecal DNA observations, the I kit produced the lowest variability. Interestingly, in the LRS library, the Q kit also demonstrated similarly low dispersion. When examining microbial community profiles, we found that unlike the strong similarity seen *in vivo*, the MN and I kits diverged considerably in the MCS. The Q method remained the most distinct from I, while Z and MN showed closer resemblance to one another and a clear separation from I. This pattern was consistent across both library types and across both software tools and databases applied (minitax with Emu db, Emu with Emu db, and minitax with the NCBI genome collection, Fig. 7a, Fig. 7b,

Supplementary Fig. 10a, b and Supplementary Fig 11a, b), highlighting that *in vitro* standards behave differently than *in vivo* samples. The I kit produced results with the highest agreement to the expected values: no bacterial species showed significant differences using either bioinformatics tool or database, except Pseudomonas aeruginosa when analyzed with Emu and the Emu db. The other extraction methods showed between 1 and 5 significant differences depending on the wet lab approach and the bioinformatics workflow.

4.4.2.2. ZymoBIOMICS Gut Microbiome Standard (GMS)

For the more complex microbial standard, a different pattern was observed. The Z kit produced the fewest deviations from the expected composition (7–8 species), while the I kit produced the highest number of deviations (11 species). The MN and Q methods fell between these extremes, each identifying 8–10 species with significantly different abundances compared to the defined community. This trend was consistent regardless of the bioinformatics pipeline or database used. Z and MN results were more similar to each other than to I, while Q differed the most from the other three. These trends were visible using minitax with Emu db, Emu with Emu db, and minitax with the NCBI genome collection (Fig. 7c, d and Supplementary Figs. 10c, d and 11c, d).

The I kit, although successful with MCS, performed differently with the more complex GMS, indicating that robustness and sample characteristics strongly affect accuracy. Interestingly, I and MN behaved similarly when analyzing the *in vivo* samples (dog feces), showing that context influences performance. DNA concentrations were especially low for the GMS I kit (Supplementary Data 7b) and for MCS using Q and I, whereas *in vivo* only the Q kit gave insufficient amounts. These results emphasize that DNA yield must be taken into account, since low concentrations can impair downstream steps. Kits that consistently produce higher yields may be more suitable for analyzing complex microbial communities. Overall, none of the kits matched theoretical expectations equally well across different sample types. Portik et al. also concluded that MCS (D6300) is more reliable than GMS (D6331)[89], suggesting that synthetic standards are not always the best indicators for selecting DNA isolation kits or sequencing strategies for *in vivo* applications. Careful evaluation is recommended before starting large-scale studies, especially when working with less familiar sample sources.

4.4.3. Difference in ratio of Gram(+) and Gram(-) bacteria

In this part of the analysis, we found that the Q kit tended to overrepresent Gram− bacteria in MCS samples and also produced higher ratios of Gram− species in *in vivo* datasets. The MN kit showed a similar overrepresentation in fecal samples, but this bias did not appear in MCS.

For both kits, the main underrepresented Gram+ taxa were Lactobacillus/Limosilactobacillus, while the main overrepresented Gram− taxa were Escherichia and Salmonella. Our results suggest that the MN method favors Gram− bacteria in a controlled environment such as MCS, but this trend may be masked in fecal samples due to their complexity, which affects the lysis of both Gram+ and Gram− species. In GMS, the I kit underestimated Gram− bacteria, whereas MN, Q, and Z provided ratios close to the expected values.

4.5. Laboratory work: key findings

Our assessment of the DNA extraction kits and library preparation protocols revealed several important points (Table 2 and Table 3). The Z kit provided very high DNA yields and the longest fragments, making it the most effective overall for high-quality diversity analysis. The MN kit produced high yields and results close to theoretical compositions but gave the shortest fragments. The I method generated consistent data of good quality but showed more variability in yield and fragment length compared with Z and MN. The Q kit produced lower yields and higher host contamination, which reduced its reliability for microbial diversity analysis (Supplementary Fig. 3b). Regarding library preparation, Illumina MiSeq with Illumina DNA Prep (WGS) produced the most accurate and high-quality data, although it required significant labor and was expensive. Illumina MiSeq with PerkinElmer V1-V3 was less consistent and often generated incorrect fragments. ONT MinION with ONT V1-V9 performed well with minimal hands-on work but produced reads of lower quality compared to Illumina and PacBio. PacBio Sequel IIe with PacBio V1-V9 generated high-quality reads but was both expensive and labor-intensive.

Kit	Advantages	Disadvantages
Qiagen	• Suitable for certain applications	• Lowest DNA yield among the kits
		• Higher levels of host DNA contamination
		• Less accurate in representing microbial diversity
Invitrogen	• Good overall quality	• Higher variability in yield and fragment length compared to Z and MN kits
	• High amount of DNA from stool samples	• Less optimal performance with GMS samples
	• Consistent results in various libraries and sequencing methods	• Less effective in certain sample types (e.g., MCS and GMS) compared to Zymo and MN
	• Closest match to theoretical composition in MCS samples	
Macherey-Nagel	• High overall DNA yield	• Shorter average DNA fragment length
	• Better performance in GMS samples	
	• Good alignment with theoretical microbial composition in MCS samples	
	• Comparable yield to Z kit when adjusted for initial volume	
Zymo Research	• High DNA yield in canine fecal and MCS samples, highest in GMS sample	• Less accurate representation in MCS
	• Provides high-quality DNA	
	• Longest average DNA fragment length	
	• Superior performance in MCS samples	
	• Closest match to theoretical composition in GMS samples	
	• Consistent results and performance for LRS and WGS libraries	
	• Stable performance with low relative dispersion	

Table 2: Summary of DNA isolation methods: yield, fragment quality, and suitability for various sample types

This table details the advantages and disadvantages of various DNA isolation kits, including Zymo Research, Qiagen, Macherey-Nagel, and Invitrogen. Key factors include DNA yield, fragment length, contamination levels, and overall performance across different sample types.

Sequencing platform / library prep kit	Advantages	Disadvantages
Illumina MiSeq/Illumina DNA Prep (WGS)	• Known for high-quality data and accuracy in determining true microbial compositions	• Labor-intensive and complex library preparation
	• Shows high consistency across libraries	• Highest hands-on time
	• High-quality libraries	• Generally more costly
Illumina MiSeq/PerkinElmer V1-V3	• Widely used V-region	• The applied kit often produces nonspecific or erroneous fragments
		• Shows the least similarity to other libraries
		• Performance can be inconsistent across different samples
ONT MinION/ONT V1-V9	• Requires less experimental work compared to other methods	• Generally lower read quality compared to Illumina and PacBio
	• High-quality, high-consistency libraries	• May not consistently match well with other libraries, particularly short-read amplicon libraries
	• Shows high similarity with PacBio V1-V9 results and moderate similarity to Illumina WGS data	
PacBio Sequel IIe/PacBio V1-V9	• Provides high-quality reads	• Labor-intensive preparation
	• High-quality, high-consistency libraries	
	• Shows strong similarity with ONT results	• Typically more costly
	• Shows good moderate similarity with V1-V2 and V1-V3 libraries	
Illumina MiSeq/Zymo Research V1-V2	• Similarity with PE V1-V3 data	• Moderate costs and labor intensity
	• High-quality libraries	
	• High consistency, reliable data quality	
Illumina MiSeq/Zymo Research V3-V4	• Strong similarity with Illumina WGS (with I and MN DNA samples)	• Moderate costs and labor intensity
	• High-quality libraries	
	• High consistency, reliable data quality	

Table 3: Comparison of library preparation methods: performance metrics and practical considerations

This table provides an overview of library preparation methods, including Illumina DNA Prep, PerkinElmer V1-V3, Zymo Research V1-V2 and V3-V4, ONT V1-V9, and PacBio V1-V9. It highlights the benefits and drawbacks related to sequencing quality, consistency, labor intensity, and cost for each preparation technique.

4.6. Comparison of bioinformatics techniques

To reduce the variability in canine microbiome composition caused by different DNA extraction methods, we used only the I kit for comparing bioinformatics approaches. To minimize differences linked to database choice, we selected the NCBI genome collection as the reference (Fig. 5e).

In this part of the study, we examined the canine microbiome at the phylum, order, genus, and species levels, comparing the influence of different databases (Emu, NCBI genomes, and SILVA) and bioinformatic tools (Emu, minitax, DADA2, and sourmash). A full-model PERMANOVA showed that databases and programs strongly affected the results, explaining 59.4% of the observed variation in microbial community composition. The "library" factor explained 20.1%, and the remaining 20.4% was unexplained. Pairwise comparisons revealed that DADA2 produced significantly different outcomes with the Emu database compared to the other tools, while minitax closely matched Emu. Using DADA2 with SILVA produced even larger differences (Fig. 5e). Although several combinations of databases and programs yielded significant differences ($p < 0.05$), the minitax and NCBI genome pairing was highly consistent. Significant differences were only observed when minitax with NCBI was compared to minitax with Emu for the ONT V1–V9 library, and to sourmash with NCBI for WGS. Aside from these, results remained stable across libraries. Therefore, the minitax and NCBI genomes combination was the most reliable choice for evaluating DNA extraction kits and library protocols (Fig. 5e).

We also analyzed MinION datasets prepared with the ONT 16S Barcoding Kit using the 16S module of the EPI2ME Labs software (EPI2ME Desktop, version 2021.09.09). This program provides a simplified workflow and rapidly generates Sankey diagrams on its online platform. Our initial analysis indicated that Blautia was the dominant genus (Fig. 8a, b), with an average abundance of around 80%. This was consistent across all DNA extraction kits but conflicted with both published data and sequencing from other methods (Illumina WGS, V1-V2, V3-V4, V1–V3, PacBio V1-V9). When analyzed with Emu and minitax, the Blautia proportion aligned with the ~15% reported in earlier studies[101], showing that EPI2ME had strongly inflated this genus. However, closer inspection showed that EPI2ME did not directly overestimate Blautia but instead excluded some genera and underrepresented others. To clarify this, we compared the ONT V1-V9 datasets processed with EPI2ME and Emu. In our canine samples, EPI2ME failed to detect Peptacetobacter, Faecalimonas, and Mediterraneibacter, discarding about 75% of the reads and altering the community ratios. To confirm whether these discrepancies were entirely due to bioinformatics filtering, we expanded our analysis beyond the original dog

sample (Fig. 8a, b). We included fecal samples from six additional dogs (Fig. 8c), neonatal[95] (Fig. 8d) and adult[105] (Fig. 8e) human stool, adult human skin[94] (Fig. 8f), and the MCS standard (Fig. 8g). All ONT V1-V9 datasets were analyzed with both EPI2ME and Emu.

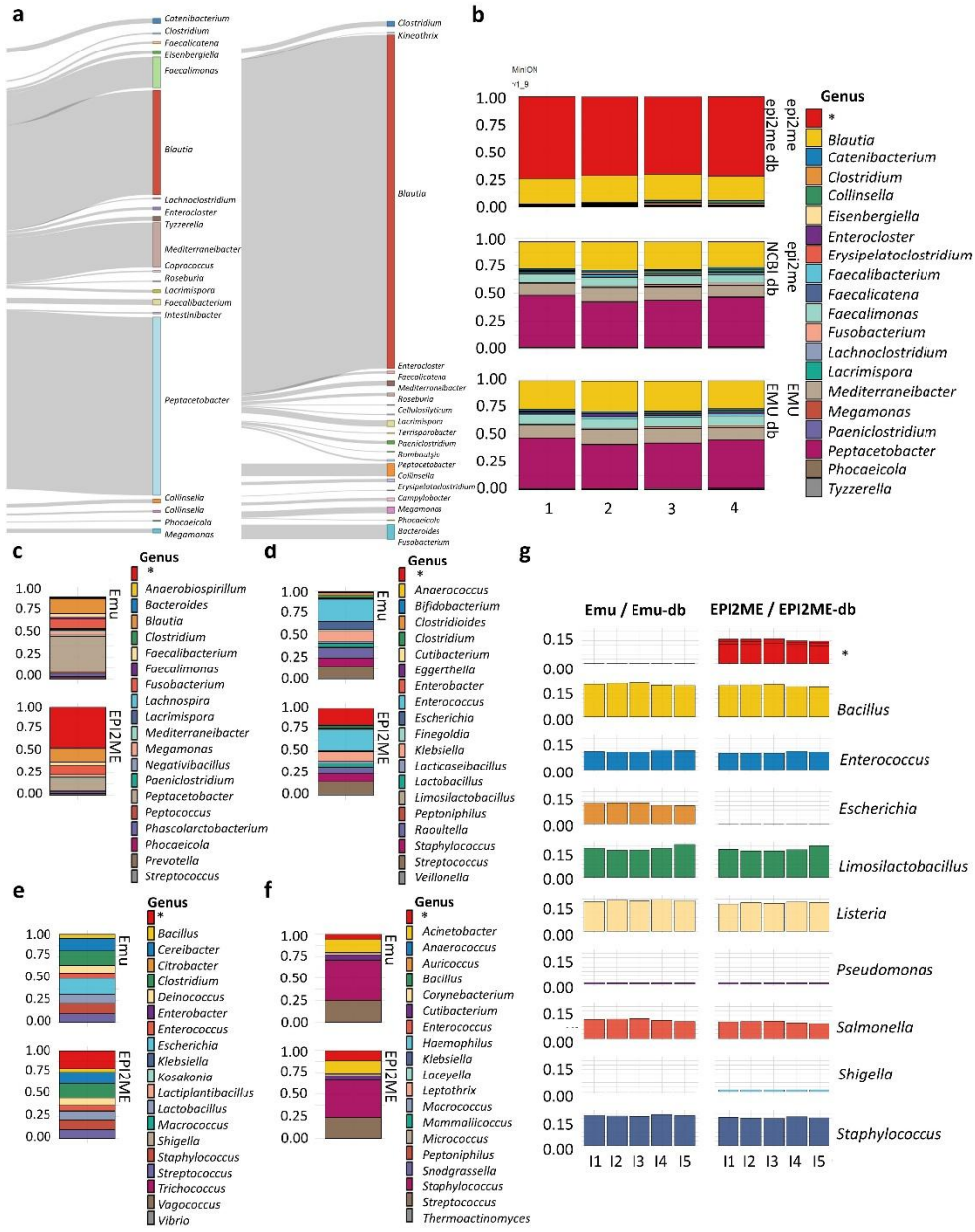


Figure 8: The EPI2ME analysis platform yields inaccurate genus-level data.

Results from the 16S module of the EPI2ME Labs program by ONT employed on canine stool sample (**a**,**b**), fecal samples from six additional dogs (**c**), neonatal[95] (**d**) and adult[105] (**e**) human stool specimens, adult human skin specimen[94] (**f**), and MCS (**g**). In our canine samples, the software failed to yield abundances to the Peptacetobacter, Faecalimonas, and Mediterraneibacter genera, discarding 75% of the reads and skewing the actual compositional ratios. * In **b-g** parts of this figure the blocks highlighted in red represent the reads that the EPI2ME program filters out and excludes from the analysis due to the LCA tag.

For the six control dog fecal samples, EPI2ME discarded around 50% of the reads (Fig. 8c). Similar to the original dog sample, *Peptacetobacter* was not detected and *Faecalimonas* and *Mediterraneibacter* were underestimated. Using ONT V1-V9 data from the European Nucleotide Archive (ENA), we compared EPI2ME and Emu again. For fecal samples, EPI2ME failed to process 15–25% of the reads and completely excluded Escherichia (Fig. 8d, e). For skin samples, the differences were minimal, with about 5% of the reads excluded by EPI2ME, most of which belonged to *Cutibacterium* and *Staphylococcus* (Fig. 8f).

In the microbial community mix from Zymo Research, EPI2ME excluded about 15% of the reads and failed to detect *Escherichia*. At the same time, it identified *Shigella*, a genus absent from the mixture (Fig. 8g).

We found that EPI2ME discarded large portions of data due to the use of Lowest Common Ancestor (LCA) tags. In the BLAST module, the LCA tag removes reads if the top three predicted genera do not match. When we reassigned these reads using their NCBI taxon IDs, the results produced by EPI2ME became very similar to those from Emu. This showed that the LCA filtering step is unnecessary and lowers accuracy, making EPI2ME less reliable when applied in its default configuration (Fig. 8b).

4.7. Evaluation of minitax: benchmarking across various sequencing methods and data types

To allow consistent comparisons across different sequencing datasets, we created minitax, a tool tailored for metagenome sequencing. We tested it thoroughly on multiple platforms and datasets, and compared its performance with other bioinformatic tools based on how accurately they could reproduce the reference microbial composition, measured by the correlation (r^2 values) between observed and theoretical communities. In addition, we applied Chi-square tests to determine whether the reconstructed and expected distributions showed statistically significant differences.

4.7.1. Comparing minitax with Emu using ONT V1-V9 sequencing of MCS and GMS

We compared minitax to Emu, which also uses minimap2 for alignment but applies an expectation-maximization step afterwards. For the MCS dataset, minitax and Emu produced highly similar results with the Emu database, confirming the robustness of minitax (Fig. 9a and Supplementary Fig. 12a). Even though full genome databases are usually not applied in 16S rRNA-Seq, minitax with the NCBI genome collection provided accurate reconstructions up to the genus level. This flexibility is valuable for researchers wishing to use the same database

across both WGS and 16S rRNA analyses. However, species-level accuracy dropped notably when the NCBI database was applied, emphasizing the importance of choosing a database that matches the resolution required. This pattern was also visible with GMS (Fig. 9a and Supplementary Fig. S12b). Since both programs performed comparably across the two synthetic standards, the differences seen between MCS and GMS reconstructions are more likely due to extraction methods or experimental variables rather than the bioinformatics tools or databases. The Chi-square results showed that only for the I kit with MCS did the reconstructed composition, whether with Emu or minitax using the Emu database, not differ significantly from the expected values (Supplementary Fig. S12a, b).

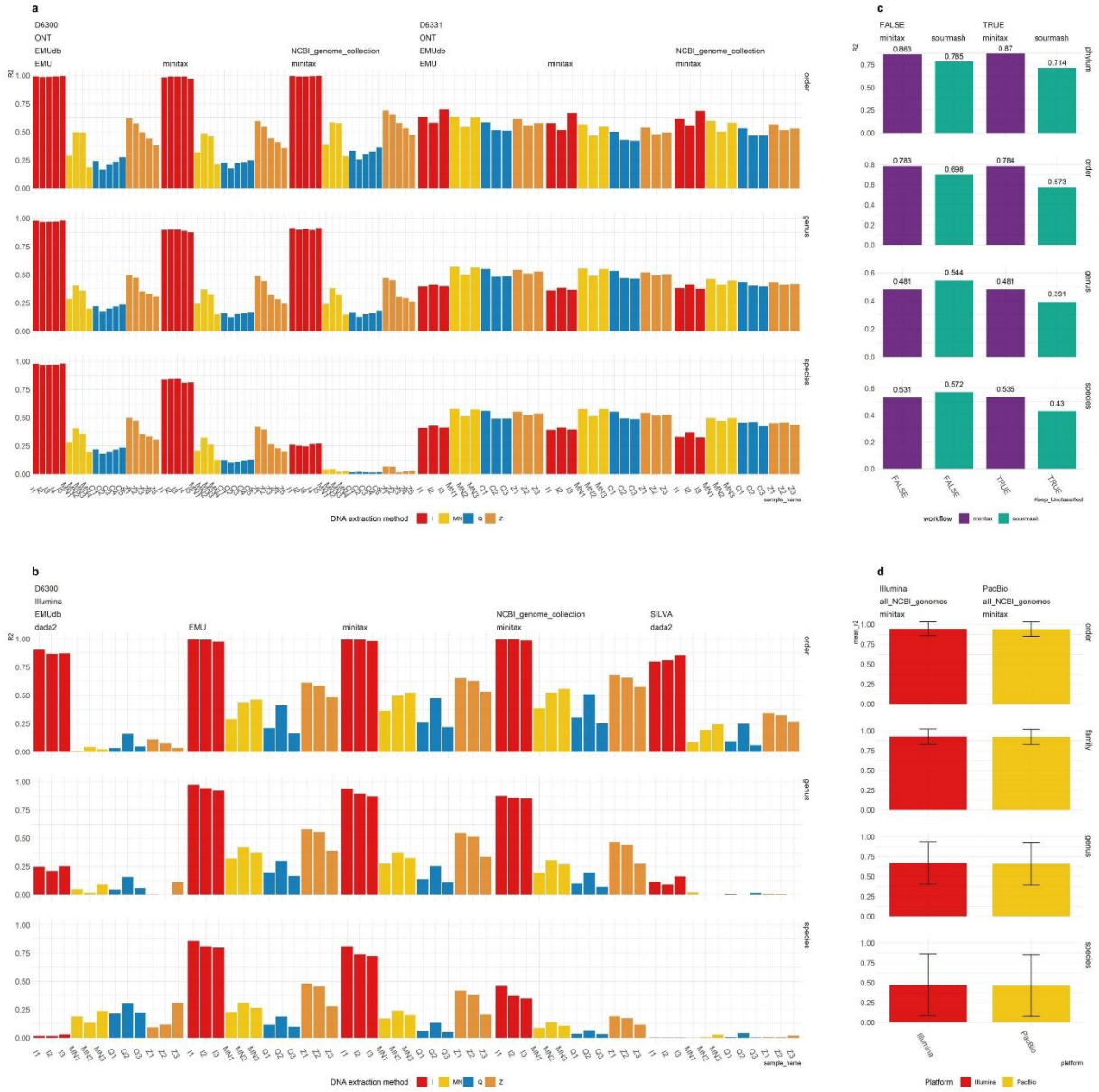


Figure 9: Comparative analysis of minitax performance across different sequencing platforms and datasets. The comparison between the theoretical and the observed microbial compositions, was generated using different software and databases. r^2 values were calculated across four taxonomic levels (phylum, order, genus, species) and plotted. **a** ONT V1-V9 Sequencing: performance comparison of minitax against Emu for the Zymo D6300 microbial community reference. **b** Illumina V1-V2 sequencing: minitax benchmarking against Emu and DADA2 for Zymo D6300 MCS. **c** PacBio HiFi WGS: minitax vs. sourmash performance comparison on Zymo D6331 MCS dataset. **d** CAMISIM mouse gut dataset: comparative analysis of minitax performance on Illumina and PacBio platforms. Ten samples were utilized from the CAMISIM database. Y-axis represent the mean r^2-value for each group, error bars show standard deviation (n = 10).

4.7.2. Comparing minitax with Emu and DADA2 using Illumina V1-V2 sequencing of MCS

We next applied Illumina-sequenced V1-V2 data from MCS DNA. Here, both minitax and Emu produced substantially higher r^2 values than DADA2 at both genus and species levels, even when DADA2 was paired with the standard SILVA database (Fig. 9b). It is worth noting that although both Emu and DADA2 are tailored for amplicon sequencing, minitax remained competitive, particularly because of its ability to handle both 16S rRNA and WGS datasets. Emu achieved slightly higher r^2 values than minitax at the species level when used with the Emu database. In the MCS samples sequenced from the V1-V2 region, Chi-square testing confirmed that both Emu and minitax produced microbial compositions not significantly different from the expected community (Supplementary Fig. 12b), similar to the results with ONT V1-V9.

4.7.3. Comparing minitax with sourmash using MCS data of PacBio HiFi WGS

We also compared minitax with sourmash, one of the most widely used tools for analyzing both LRS and SRS WGS data[106]. This comparison was performed on a PacBio HiFi dataset (NCBI accession: SRX9569057) from GMS. We calculated r^2 values both including and excluding unclassified reads. Minitax surpassed sourmash when unclassified reads were included in abundance estimates but performed slightly worse at the species level when these reads were excluded (Fig. 9c), since this adjustment shifted the relative abundances of identified taxa. Chi-square tests with three different detection thresholds showed that with minitax, the reconstructed and expected compositions were not significantly different when 0.1% and 0.01% thresholds were applied (Supplementary Fig. 12c).

4.7.4. CAMISIM: simulated mouse gut datasets

To broaden the evaluation, we also tested CAMISIM-simulated mouse gut datasets, which included 10 samples each from PacBio and Illumina. In both cases, minitax showed stable performance, achieving $r^2 = 0.96$ at the phylum level. This value decreased to a mean of $r^2 = 0.46$ with Illumina and $r^2 = 0.55$ with PacBio at the species level (Fig. 9d).

5. Discussion

In this comprehensive study, we compared metagenomic strategies using canine fecal samples alongside two synthetic microbial community standards. Our main goal was to assess the efficiency of different methodological steps, including DNA extraction, library preparation, and bioinformatic processing. To support this, we developed "minitax," a flexible bioinformatic tool designed to work with a wide range of metagenomic laboratory protocols.

Selecting the right DNA extraction method is critical for maximizing yield, minimizing fragmentation, and maintaining the integrity needed for downstream analyses.[20] An ideal extraction kit produces high-quality, high-yield DNA to lower the chance of false negatives[96], removes PCR inhibitors that are common in fecal material[107–111], ensures consistent results, and efficiently lyses Gram+ bacterial cell walls to give a faithful picture of the microbial community.[21,98,99,112,113]

Our evaluation of four DNA isolation kits—Zymo (Z), Qiagen (Q), Macherey-Nagel (MN), and Invitrogen (I)—showed significant differences in yield, fragment size, and overall performance. In canine fecal samples, the Z kit provided the highest DNA yield and the longest fragments with very low variation, highlighting its stability (Table 1, Fig. 4a–f). The Q kit yielded the least DNA and had a higher proportion of host contamination. The MN kit produced the largest overall yield but shorter fragments, while the I kit generated acceptable quality but with more variability in both yield and fragment length. For MCS, the Z kit produced considerably more DNA with longer fragments, outperforming MN, which had shorter peaks (Table 1). For GMS, MN slightly outperformed Z in yield, although both performed strongly (Table 1). Altogether, Z and MN showed the best yields, with Z excelling in MCS and MN performing better in GMS.

When examining microbial diversity in canine fecal samples, I and MN consistently provided similar outcomes across different libraries, sequencing methods, and bioinformatic pipelines. Z differed somewhat from these, while Q consistently performed the weakest in capturing microbial diversity (Fig. 4g–i).

For MCS, I provided the closest match to the expected community, while Z and MN were less accurate. Q deviated most strongly from the theoretical composition. In GMS, MN and Z produced results closest to expectations, although neither was ideal. I, which performed well with MCS, did not perform as effectively with GMS (Fig. 7).

These findings suggest that sample type and purification strategy play a key role in kit performance. The protective medium used for MCS and GMS may not be optimal for any kit, possibly reducing efficiency. Extremely low yields from some kits may also explain discrepancies in microbial representation. This shows that synthetic standards are not always appropriate for choosing DNA extraction kits.

Overall, Z is recommended for most purposes, since it delivers high yield, stable output, and long DNA fragments, which are particularly important for long-read sequencing and diversity analysis. MN and I also perform well, but each has strengths and weaknesses that vary with sample type and application (Table 2). Q proved to be the least reliable, with low-quality DNA and poor representation of the microbial community.

Therefore, the choice of extraction kit should depend on the type of sample and the requirements of the sequencing application.

We also observed that ONT 16S rRNA amplicon libraries required the least experimental work but had lower read quality than Illumina or PacBio. Illumina WGS and PacBio amplicon libraries required more effort but consistently produced high-quality results. The PerkinElmer (PE) V1–V3 protocol often generated nonspecific DNA fragments, reducing performance, while the other methods showed strong quality and consistency.

Our analysis, consistent with other reports[19,114], showed that WGS captures more taxonomic diversity than 16S sequencing. However, full-length 16S sequencing provides a clearer view of bacterial communities compared with SRS targeting only one or a few variable regions. For this reason, we included both SRS and LRS approaches, as well as 16S rRNA amplicon sequencing and metagenomic WGS.

We found that WGS data obtained from I and MN extractions closely matched the microbial profiles seen in the Z V3-V4 libraries. Similarly, ONT V1-V9 libraries based on I and MN DNA showed comparable community structures. Libraries made from Z DNA were generally consistent with those from the other kits, demonstrating that Z produces stable microbial community profiles across different library preparations.

The V1-V3 libraries showed considerable variability across all extraction kits, regardless of the method used. This highlights the importance of carefully selecting library preparation methods for reliable microbial profiling. A detailed evaluation of 16S V1-V9 sequencing with ONT and PacBio showed that only the Z kit gave consistent results across both platforms. We also

compared V1-V2 and V1-V9 libraries on MCS and found that the I kit produced the closest match to theoretical expectations in both. MN resembled Z more closely and differed significantly from I in both SRS and LRS. These observations emphasize the different outcomes seen between *in vitro* and *in vivo* experiments and suggest that MCS may not be suitable for validation in other systems.

The α-diversity analysis showed that SRS libraries produced generally consistent values, except where differences came from library preparation. PacBio and Illumina methods displayed distinct diversity profiles that influenced microbial community composition. β-diversity analyses further indicated that library preparation shaped clustering more than DNA extraction. The clear separation between V1-V2 and V3-V4 underlined how strongly library methods affect microbial profiles. The contrasting patterns in microbial diversity and composition between V1-V3 and V3-V4 showed that sequencing method plays a decisive role. The consistent clustering by library rather than extraction method highlighted the importance of selecting the right preparation protocol for accurate analysis.

The importance of the bioinformatic pipeline in metagenomics has been noted before[89]. Our findings also show that roughly 60% of the variation in microbial profiles comes from computational choices. Emu and minitax consistently produced closely matching results, and for MCS they aligned well with theoretical compositions of all amplicon libraries. Sourmash and minitax also showed strong agreement for WGS, demonstrating the broad applicability of minitax. In contrast, results from DADA2 differed greatly from both minitax and Emu *in vivo* and *in vitro*.

For specific sequencing data types, Emu was the most reliable for amplicon sequencing, performing well across V-regions. Although this study did not focus in depth on WGS bioinformatic pipelines, sourmash—previously highlighted by Portik and colleagues[89] as highly accurate—performed effectively in our work. This suggests that Emu is best suited for amplicon data, while sourmash is highly reliable for WGS. Minitax proved to be a versatile option, consistently working across library types and sequencing methods. Although it did not always give the very highest correlations, it often outperformed other pipelines by producing stable results. It was also effective in handling both amplicon and WGS data, reaching genus-level resolution even with genome-wide databases. This flexibility makes it a strong candidate for comparative studies spanning different sequencing strategies.

In conclusion, our study provides a detailed evaluation of gut microbiome analysis and shows that optimizing and standardizing methods is crucial for accuracy and reproducibility. At the same time, applying multiple complementary methods can help overcome the limitations of each individual approach, leading to a deeper understanding of the microbiome[115].

We recommend the following wet-lab pipelines for gut microbiome profiling:

1. For cost-effective studies, use Z DNA extraction with ONT V1-V9 libraries. This provides an affordable option, suitable for longitudinal designs where maximum precision is not essential.
2. For a balanced approach, apply I DNA extraction with V3-V4 libraries on Illumina. This combines moderate cost with solid accuracy, offering dependable microbial profiling.
3. For the most precise projects, use MN DNA extraction with Illumina DNA Prep (WGS). Though more expensive, this provides the most detailed and accurate results.

For bioinformatics analysis, we recommend:

1. For cross-platform comparisons, use minitax with the NCBI genome collection, as it performed well across both amplicon and WGS data.
2. For amplicon sequencing, Emu provides the most reliable results.
3. For WGS, sourmash is highly effective, as noted by previous studies and supported by our findings.

We did not fully exhaust the dataset in all analyses, making it a potentially valuable resource for further investigation by the research community.

4. Summary

In this study, various laboratory and bioinformatic methods for metagenomic profiling of the gut microbiome were compared in order to determine how technical decisions influence the obtained results. The investigations were carried out on different sample types, including dog fecal samples and artificially constructed microbial communities, using several DNA extraction kits, library preparation strategies, and sequencing platforms. The data obtained in this way were processed with different bioinformatic approaches, allowing for a detailed evaluation of the effects of laboratory and computational steps. Based on the results, it can be concluded that a universally best method cannot be identified, as the chosen strategy depends on the sample type and the research objective.

During DNA extraction, the Zymo Research Quick-DNA HMW MagBead Kit proved to be the most reliable, as with its high yield and good quality it provided a stable basis for both short- and long-read sequencing procedures. Other kits, such as Qiagen QIAamp Fast DNA Stool Mini Kit, produced less accurate results. In the case of comparing sequencing platforms, the accuracy of 16S sequencing, combined with long-read sequencing, enables more in-depth analyses.

In terms of bioinformatic processing, it was demonstrated that the applied software and reference database explain a significant part of the variability, meaning that the computational methodology plays a decisive role in the final results. Our developed tool, “minitax,” showed consistent and stable performance, while with the Oxford Nanopore EPI2ME software, frequent biases were observed.

Overall, it can be concluded that in microbiome research, the combined application of multiple complementary methods is justified, since only in this way can sufficient accuracy and reproducibility be ensured. The results highlight that the success of metagenomic studies is determined by the entire methodological chain, therefore in future research standardized procedures and the critical comparison of different techniques are of key importance.

5. Funding

The project was funded by the HAS (MTA) Lendület Programme LP2020-8/2020.

6. Acknowledgements

First and foremost, I wish to express my sincere gratitude to my supervisor, Dr. Dóra Tombácz, for her invaluable guidance and support both throughout and prior to my doctoral studies. I am equally grateful to Prof. Dr. Zsolt Boldogkői, Professor and Head of the Institute of Medical Biology at University of Szeged Albert Szent-Györgyi Medical School, for granting me the opportunity to conduct my research within his institute, as well as for the generous support and numerous opportunities he has provided.

I would like to express my gratitude to Ákos Dörmő, Tamás Járay, Dr. Zsolt Csabai, Dr. Balázs Kakuk, and Dr. István Prazsák for their contributions and assistance, which greatly supported the realization of the publication forming the foundation of this thesis.

I wish to extend my thanks to all the staff members of the Institute of Medical Biology at University of Szeged Albert Szent-Györgyi Medical School for their work and assistance.

And finally, above all, I wish to express my deepest gratitude and heartfelt thanks to my love, Dr. Diána Szűcs, and to my entire loving family, without whose endless love and support I could not have reached this point. Thank you.

7. References

1. Ursell, L. K. *et al.* The intestinal metabolome: An intersection between microbiota and host. *Gastroenterology* **146**, 1470–1476 (2014).
2. Grice, E. A. & Segre, J. A. The human microbiome: Our second genome. *Annual Review of Genomics and Human Genetics* vol. 13 151–170 (2012).
3. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
4. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
5. Nayfach, S. *et al.* A genomic catalog of Earth's microbiomes. *Nat Biotechnol* **39**, 499–509 (2021).
6. Almeida, A. *et al.* A new genomic blueprint of the human gut microbiota. *Nature* **568**, 499–504 (2019).
7. Franzosa, E. A. *et al.* Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat Rev Microbiol* **13**, 360–372 (2015).
8. Hou, K. Microbiota in health and diseases. *Sig. Transduct. Target Ther.* **7**, (2022).
9. Mendes, R. *et al.* Deciphering the Rhizosphere Microbiome for Disease-Suppressive Bacteria. *Science* **332**, 1097–1100 (2011).
10. Adrio, J. L. & Demain, A. L. Microbial Enzymes: Tools for Biotechnological Processes. *Biomolecules* **4**, 117–139 (2014).
11. The Microbial Engines That Drive Earth's Biogeochemical Cycles | Science. https://www.science.org/doi/10.1126/science.1153213.
12. Hillman, E. T., Lu, H., Yao, T. & Nakatsu, C. H. Microbial ecology along the gastrointestinal tract. *Microbes and Environments* vol. 32 300–313 (2017).

13. Laterza, L., Rizzatti, G., Gaetani, E., Chiusolo, P. & Gasbarrini, A. The gut microbiota and immune system relationship in human graft-versus-host disease. *Mediterranean Journal of Hematology and Infectious Diseases* vol. 8 (2016).
14. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
15. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
16. Roberfroid, M. B., Bornet, E., Bouley, C. & Cummings, J. H. *Colonic Microflora: Nutrition and Health Summary and Conclusions of an International Life Sciences Institute (ILSI) [Europe] Workshop Held in Barcelona, Spain*. (1995).
17. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology* vol. 19 55–71 (2021).
18. Coelho, L. P. Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome* **6**, (2018).
19. Lewis, S. Comparison of 16S and whole genome dog microbiomes using machine learning. *BioData Min.* **14**, (2021).
20. Costea, P. I. Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* **35**, (2017).
21. Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z. & Forney, L. J. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* **7**, (2012).
22. Zhang, B. *et al.* Impact of Bead-Beating Intensity on the Genus- and Species-Level Characterization of the Gut Microbiome Using Amplicon and Complete 16S rRNA Gene Sequencing. *Frontiers in Cellular and Infection Microbiology* **11**, (2021).
23. 1,500 scientists lift the lid on reproducibility | Nature. https://www.nature.com/articles/533452a.

24. Ducarmon, Q. R., Hornung, B. V. H., Geelen, A. R., Kuijper, E. J. & Zwittink, R. D. Toward standards in clinical microbiota studies: comparison of three DNA extraction methods and two bioinformatic pipelines. *mSystems* **5**, (2020).
25. Tourlousse, D. M. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. *Microbiome* **9**, (2021).
26. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **35**, (2017).
27. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, (2014).
28. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, (2019).
29. Lu, J. & Salzberg, S. L. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome* **8**, (2020).
30. Brown, C. T. & Irber, L. sourmash: a library for MinHash sketching of DNA. *J. Open Source Softw.* **1**, (2016).
31. Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biology* **21**, 30 (2020).
32. Watson, M. & Warr, A. Errors in long-read assemblies can critically affect protein prediction. *Nat Biotechnol* **37**, 124–126 (2019).
33. Nicholls, S. M., Quick, J. C., Tang, S. & Loman, N. J. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience* **8**, giz043 (2019).
34. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).

35. Ju, J. *et al.* Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proceedings of the National Academy of Sciences* **103**, 19635–19640 (2006).

36. Mardis, E. R. Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry* **6**, 287–303 (2013).

37. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135–1145 (2008).

38. Metzker, M. L. Sequencing technologies — the next generation. *Nat Rev Genet* **11**, 31–46 (2010).

39. Kircher, M. & Kelso, J. High-throughput DNA sequencing – concepts and limitations. *BioEssays* **32**, 524–536 (2010).

40. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333–351 (2016).

41. Quail, M. A. *et al.* A large genome center's improvements to the Illumina sequencing system. *Nat Methods* **5**, 1005–1010 (2008).

42. Eid, J. *et al.* Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**, 133–138 (2009).

43. Travers, K. J., Chin, C.-S., Rank, D. R., Eid, J. S. & Turner, S. W. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* **38**, e159 (2010).

44. Korlach, J. *et al.* Long, Processive Enzymatic Dna Synthesis Using 100% Dye-Labeled Terminal Phosphate-Linked Nucleotides. *Nucleosides Nucleotides Nucleic Acids* **27**, 1072–1083 (2008).

45. Levene, M. J. *et al.* Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations. *Science* **299**, 682–686 (2003).

46. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* **10**, 563–569 (2013).

47. Fuller, C. W. *et al.* The challenges of sequencing by synthesis. *Nat Biotechnol* **27**, 1013–1023 (2009).

48. Wenger, A. M. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* **37**, 1155–1162 (2019).

49. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences* **93**, 13770–13773 (1996).

50. Bayley, H. Nanopore Sequencing: From Imagination to Reality. *Clin Chem* **61**, 25–31 (2015).

51. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotech* **4**, 265–270 (2009).

52. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nat Biotechnol* **26**, 1146–1153 (2008).

53. Manrao, E. A. *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol* **30**, 349–353 (2012).

54. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* **17**, 239 (2016).

55. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* **36**, 338–345 (2018).

56. Wick, R. R., Judd, L. M. & Holt, K. E. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLOS Computational Biology* **14**, e1006583 (2018).

57. Rand, A. C. *et al.* Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* **14**, 411–413 (2017).

58. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**, 201–206 (2018).

59. Biada, I., Santacreu, M. A., González-Recio, O. & Ibáñez-Escriche, N. Comparative analysis of Illumina, PacBio, and nanopore for 16S rRNA gene sequencing of rabbit's gut microbiota. *Front. Microbiomes* **4**, (2025).

60. Wagner, J. *et al.* Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. *BMC Microbiology* **16**, 274 (2016).

61. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, (2007).

62. Huson, D. H. MEGAN community edition: interactive exploration and analysis of large-scale microbiome sequencing data. *PLOS Comput. Biol.* **12**, (2016).

63. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–D745 (2016).

64. Quast, C. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, (2013).

65. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**, 610–618 (2012).

66. Cole, J. R. *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* **42**, D633–D642 (2014).

67. Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* **38**, 1079–1086 (2020).

68. Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* **12**, 902–903 (2015).

69. Franzosa, E. A. *et al.* Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* **15**, 962–968 (2018).

70. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457–D462 (2016).

71. Huerta-Cepas, J. *et al.* eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44**, D286–D293 (2016).

72. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489 (2021).

73. Proctor, L. M. *et al.* The Integrative Human Microbiome Project. *Nature* **569**, 641–648 (2019).

74. Hou, K. *et al.* Microbiota in health and diseases. *Signal Transduct Target Ther* **7**, 135 (2022).

75. Liang, Y. *et al.* Systematic Analysis of Impact of Sampling Regions and Storage Methods on Fecal Gut Microbiome and Metabolome Profiles. *mSphere* **5**, (2020).

76. Tukey, J. W. *Exploratory Data Analysis*. (Addison-Wesley, Reading, MA, 1977).

77. Tukey, J. W. The problem of multiple comparisons. in *Proceedings of the 1953 Biennial Research Conference on the Teaching of Statistics* 15–20 (Princeton University, 1953).

78. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, (2018).

79. Morgan, M., Pagès, H., Obenchain, V. & Hayden, N. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. *Bioconductor* Bioconductor https://doi.org/10.18129/B9.bioc.Rsamtools (2025).

80. Barrett, T. *et al.* data.table: Extension of 'data.frame'. CRAN (2024).
81. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, (2013).
82. Callahan, B. J. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, (2016).
83. Curry, K. D. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat. Methods* **19**, (2022).
84. Oxford Nanopore Technologies. Guppy: Oxford Nanopore basecalling software (version 6.1.5, MinKNOW 20.05.8). (2025).
85. Oxford Nanopore Technologies. EPI2ME pipeline (version 3.6.1). (2025).
86. Biosciences, P. SMRT Link (version 10.2.0.133434).
87. Felix Krueger. Trim Galore.
88. Institute, I. L. R. BMTagger.
89. Portik, D. M., Brown, C. T. & Pierce-Ward, N. T. Evaluation of taxonomic classification and profiling methods for long-read shotgun metagenomic sequencing datasets. *BMC Bioinforma.* **23**, (2022).
90. Fritz, A. CAMISIM: simulating metagenomes and microbial communities. *Microbiome* **7**, (2019).
91. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
92. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, (2016).
93. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

94. Rozas, M., Brillet, F., Callewaert, C. & Paetzold, B. MinION^{TM} nanopore sequencing of skin microbiome 16S and 16S-23S rRNA gene amplicons. *Front. Cell Infect. Microbiol.* **11**, (2022).

95. Cha, T. Gut microbiome profiling of neonates using nanopore MinION and Illumina MiSeq sequencing. *Front. Microbiol.* **14**, (2023).

96. Salonen, A. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J. Microbiol. Methods* **81**, (2010).

97. Kerkhof, L. J. Is Oxford Nanopore sequencing ready for analyzing complex microbiomes? *FEMS Microbiol. Ecol.* **97**, (2021).

98. Boer, R. Improved detection of microbial DNA after bead-beating before DNA isolation. *J. Microbiol. Methods* **80**, (2010).

99. Knudsen, B. E. Impact of sample type and DNA isolation procedure on genomic inference of microbiome composition. *mSystems* **1**, (2016).

100. You, I. & Kim, M. J. Comparison of gut microbiota of 96 healthy dogs by individual traits: breed, age, and body condition score. *Animals* **11**, (2021).

101. Söder, J. Composition and short-term stability of gut microbiota in lean and spontaneously overweight healthy Labrador retriever dogs. *Acta Vet. Scand.* **64**, (2022).

102. Thomson, P., Santibáñez, R., Rodríguez-Salas, C., Flores-Yañez, C. & Garrido, D. Differences in the composition and predicted functions of the intestinal microbiome of obese and normal weight adult dogs. *PeerJ* **10**, (2022).

103. Li, Z. Analysis and comparison of gut microbiome in young detection dogs. *Front. Microbiol.* **13**, (2022).

104. Xu, J. The response of canine faecal microbiota to increased dietary protein is influenced by body condition. *BMC Vet. Res.* **13**, (2017).

105. Matsuo, Y. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION^{TM} nanopore sequencing confers species-level resolution. *BMC Microbiol.* **21**, (2021).

106. Sinha, R. The microbiome quality control project: baseline study design and future directions. *Genome Biol.* **16**, (2015).

107. Monteiro, L., Cabrita, J. & Mégraud, F. Evaluation of performances of three DNA enzyme immunoassays for detection of Helicobacter pylori PCR products from biopsy specimens. *J. Clin. Microbiol.* **35**, (1997).

108. Flekna, G., Schneeweiss, W., Smulders, F. J. M., Wagner, M. & Hein, I. Real-time PCR method with statistical analysis to compare the potential of DNA isolation methods to remove PCR inhibitors from samples for diagnostic PCR. *Mol. Cell Probes* **21**, (2007).

109. Nechvatal, J. M. Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *J. Microbiol. Methods* **72**, (2008).

110. Li, X. Efficiency of chemical versus mechanical disruption methods of DNA extraction for the identification of oral Gram-positive and Gram-negative bacteria. *J. Int. Med. Res.* **48**, (2020).

111. Josefsen, M. H., Andersen, S. C., Christensen, J. & Hoorfar, J. Microbial food safety: potential of DNA extraction methods for use in diagnostic metagenomics. *J. Microbiol. Methods* **114**, (2015).

112. Santiago, A. Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol.* **14**, (2014).

113. Teng, F. Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci. Rep.* **8**, (2018).

114. Johnson, J. S. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **10**, (2019).

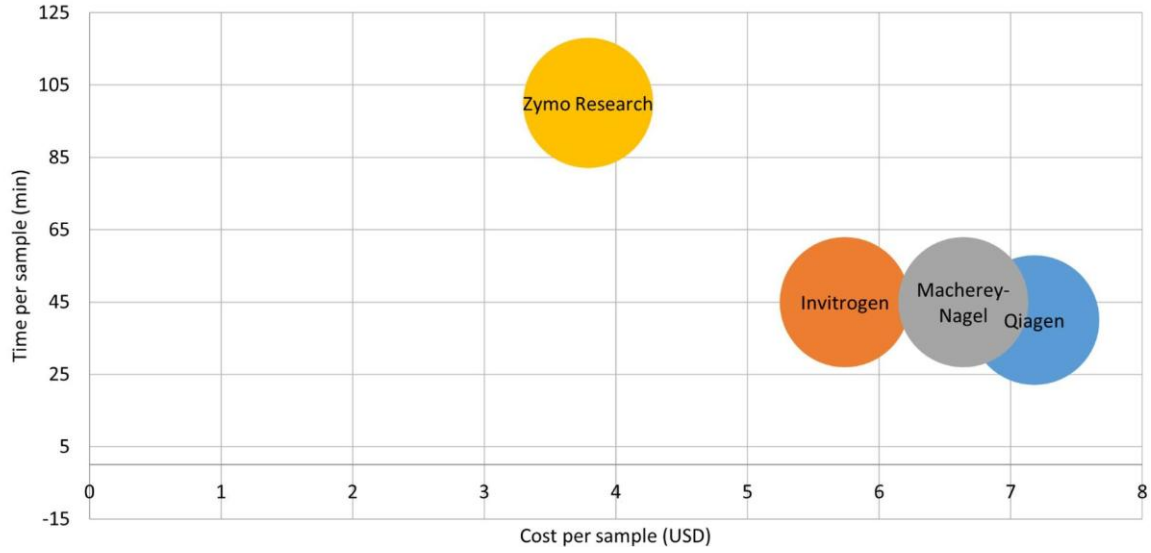
115. Walker, A. W. & Hoyles, L. Human microbiome myths and misconceptions. *Nat. Microbiol.* **8**, (2023).

8. Supplementary figures

8.1. Supplementary Figure 1

Supplementary Figure 1: Comparative analysis of cost and hands-on time for DNA isolation kits.

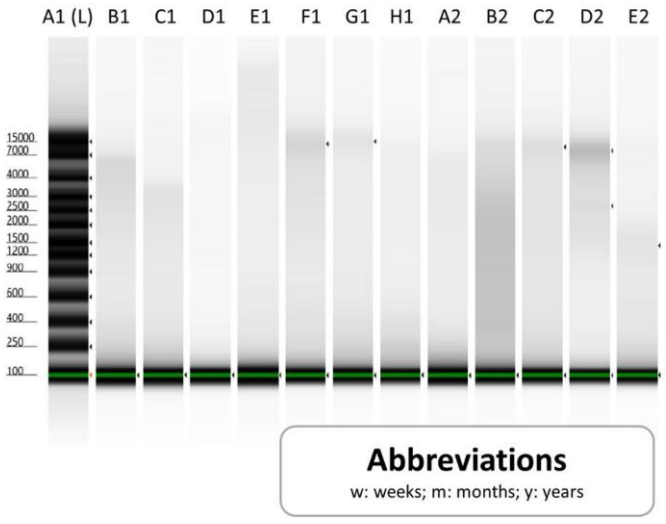
This figure contrasts the cost per sample with the hands-on time required for different DNA isolation kits, with prices presented in USD and based on the summer 2023 sales in Hungary.



8.2. Supplementary Figure 2

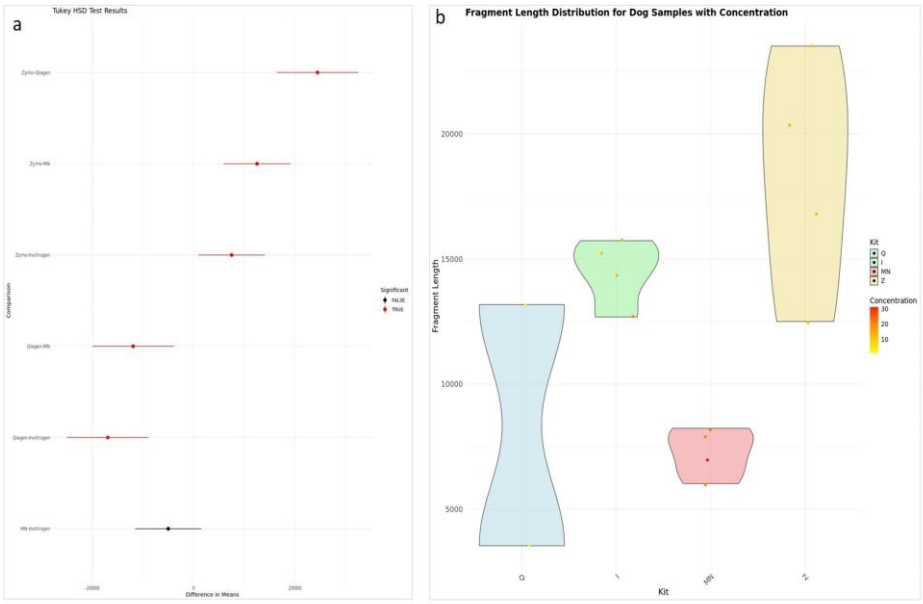
Supplementary Figure 2: Analysis of canine stool DNA extraction using Qiagen kit and Agilent ScreenTape Assay.

The figure illustrates DNA extracted from control canine fecal samples utilizing the Qiagen kit, which is then analyzed using the Agilent genomic ScreenTape assay. Lane A1 (L) represents the DNA ladder. Abbreviations are as follows: y: years, w: weeks.



	Sample	Age
B1	Female 1	6,5 y
C1	Female 1	6,5 y
D1	Male 1	6 y
E1	Male 1	6 y
F1	Female 2	4 w
G1	Female 2	5 w
H1	Male 2	5 w
A2	Male 2	5 w
B2	Male 3	7 w
C2	Male 3	5 w
D2	Female 3	4 w
E2	Female 3	5 w

8.3. Supplementary Figure 3



Supplementary Figure 3: Fragment length analysis across different DNA extraction kits: Tukey HSD test results and distribution

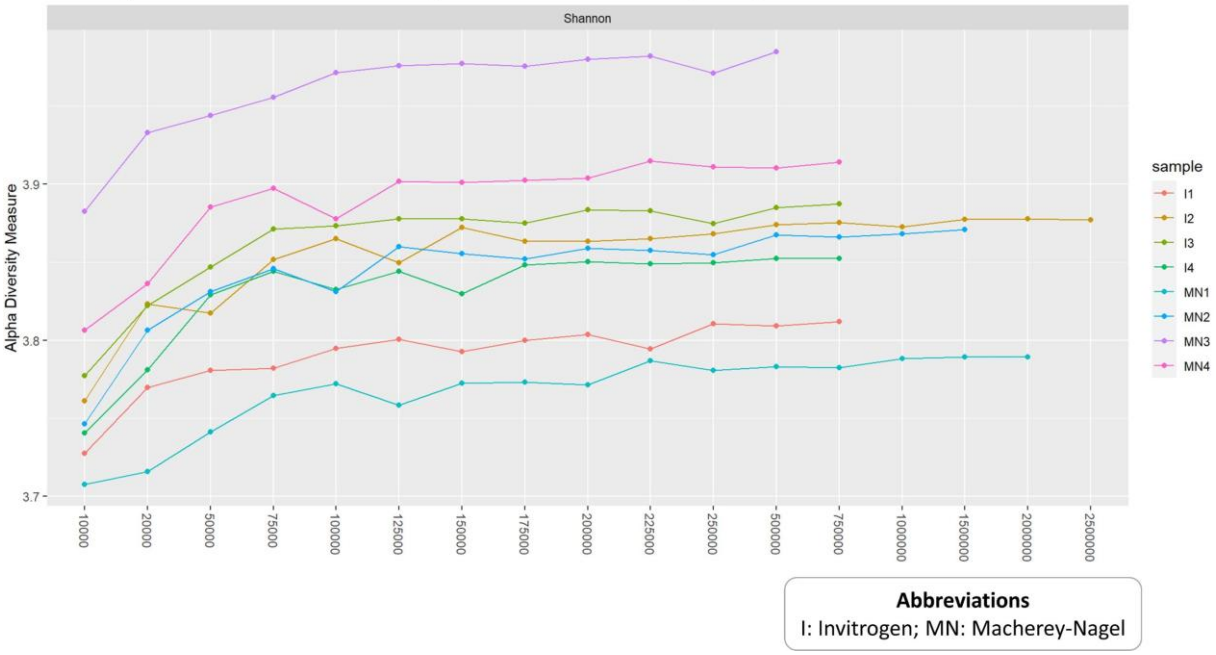
a, The results of the Tukey HSD test on DNA fragment lengths across various extraction kits are presented, highlighting the differences in mean DNA fragment lengths between pairs of extraction kits. The horizontal lines represent the 95% confidence intervals for each comparison. Points on the plot indicate the mean differences, with colors denoting statistical significance: red indicates comparisons with p-values less than 0.05, and black indicates non-significant comparisons. Significant differences are found between MN and Q, MN and Z, Q and I, Z and Q, and Z and I. No significant difference was observed between MN and I kits.

b, The distribution of DNA fragment lengths for each extraction kit is illustrated, with violin plots depicting the range and density of fragment lengths and colors indicating concentration levels. Jittered points overlaying the violin plots represent individual fragment length measurements and are colored according to concentration. This visualization helps to understand the distribution and variation of fragment lengths within each kit.

8.4. Supplementary Figure 4

Supplementary Figure 4: Correlation between downsampling and diversity.

The relationship between read count and microbial diversity is explored in this figure, underscoring the opportunity for cost-efficiency in shotgun sequencing. Even with a reduction to approximately 200,000 reads, the diversity across samples remains stable. For assessments focusing on bacterial composition rather than detection of rare species, a read count of 200,000 per sample is sufficient.



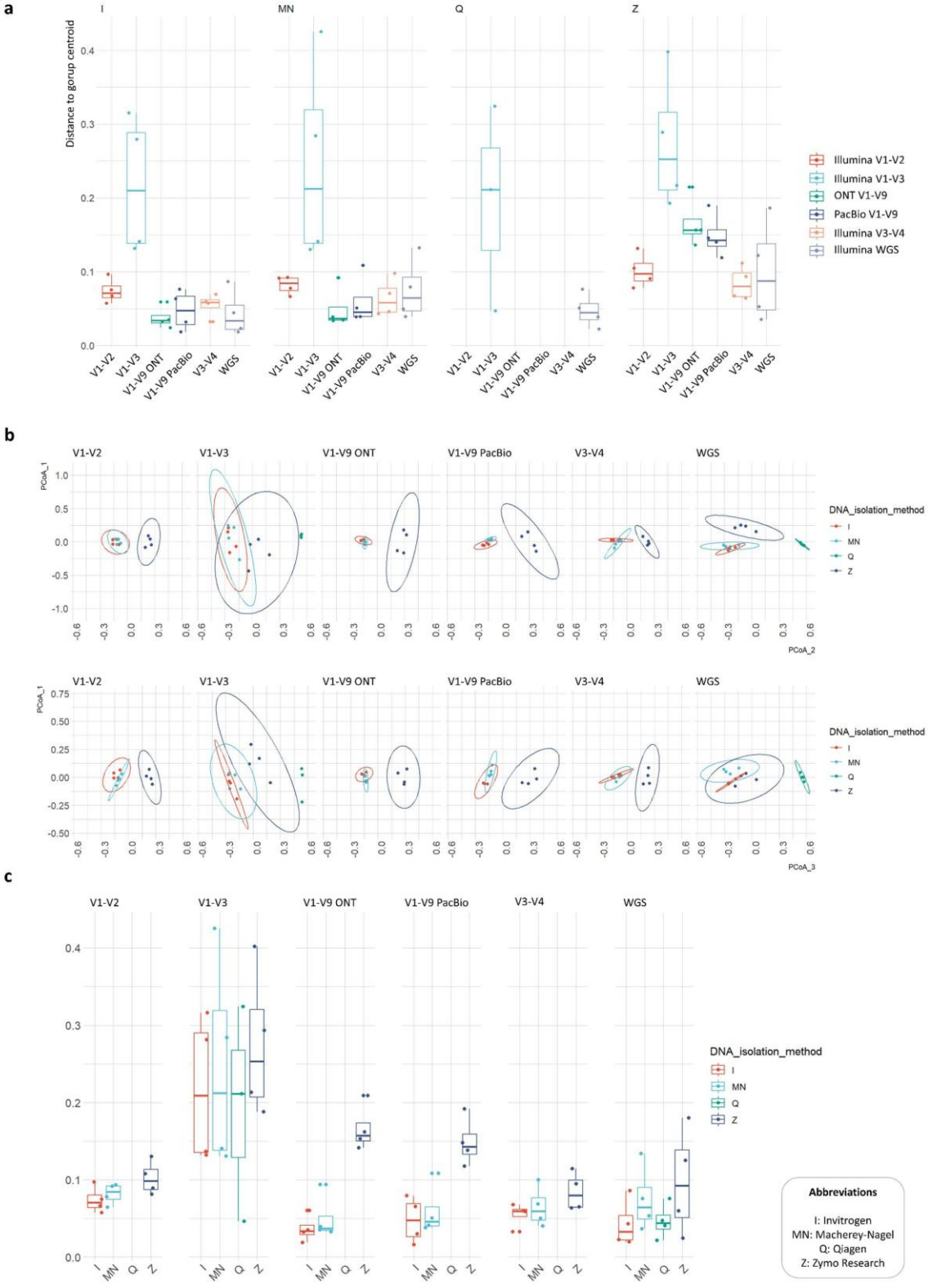
8.5. Supplementary Figure 5

Supplementary Figure 5: β-diversity assessment using Bray-Curtis distance and PERMIDISP analysis

a, The PERMIDISP results display a multivariate analysis of group dispersion homogeneity, indicating the distance to the centroid for each sample in relation to DNA isolation kits and library preparation protocols.

b, The PCoA results depict the variance and resemblance among the DNA isolation protocols for each library preparation.

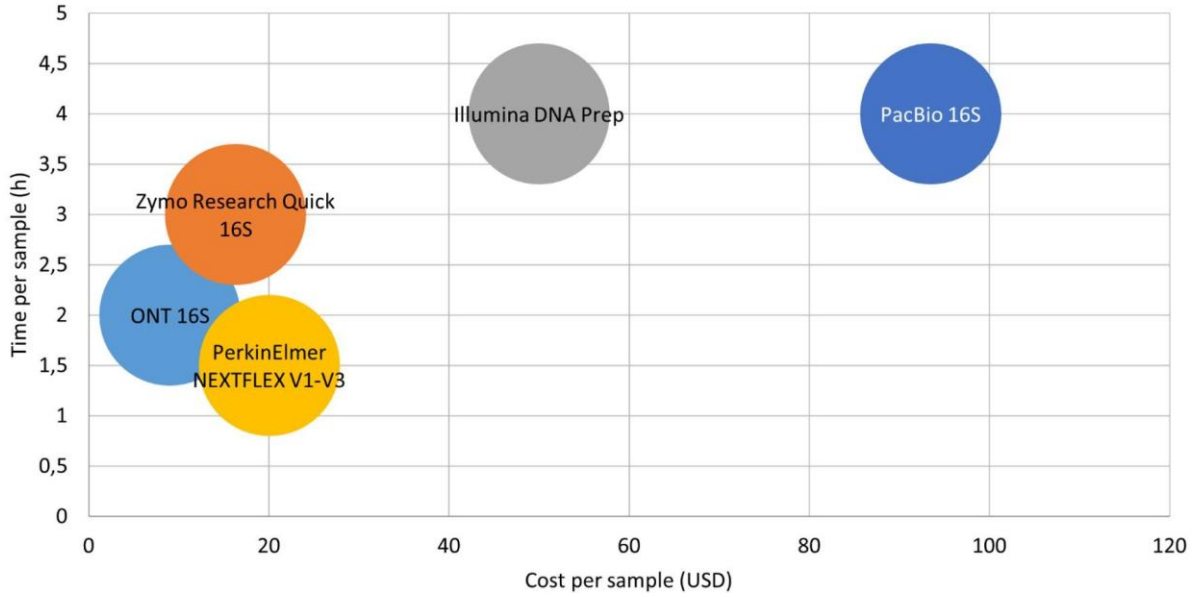
c, The results of the PERMIDISP analysis are presented, categorized by library preparation protocols and DNA isolation kits.



8.6. Supplementary Figure 6

Supplementary Figure 6: Assessment of cost and hands-on time for different library preparation kits

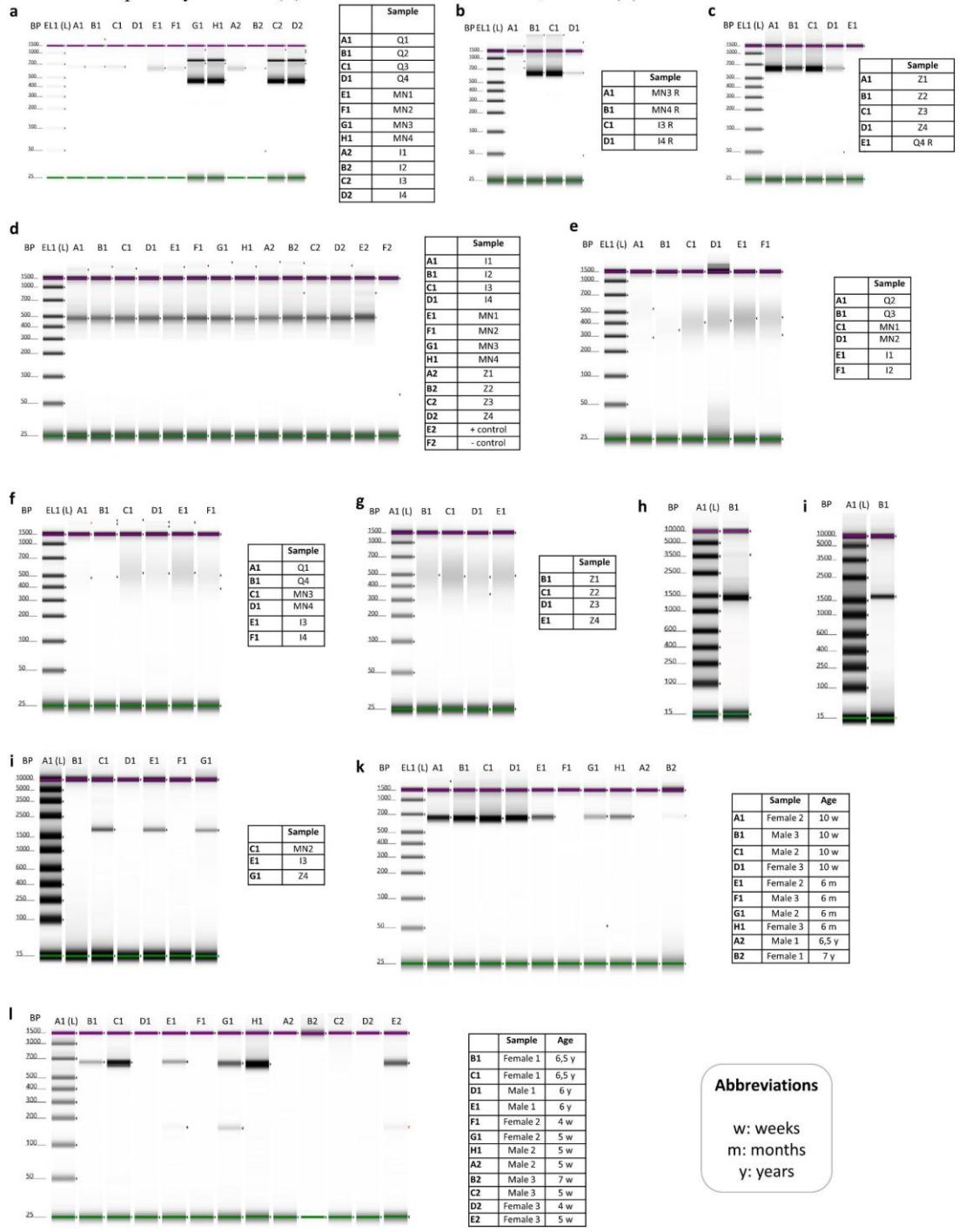
This figure contrasts the cost per sample and the hands-on time associated with various library preparation kits, denoted in USD, based on the summer 2023 sales prices in Hungary.



8.7. Supplementary Figure 7

Supplementary Figure 7: The TapeStation images of the various libraries.

a-j, represent samples from the primary test dog
k-l, indicate samples from the control dogs.
a-c, 16S V1-V3 libraries created using the PerkinElmer NEXTFLEX® Kit and analyzed with the Agilent D1000 ScreenTape assay. Lanes EL1 (L) represent electronic DNA ladders. R indicates a technical replicate (Dog stool).
d, Zymo Research V1-V2 libraries examined with the Agilent D1000 ScreenTape assay. Lane EL1 (L) represents the electronic DNA ladder.
e-g, Shotgun (mWGS, Illumina DNA Prep) libraries derived from canine stool samples.
h-j, V1-V9 library pools from dog stool prepared for ONT (h) and PacBio (i) sequencing. Panel h displays a pool of sixteen libraries, with four each prepared from DNA isolated using Q, I, MN, and Z DNA purification kits for ONT sequencing. Panel i shows twelve PacBio libraries (four each from I, MN, and Z-prepped DNAs). The PCR products were approximately 1.5 kb in size. Abbreviations: MN: Macherey-Nagel, I: Invitrogen, Z: Zymo Research.
k-l, V1-V3 libraries made using the PerkinElmer NEXTFLEX® 16S Amplicon-Seq Kit and analyzed with the Agilent D1000 ScreenTape assay. Lane E1 (L) is the electronic DNA ladder, and A1 (L) is the ladder.



	Sample
A1	Q1
B1	Q2
C1	Q3
D1	Q4
E1	MN1
F1	MN2
G1	MN3
H1	MN4
A2	I1
B2	I2
C2	I3
D2	I4

	Sample
A1	MN3 R
B1	MN4 R
C1	I3 R
D1	I4 R

	Sample
A1	Z1
B1	Z2
C1	Z3
D1	Z4
E1	Q4 R

	Sample
A1	I1
B1	I2
C1	I3
D1	I4
E1	MN1
F1	MN2
G1	MN3
H1	MN4
A2	Z1
B2	Z2
C2	Z3
D2	Z4
E2	+ control
F2	- control

	Sample
A1	Q2
B1	Q3
C1	MN1
D1	MN2
E1	I1
F1	I2

	Sample
A1	Q1
B1	Q4
C1	MN3
D1	MN4
E1	I3
F1	I4

	Sample
B1	Z1
C1	Z2
D1	Z3
E1	Z4

	Sample
C1	MN2
E1	I3
G1	Z4

	Sample	Age
A1	Female 2	10 w
B1	Male 3	10 w
C1	Male 2	10 w
D1	Female 3	10 w
E1	Female 2	6 m
F1	Male 3	6 m
G1	Male 2	6 m
H1	Female 3	6 m
A2	Male 1	6,5 y
B2	Female 1	7 y

	Sample	Age
B1	Female 1	6,5 y
C1	Female 1	6,5 y
D1	Male 1	6 y
E1	Male 1	6 y
F1	Female 2	4 w
G1	Female 2	5 w
H1	Male 2	5 w
A2	Male 2	5 w
B2	Male 3	7 w
C2	Male 3	5 w
D2	Female 3	4 w
E2	Female 3	5 w

Abbreviations

w: weeks
m: months
y: years

8.8. Supplementary Figure 8

Supplementary Figure 8: Ratio of Gram-positive and Gram-negative bacteria in canine stool as a result of various laboratory methods

The barplots show the ratio of Gram(+) and Gram(-) bacteria in each sample, according to library preparation protocols and DNA isolation methods. Zymo's results indicate a slight overrepresentation of Gram negatives compared to the Macherey-Nagel and Invitrogen methods, with the Qiagen method showing a significant overrepresentation.

8.9. Supplementary Figure 9

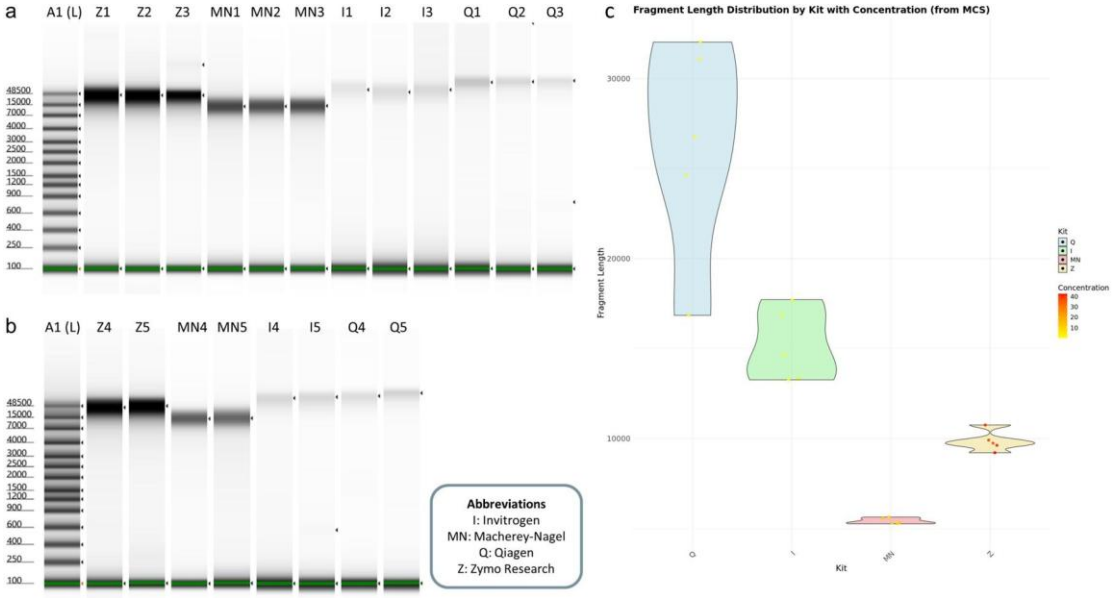
Supplementary Figure 9: Comparing DNA yields and fragment lengths from MCS samples

a, b The images depict the Agilent genomic DNA ScreenTape assay results for DNA extracted from MCS. Lanes A1 (L) represent DNA ladders. The Invitrogen and Qiagen samples showed lower DNA amounts compared to others, with Zymo yielding the highest amount. Despite Zymo Kit being designed for HMW DNA isolation, Invitrogen and Qiagen yielded longer DNA fragments from Zymo MCS samples.

a, First three replicates.

b, Two subsequent replicates.

c, The distribution of DNA fragment lengths for each extraction kit is depicted through a violin plot. The shape and width of each violin indicate the range and density of fragment lengths, with colors representing the concentration levels. Overlaying jittered points show individual measurements and are colored according to concentration.



8.10. Supplementary Figure 10

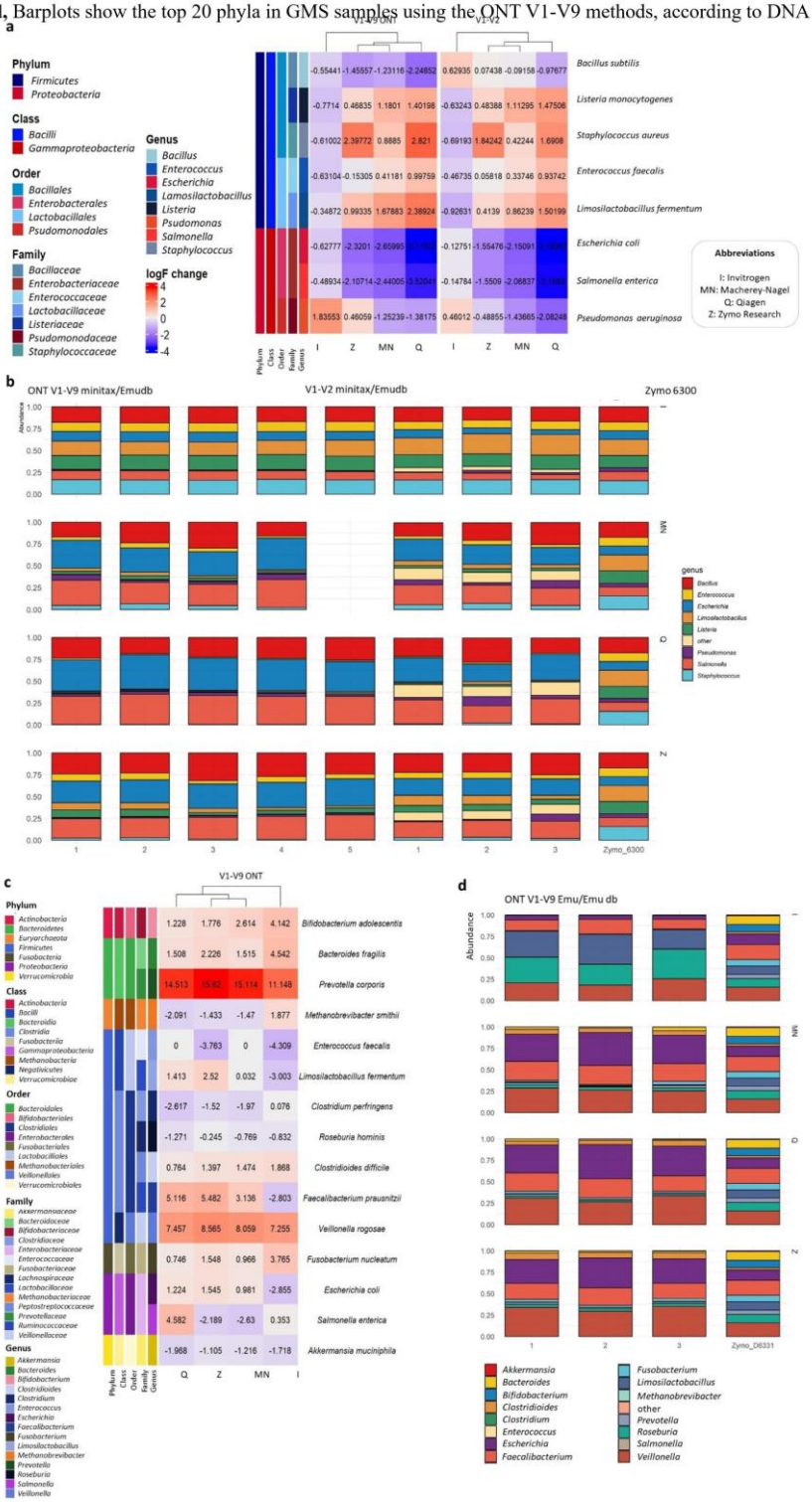
Supplementary Figure 10. Comparative analysis of DNA isolation and library preparation protocols in MCS and GMS using *Emu* with the *Emu database*

a, Heatmap of differences between the theoretical and experimental abundances on according to each species in the MCS. The abundance values were compared to the theoretical values provided by Zymo and *log2* fold differences were estimated and are shown within the boxes. Deeper blue colors indicate lower experimental values compared to the theoretical, while more red colors indicate higher experimental values.

b, Barplots showing the top 20 phyla in MCS samples using the Illumina V1-V2 and ONT V1-V9 methods, according to DNA isolation kit.

c, Heatmap of differences between the theoretical and experimental abundances on according to each species in the GMS. The abundance values identified by the Emu tool were compared to the theoretical values provided by Zymo Research and *log2* fold differences were estimated and are shown within the boxes. Darker blue colors represent lower experimental values compared to the theoretical, while dark red colors indicate higher experimental values.

d, Barplots show the top 20 phyla in GMS samples using the ONT V1-V9 methods, according to DNA isolation kit.



8.11. Supplementary Figure 11

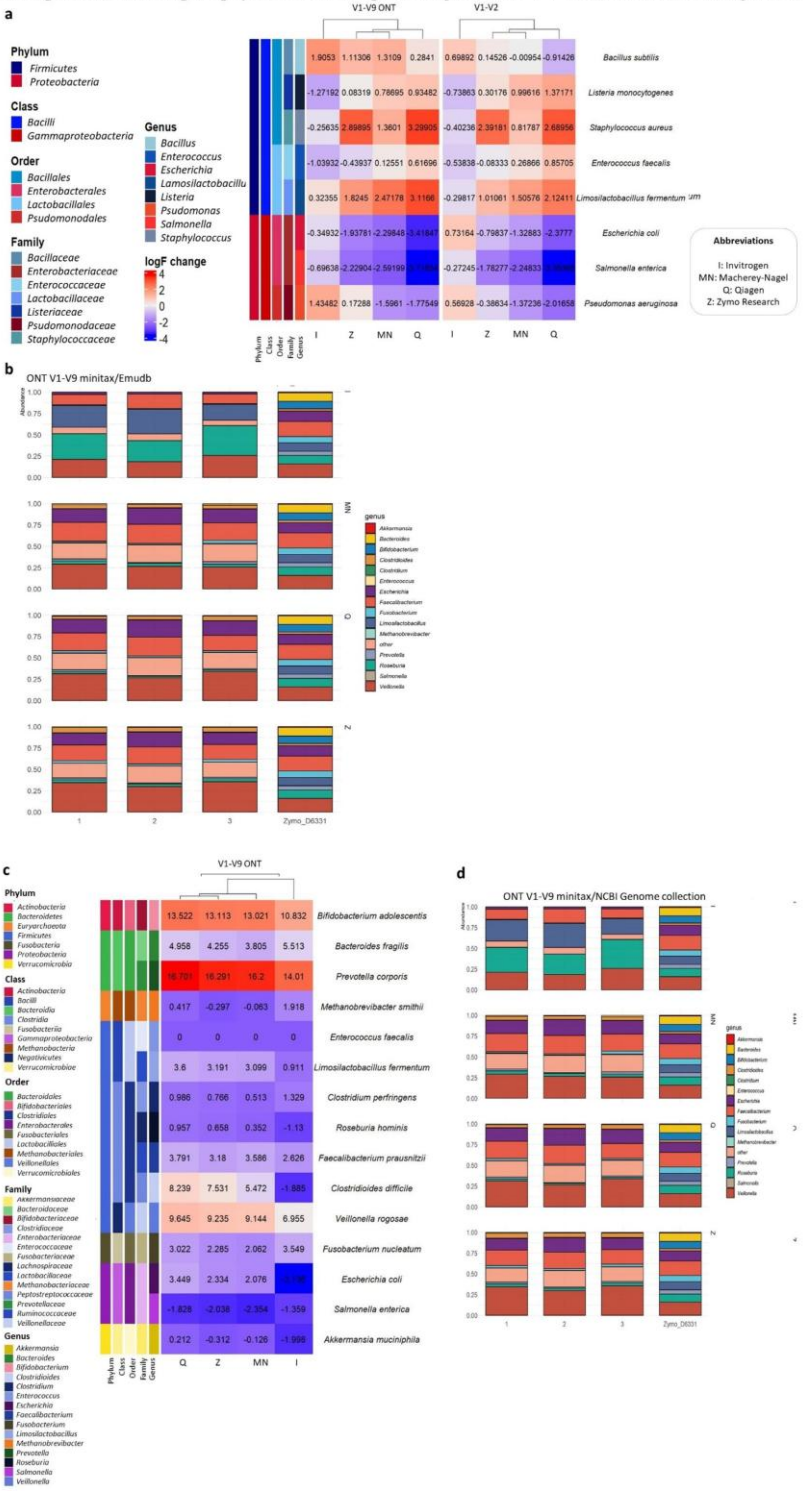
Supplementary Figure 11. Comparative analysis of DNA isolation and library preparation protocols in MCS and GMS using *minitax* with the *NCBI genome collection database*

a, Heatmap of differences between the theoretical and experimental abundances on according to each species in the MCS. The abundance values were compared to the theoretical values provided by Zymo and *log2* fold differences were estimated and are shown within the boxes. Deeper blue colors indicate lower experimental values compared to the theoretical, while more red colors indicate higher experimental values.

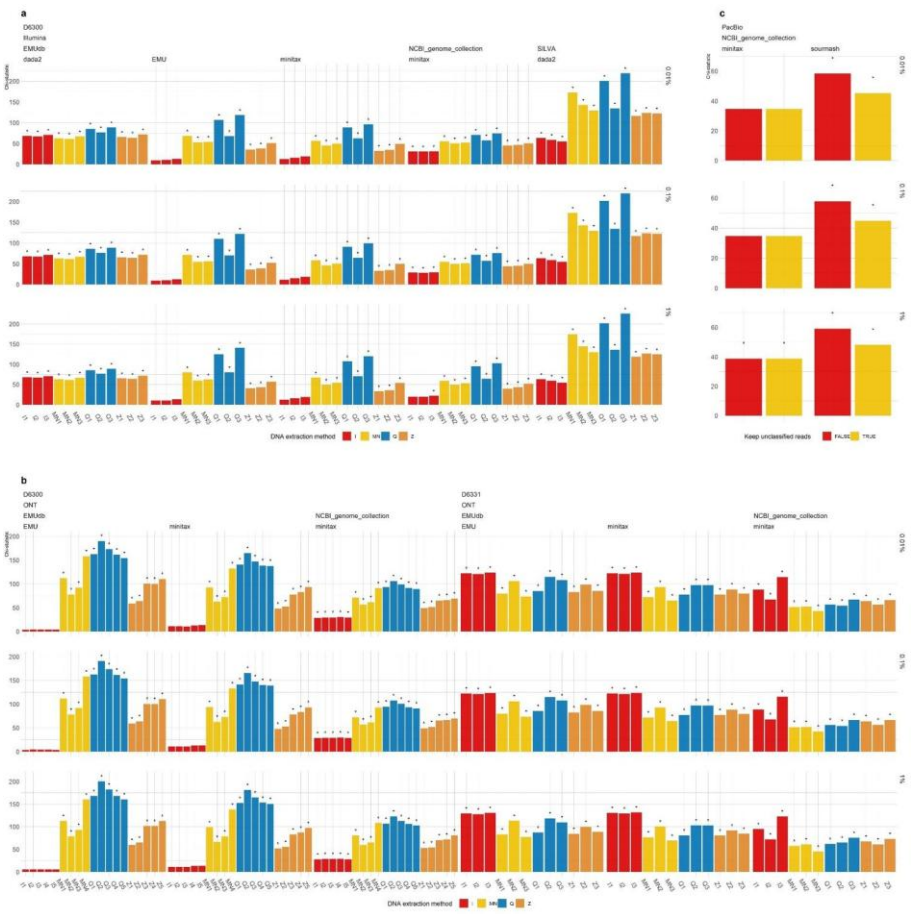
b, Barplots showing the top 20 phyla in GMS samples using the ONT V1-V9 methods, according to DNA isolation method.

c, Heatmap of differences between the theoretical and experimental abundances on according to each species in the GMS. The abundance values were compared to the theoretical values provided by Zymo Research and *log2* fold differences were estimated and are shown within the boxes. Darker blue colors represent lower experimental values compared to the theoretical, while dark red colors indicate higher experimental values.

d, Barplots show the top 20 phyla in GMS samples using the ONT V1-V9 methods, according to DNA isolation method.



8.12. Supplementary Figure 12



Supplementary Figure 12: Evaluating minitax: a comparison with other methods based on Pearson's correlations.
This figure compares minitax with other methods based on the Pearson's correlations between theoretical and observed compositions (r2 values).
a, Comparison with Emu on ONT V1-V9 Sequencing of Zymo D3600 MCS data.
b, Comparison with DADA2 and Emu on Illumina V1-V2 Sequencing of Zymo D3600 MCS.
c, Comparison with sourmash on PacBio HiFi WGS of Zymo MCS D6331.

9. Supplementary methods

9.1. QIAGEN QIAamp Fast DNA Stool Mini Kit

Canine stool: A 200 mg stool sample was placed in a 2 ml microcentrifuge tube and kept on ice. InhibitEX Buffer was added to each sample, and they were mixed using a vortex until completely homogenized. The large stool pieces were removed by centrifugation. Six hundred µl of the supernatant was combined with 25 µl of proteinase K and 600 µl of Buffer AL. The mixture was thoroughly vortexed, then heated to 95°C for 5 min (this is a 95°C lysis incubation for 5 minutes, diverging from the 70°C recommended by the QIAamp kit). Six hundred µl of 100% ethanol was added to the lysate and then mixed. Six hundred µl of the lysate was loaded onto a QIAamp spin column and centrifuged at 20,000 x g for 1 min. The QIAamp spin column was placed into a new 2 ml tube. The remainder of the lysate was then loaded onto the column. After centrifugation, 500 µl of Buffer AW1 was added to the column. This was followed by a 20,000 x g centrifugation for 1 min, then we discarded the collection tube. Next, 500 µl of Buffer AW2 was added to the column, which was then placed into a new collection tube. A full-speed centrifugation (20,000 x g) was performed for 3 min. To avoid any Buffer AW2 carryover, the spin column was set in a fresh 2 ml collection tube and the samples were spun down at full speed for 3 min. The spin columns were then transferred to new Eppendorf tubes, and 100 µl of Buffer ATE was directly loaded onto the QIAamp membrane. After letting it incubate at room temperature for 1 min, a centrifugation (20,000 x g) step was carried out for 1 min to elute the DNA in 50 µl in elution buffer, followed by storing the DNA solution at -20°C.

MCS and GMS samples: 75 µl from the mixtures was used for DNA purification, following the same protocol as for the dog sample, with the DNA being eluted in a final volume of 50 µl.

9.2. Invitrogen PureLink™ Microbiome DNA Purification Kit

Canine stool: A 200 mg sample was combined with 600 µl of S1 Lysis Buffer in the Bead Tube (provided in the kit) and the mixture was homogenized by vortexing. Subsequently the 100 µL of S2 Lysis Enhancer (from the kit) was added, and the samples were vortexed again. The mixtures were then incubated at 65°C for 10 minutes. For homogenization, the samples were subjected to bead beating using a vortex mixer with horizontal agitation at maximum speed for 10 min. The samples were then centrifuged at 14,000 × g for 5 min. Afterwards, 400 µL of the supernatant was transferred to a new Eppendorf tube and mixed with 250 µL of S3 Cleanup Buffer. The samples were centrifuged again at 14,000 × g for 2 min, and 500 µL of the resulting supernatant was transferred to a clean tube and mixed with 900 µL of S4 Binding Buffer. After

brief vortexing, 700 µL of the sample mixture was loaded onto a spin column and centrifuged at 14,000 × g for 1 min. The spin column was then placed in a new tube, and the remaining sample mixture was loaded onto it for an additional 1 min centrifugation. The spin column was subsequently placed in a clean collection tube, and 500 µL of S5 Wash Buffer was added, followed by centrifugation at 14,000 × g for 1 min. To remove any residual S5 Wash Buffer, a second centrifugation was carried out at 14,000 × g for 30 sec. Finally, DNA was eluted from the spin columns using 100 µL of S6 Elution Buffer and stored at -20°C.

MCS and GMS samples: 75 µl of the microbial mixes was utilized for DNA extraction, following the same steps as described above. The DNA was eluted in 50 µl S6 Elution Buffer.

9.3. Macherey-Nagel NucleoSpin DNA Stool Mini kit

Canine stool: DNA isolation was performed using 200 mg of fecal sample, which was transferred to a Macherey-Nagel Bead Tube Type A, and then 850 µL Buffer ST1 was added. The mixtures were shaken horizontally for 3 seconds before being placed in a heat incubator. Subsequently, the samples were incubated for 5 min at 70 °C, agitated on a Vortex-Genie® 2 at full speed and room temperature for 10 min, and then centrifuged for 3 min at 13,000 x g. Six hundred µl of the supernatant was transferred to a new 2 ml tube, and 100 µl of Buffer ST2 was added and briefly vortexed. The mixtures were incubated for 5 min at 4 °C and then centrifuged for 3 min at 13,000 x g. Five hundred fifty µl of the lysate was loaded onto a NucleoSpin® Inhibitor Removal Column and centrifuged for 1 min at 13,000 x g. The Inhibitor Removal Column was discarded. Two hundred µl of Buffer ST3 was added to the samples, which were then mixed. Seven hundred µl of the sample mixture was loaded onto a NucleoSpin® DNA Stool Column and centrifuged for 1 min at 13,000 x g. The column was placed in a new tube. The sample was washed four times: first, 600 µl of Buffer ST3 was added to the NucleoSpin® DNA Stool Column and centrifuged for 1 min at 13,000 x g. The column was then placed in a new tube, and 550 µl of Buffer ST4 was added. After a 1 min centrifugation at 13,000 x g, the column was placed into a new tube, and 700 µl of Buffer ST5 was added. Following a brief vortexing, the samples were centrifuged for 1 min at 13,000 x g. The column was placed in a new tube, and 700 µl of Buffer ST5 was added, followed by a 1 min centrifugation at 13,000 x g. The flow- through was discarded, and the column was placed back onto the tube. The silica membrane of the column was dried by a 2 min centrifugation at 13,000 x g. One hundred µl of Buffer SE was loaded onto the center of the column, and the DNA was eluted by centrifugation for 1 min at 13,000 x g. DNA samples were stored at -20°C.

MCS and GMS samples: 75 µl of the MCS or the GMS mixture was utilized for DNA isolation, with the sample being eluted in a final volume of 50 µl.

9.4. Zymo Research Quick-DNA™ HMW MagBead Kit

Canine stool: One hundred mg of the fecal sample was used as initial weight and resuspended in 200 µl of DNA/RNA Shield™, followed by incubation at room temperature (20-30°C) on a tube rotator for 5 minutes. Subsequently, 33 µl of MagBinding Beads were added to each sample, mixed and placed on a shaker for a 10 min. The sample was then placed on a magnetic stand until a clear separation of the beads and the solution was observed, after which the supernatant was removed. For the washing step, 500 µl of Quick-DNA™ MagBinding Buffer was added and the beads were resuspended and shaken for 5 min. The sample was returned to the magnetic stand, and the supernatant was discarded. Next, 500 µl of DNA Pre-Wash Buffer was added, and the beads were resuspended. The sample was placed on the magnetic stand again, and the supernatant was discarded. In the subsequent step, 900 µl of g-DNA Wash Buffer was added and mixed, and the entire liquid was transferred to a new tube. The magnetic stand was used to separate the beads from the solution, and the supernatant was discarded. These washing steps were repeated once more. The sample were left to air dry for 20 min. For the elution step, 50 µl of DNA Elution Buffer was added to each sample. After mixing, the solution was incubated at room temperature for 5 min. Finally, the sample was placed back on the magnetic stand until the beads separated from the solution. The eluted DNA was carefully transferred to a new tube and stored at -20ºC for future use.

MCS and GMS samples: 75 µl of the mix was used for DNA purification, following the same protocol as described for the dog sample. The DNA was eluted in a final volume of 50 µl.

9.5. LIBRARY PREPARATION

9.5.1. From partial regions of the 16S rRNA gene

9.5.1.1. Zymo Research V1-V2

DNA from canine stool: Ten µl of Quick-16S™ qPCR Premix were mixed with 4 µl of Quick-16S™ Primer Set V1-V2 and 4 µl of ZymoBIOMICS® DNase/RNase Free Water. Additionally, 2 µl of DNA samples (2.5 ng/µl) were added. PCR was conducted in a Verity Thermal Cycler (Applied Biosystems) as per the Zymo Research Manual (Supplementary Data 9). After amplification, 1 µl of Reaction Clean-up Solution was added to the samples, which were then incubated at 37°C for 15 min. The reactions were terminated by heating to 95°C for

10 min, and the samples were subsequently cooled to 4°C. Next, 10 µl of Quick-16S™ qPCR Premix and 4 µl of ZymoBIOMICS® DNase/RNase Free Water were combined. Index primers (2 µl each from ZA5 and ZA7, Supplementary Data 10 for detailed pairs and sequences) and 2 µl of the amplified DNA were also measured into the mixture. Barcoded PCR reactions were performed as recommended by the manual (Supplementary Data 11). For purification of the PCR products, Select-a-Size MagBeads were used. First, the MagBeads were resuspended by shaking, and then 16 µl of the Select-a-Size MagBeads were mixed with each sample. The mixture was incubated at room temperature for 5 min and then placed on a magnetic rack for 3 to10 min. The supernatant was discarded and the beads were washed twice with 200 µl of DNA Wash Buffer. The samples were removed from the magnet and were incubated for 3 min at room temperature to eliminate all traces of buffer. Libraries were eluted in 25 µl of DNA Elution Buffer and stored at -20°C until further use.

DNA from MCS samples: 2 ng/µl DNA was used to prepare the V1-V2 libraries from microbial mixture

9.5.1.2. Zymo Research V3-V4

The protocol used was the same as described in the 'Zymo Research V1-V2' section with the following modifications: In the initial PCR step, V3-V4 primers were utilized. Supplementary Data 9 provides details of the barcoded primers used, including the pairs and their sequences.

9.5.1.3. PerkinElmer NEXTFLEX® 16S V1-V3 Amplicon-Seq Kit for Illumina

Genomic DNA, having concentrations between 1.6 ng and 36 ng as specified in Supplementary Data 12, was diluted using Nuclease-free Water to maintain a total volume no greater than 36 µL. Subsequently, 12 µL of NEXTflex™ PCR Master Mix and 2 µL of the 16S V1-V3 PCR I Primer Mix were added to the solution. The final reaction volume was adjusted to 50 µL. First amplification step of the PCR cycling was carried out using the settings outlined in Supplementary Data 13. PCR cleanup: fifty µL AMPure XP Beads was added to each sample. After mixing, the samples were incubated at room temperature for 5 min. Then, using a magnetic stand, the samples were left until the supernatant clarified. The supernatant was then discarded, and the beads were washed twice with 200 µL of freshly prepared 80% ethanol. Next, the samples were air-dried for 3 min and resuspended in 38 µL of Resuspension Buffer. After a further incubation of 2 minutes at room temperature, 36 µL from the clear supernatant was transferred to fresh tubes. This sample was then subjected to PCR amplification with the addition of 12 µL of NEXTflex™ PCR Master Mix and 2 µL of NEXTflex™ PCR II Barcoded

Primer Mix. The procedure was executed following the guidelines specified in Supplementary Data 14. PCR cleanup was carried out in accordance with the purification after the first PCR.

9.5.2. For the analysis of full-length 16S rRNA gene sequencing

9.5.2.1. ONT Rapid Sequencing 16S Barcoding Kit (SQK-RAB204)

Ten ng of high molecular weight genomic DNA (in a 10 ul volume) was used for library preparation from both the canine, MCS and GMS samples. DNA isolated using the QIAGEN kit did not meet this criterion. The input DNA was mixed with 14 µl of Nuclease-free water (Invitrogen), 1 µl of 16S Barcode (1 µM; Supplementary Data 15)) and 25 µl of LongAmp Taq 2X master mix (New England Biolabs). PCR Amplification of the samples was carried out according Supplementary Data 16. Amplified DNA samples were transferred to clean 1.5 ml Eppendorf DNA LoBind tubes and mixed with 30 µl of resuspended AMPure XP beads (Beckman Coulter). Next, they were incubated on a Hula mixer for 5 minutes at room temperature. Tubes were placed on a magnetic rack then the supernatant was discarded. The beads were washed with 200 µl of freshly prepared 70% ethanol. Ethanol was removed and the washing was repeated once. After air drying of the beads, samples were removed from the magnet and beads were resuspended in 10 µl of 10 mM Tris-HCl pH 8.0 with 50 mM NaCl. After 2 min incubation at room temperature, samples were placed on the magnet. Ten µl of the clean supernatant, containing the ONT libraries, was transferred to a new Eppendorf DNA LoBind tube. Barcoded libraries were pooled in equal molar ratio and then, 1 µl of RAP was added. The reaction was incubated for 5 minutes at room temperature. One hundred fmoles were loaded on a MinION flow cell.

9.5.2.2. PacBio Full-Length 16S Library Preparation Using SMRTbell Express Template Prep Kit 2.0 Sequel IIe System ICS v10.0 / Sequel II Chemistry 2.0 / SMRT Link v10.0

For each sample, 1.5 µL of PCR-grade water and 12.5 µL of 2X KAPA HiFi HotStart ReadyMix were mixed. Subsequently, 3 µL of barcoded forward primer solution (2.5 µM, sequences in Supplementary Data 17) was added. This was followed by the addition of 3 µL of the respective reverse primer solution (Supplementary Data 17) and 5 µL of the DNA sample. DNA amplification was carried out according to the parameters listed in Supplementary Data 18.

9.5.3. Shotgun sequencing

9.5.3.1. Illumina DNA Prep

Sixty ng of DNA (in 30 µl) was used as total input per sample. Ten µl Tagmentation Buffer 1 (TB1) was mixed with 10 µl Bead-Linked Transposomes (BLT), and 20 µl of this mixture was added to a DNA sample. The mixture was incubated at 55°C for 15 min and then held at 10 °C. Following this, 10 µl of Tagment Stop Buffer (TSB) was added to the sample and gently mixed. The samples were incubated at 37°C for 15 min, and then kept at 10°C. Next, the samples were placed on a magnetic stand for 3 min, and the supernatant was discarded. The sample was removed from the magnet and 100 µl of Tagment Wash Buffer (TWB) was slowly added directly onto the beads. The sample was placed back on the magnetic stand, the supernatant was discarded, and the wash step was performed again. A mixture of 20 µl of Enhanced PCR Mix (EPM) and 20 µl of nuclease-free water was prepared, and 40 µl of this mixture was added to the washed beads. Index adapters (i5 and i7, 5 µl each; Supplementary Data 19) were added, and PCR was carried out according to Supplementary Data 20. After amplification, the libraries were cleaned up. First, the samples were placed on a magnet for approximately 5 min. Forty-five µl of supernatant from each PCR product was transferred to a new tube. Forty µl of nuclease- free water and 45 µl of Sample Purification Beads (SPB) were added to the supernatant and the samples were mixed at 1600 rpm for 1 min. They were incubated at room temperature for 5 min. After this step, the samples were placed on a magnet, and 15 µl of SPB were added to new tubes. Next, 125 µl of supernatant from each sample was added to the tubes containing 15 µl of undiluted SPB. The samples were mixed at 1600 rpm for 1 min, and then incubated at room temperature for 5 min. The supernatant was discarded, and the washing step was carried out twice with 200 µl of freshly prepared 80% ethanol. The sample was stored on magnetic stand for 30 secs, and then the ethanol was removed. After the second washing step, the pellet was air dried. Next, 32 µl RSB was added to the beads, and they were mixed and incubated for 2 min. Finally, the sample was placed on a magnetic stand for 2 min, and the supernatant containing the prepared library was transferred to a new tube.

9.6. Data availability

All our sequencing data have been submitted to the ENA under the accession PRJEB59610. The datasets from other sources were downloaded from the ENA with the following accession numbers: PRJNA783735, PRJNA678365, PRJNA871395. Supplementary Data files (1–21): figshare, https://doi.org/10.6084/m9.figshare.2723262673. In addition, data used to generate figures can be found under the project's github repository:

https://github.com/Balays/Microbiome-Method-Comparison, specifically:—Figure 4 g: Fig4.G-H_MCM_all_Adiversity_estimates.tsv - Figure 4 i: Fig4_statistics_data.rds—Figure 5 a: Fig_5A_BrayCurtis_distances.tsv—Figure 5 b: Fig_5B_PCoA_data.tsv—Figure 5 c: Fig_5C_Heatmap_data.tsv - Figure 5 d: ./Dog_Feces/Microbial_abundances for each taxon level—Figure 5 e: DogFeces_all_methods_PS.rds - Figure 7: sig.freq.tsv and l2F_diff.dt.tsv separately for each group—Figure 7: ./MCM/…/sig.freq.tsv and l2F_diff.dt.tsv separately for each group - Figure 9 a-b: ./MCM/…/Detection statistics.tsv for each each group—Figure 9 c: .//Portik_etal_2022/ Detection statistics.tsv - Figure 9 d: ./CAMISIM/Detection statistics.tsv. In addition, phyloseq objects that can be imported into R and contain the microbial abundances for the respective samples are also available: - MCM_ZymoD6300_all_PS.rds - MCM_ZymoD6331_all_PS.rds - DogFeces_all_methods_PS.rds. The data used to generate Fig. 4e, f are available in Supplementary Data 3, the data for Fig. 6 can be found in Supplementary Data 6, while the data for generate Fig. 8 are in Supplementary Data 21.

9.7. Code availability

minitax: https://github.com/Balays/minitaxStatistics: https://github.com/Balays/Microbiome-Method-Comparison other in-house scripts - https://github.com/Balays/Microbiome-Method-Comparison - https://github.com/gabor-gulyas/Technical-article-downsample DADA2: https://benjjneb.github.io/dada2/ Emu: https://gitlab.com/treangenlab/emu Trim Galore: https://github.com/FelixKrueger/TrimGalore BMTagger: https://hpc.ilri.cgiar.org/bmtagger-software

10. Suppleementary datas

10.1. Suppleementary Data 1

Supplementary Data 1. This table summarizes the previously published articles focusing on the comparison of library prep kits or bioinformatic methods used at various stages of metagenomic analyses.

a. human and mammalian stool

#	sample	DNA isolation kits	Method	Library preparation kit	Approach	Sequencer	read length	Bioinformatics	Reference
1	**human stool and pure culture**	FastDNAR kit (Bio 101) NucleospinR C + T kit1 (Macherey-Nagel) Quantum PrepR Aquapure Genomic DNA isolation kit QIAampR DNA stool minikit Boom et al., 1990	PCR						71. McOrist, A. L., Jackson, M. & Bird, A. R. A comparison of five methods for extraction of bacterial DNA from human faecal samples. *J Microbiol. Methods* **50**, 131-9 (2002).
2	**human stool**	RNA/DNA Mini kit (Qiagen) QIAamp DNA Stool Mini kit (Qiagen) Fecal DNA Isolation kit (Mo BIO)	Real-time PCR						72. Nechvatal, J. M. et al. Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *J. Microbiol. Methods* . **72**, 124-32 (2008).
3	**human stool**	Differential Centrifugation and Lysis (Apajalahti et al., 1998) Promega Genomic Wizard DNA Purification Kit (Promega) Repeated Bead Beating (Yu and Morrison QiaAmp DNA Stool Mini Kit (Qiagen)	Real-time PCR, microarray						58. Salonen, A. et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J. Microbiol. Methods* **81**, 127-134 (2010).
4	**stool from 6 months to 4 years kids with diarrhoea**	QIAamp DNA stool MiniKit with or without Mini BeadBeater 8 (BioSpec Products Inc) or TissueLyser system (Qiagen Retsch GmbH)	PCR		V3				73. Smith, B., Li, N., Andersen, A. S., Slotved, H.C. & Krogfelt, K. A. Optimising bacterial DNA extraction from faecal samples: comparison of three methods. *Open Microbiol. J.* **5**, 14-7 (2011).
5	**human stool**	NucliSENS® easyMag® (BioMérieux) - semi-automatic QIAamp DNA Stool Mini Kit (Qiagen) - manual	PCR		V2-V3				74. Mirsepasi, H. et al. Microbial diversity in fecal samples depends on DNA extraction method: easyMag DNA extraction compared to QIAamp DNA stool mini kit extraction. BMC *Res. Notes* **7**, 50 (2014).

6	**human stool**	sample from outer an innel layer and from homogenized stool, with or without bead-beating	pyrosequencing		V4				9. Santiago, A. et al. Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol*. **14,** 112 (2014).
7	**Human stool**	American Human Microbiome Project European MetaHIT project	SRS		WGS	Illumina HiSeq 2000	2x100	Mocat Blast+	14. Wesolowska-Andersen, A. et al. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* **2**, 19 (2014).
8	**human stool and synthetic DNA mock community**	NA	SRS	Illumina Nextera XT kit Illumina TruSeq DNA PCR-free kit KAPA Biosystems Hyper Prep PCR system KAPA Biosystems Hyper Prep PCR-free system	WGS	Illumina HiSeq (v4 chemistry)			15. Jones, M. B. et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc. Natl. Acad. Sci. U S A*. **112**, 14024-9 (2015).
9	**Sewage water, soil & human stool, biopsy**	seems like an inhouse protocol DNA isolation kit (Qiagen, Germany) MagNA pure, Roche Diagnostics, Swizerland	SRS		V1-V3 V1-V5	454 GS FLX+ pyrosequencer			16. Bag, S. et al. An Improved Method for High Quality Metagenomics DNA Extraction from Human and Environmental Samples. *Sci. Rep*. **6**, 26775 (2016).
10	**human stool**	seven most commonly used (2017!) kit with modifications (altogether 21 extraction protocols) PSPStool (Invitek) PowerSoil (Mobio) EZNAstool (Omega Bio Tek) Maxwell (Promega) QIAampStool Minikit (Qiagen) G'Nome (Bio101) FastDNAspinSoil (MP-Biomedicals) MagNAPureIII (Roche)	SRS		WGS	Illumina HiSeq	2x100		17. Costea, P.I. et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol*. **35,** 1069-1076 (2017).

11	**human stool**	TianLong Stool DNA/RNA Extraction Kit (Xi'an TianLong Science and Technology Co., Ltd.) with OR without bead-beating QIAamp DNA Stool mini kit (Qiagen) with OR without bead-beating QIAamp PowerFecal DNA Isolation kit (Qiagen)	SRS	16S Metagenomic Sequencing Library Preparation Illumina	V3-V4	Illumina MiSeq	2x300	QIIME	75. Lim, M. Y., Song, E. J., Kim, S. H., Lee, J. & Nam, Y. D. Comparison of DNA extraction methods for human gut microbial community profiling. *Syst. Appl. Microbiol* . **41**, 151-157 (2018).
12	**human stool**	QIAamp DNA Mini Kit (Qiagen) includig bead-beating PowerFecal® DNA Isolation Kit (MO BIO)	SRS	PCR amplified	V3-V4	Illumina MiSeq	2x300	QIIME, UCHIME	76. Szopinska, J. W. et al. Reliability of a participant-friendly fecal collection method for microbiome analyses: a step towards large sample size investigation. *BMC Microbiol* . **18**, 110 (2018).
13	**human stool and germ-free mice feces spiked with bacterial or fungal strains**	QIAamp DNA Stool Mini Kit (Qiagen), and PureLinkTM Microbiome DNA Purification Kit (TermoFisher) Fecal DNA MiniPrepTM Kit (ZymoResearch) NucleoSpin® DNA Stool Kit (Macherey-Nagel) IHMS protocol Q (QIAamp DNA Stool Kit with bead beating)	qPCR and SRS	16S Metagenomic Sequencing Library Preparation protocol	V3-V4, ITS1F+ ITS2	Illumina MiSeq	2x300	QIIME	77. Fiedorová, K. et al. The Impact of DNA Extraction Methods on Stool Bacterial and Fungal Microbiota Community Recovery. *Front. Microbiol* . **10**, 821 (2019).

14	**paleofeces human and dog!!!**	the Human Microbiome Project standard protocol using the PowerSoil kit (Qiagen) an aDNA-optimized modified MinElute (Qiagen) protocol for bone extraction following Dabney et al. (2013) Phenol-chloroform + modified MinElute protocol Split modified MinElute protocol HMP protocol + modified MinElute protocol	SRS	NEBNext DNA Library Prep Master Set	WGS	Illumina HiSeq	2x100	QIIME	78. Hagan, R. W., et al. Comparison of extraction methods for recovering ancient microbial DNA from paleofeces. *Am. J. Phys. Anthropol.* **171**, 275-284 (2020).
15	**human stool and 36 bacterial strains represent microbes prevalent in the human body sites**	DNeasy Blood and Tissue kit (Qiagen; bacteria) PowerSoil DNA Isolation Kit (Mobio; human)	SRS LRS	TruSeq Nano DNA HT kit (Illumina) SMRTbell 1.0 Template Prep Kit (Pacific Biosciences)	WGS WGS V1-V9	Illumina MiSes and HiSeq (bacteria) Illumina NextSeq (human) PacBio RSII (bacteria)	2x150	MUSCLE	12. Johnson, J. S. et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **10**, 5029 (2019).
16	**capuchin monkey stools**	QIAamp DNA Stool Mini Kit (Qiagen) PowerSoil DNA Isolation kit (Mo Bio)	two-step PCR protocol amplicon library preparation on the Fluidigm Access Array	MyTaq HS Red Mix (Bioline) Accustart II PCR ToughMix (Quantiabio)	V3-V4 V4-V5	Illumina MiSeq		QIIME	31. Mallott, E. K., Malhi, R. S. & Amato, K. R. Assessing the comparability of different DNA extraction and amplification methods in gut microbial community profiling. *Access Microbiol.* **1**, e000060 (2019).
17	**eight different biological specimens, including human stool, salive, tissues etc.** **zymo microbial community standard**	Quick-DNA fecal/soil microbe kit (Zymo Research) QIAampStool Minikit (Qiagen) MagNA Pure 96 (Roche Diagnostics)	SRS		V4	Illumina NextSeq	2x150	QIIME 2, NG-Tax 0.4	18. Ducarmon, Q. R., Hornung, B. V. H., Geelen, A. R., Kuijper, E. J. & Zwittink, R. D. Toward Standards in Clinical Microbiota Studies: Comparison of Three DNA Extraction Methods and Two Bioinformatic Pipelines. *mSystems* **5**, e00547-19 (2020).

18	**human stool & mock community (Zymo)**	Mag-Bind® Universal Metagenomics Kit (Omega Bio tek) DNeasy PowerSoil Kit (Qiagen)	SRS shotgun	KAPA Hyper Prep Kit (Kapa HiFi) with different sample inputs TruePrep DNA Library Prep Kit V2 (Vazyme Biotech) with different sample	WGS	Illumina HiSeq 4000	2x250 2x350	alpha-, beta-diversity, PCA	79. Peng, Z. et al. Comparative Analysis of Sample Extraction and Library Construction for Shotgun Metagenomics. *Bioinform. Biol. Insights* . 14, 1177932220915459 (2020).
19	**human stool**	QIAamp DNA Stool Mini Kit (Qiagen) QIAamp PowerFecal Pro DNA Kit (Qiagen) QIAamp DNA Stool Mini Kit (Qiagen) with additional bead-beating step [5]	SRS	16S Metagenomic Sequencing Library Preparation Illumina	V3-V4	Illumina MiSeq	2x300	DADA2 plugin [15] within QIIME2	80. Lim, M., et al. Evaluation of fecal DNA extraction protocols for human gut microbiome studies. *BMC Microbiol* . **20**, 212 (2020).
20	**simulation and human stool**	PureLink Microbiome DNA Purification Kit (Invitrogen) Qiamp PowerFecal DNA Kit (Qiagen) ZimoBiomics DNA Mini Kit (Zymo Research) Power Microbiome RNA/DNA Isolation Kit (Mo Bio)	SRS LRS		WGS WGS	ONT MinION Illumina NovaSeq SOLiD	2x150		81. Alili, R. et al. Exploring Semi-Quantitative Metagenomic Studies Using Oxford Nanopore Sequencing: A Computational and Experimental Protocol. *Genes* **12**, 1496 (2021).
21	**human stool - 200 healthy Japanese**	DNEasy Power Soil Kit (Qiagen)	SRS qPCR	16S library preparation protocol provided by Illumina	V1-V2 V3-V4	Illumina MiSeq	2x250 2x300	QIIME1 & QIIME2 UCLUST, DADA2	82. Kameoka, S., et al. Benchmark of 16S rRNA gene amplicon sequencing using Japanese gut microbiome data from the V1–V2 and V3–V4 primer sets. *BMC Genomics* **22**, 527 (2021).
22	**dog stool (datasets)**	PowerFecal DNA Isolation Kit (MoBio)	SRS		WGS V3-V4	Illumina HiSeq 2500 Ilumina MiSeq	2x125 2x300	MetaPhlAn2 DADA2, QIIME2	20. Lewis, S. et al. Comparison of 16S and whole genome dog microbiomes using machine learning. *BioData Min* . **14**, 41 (2021).
23	**human stool**	Mechanical-Enzymatic Lysis method: Phenol: Chloroform: Isoamyl Alcohol	LRS		WGS	ONT MinION		EPI2ME	83. Sahu, S. et al. Fecal genomic DNA extraction method impacts outcome of MinION based metagenome profile of tuberculosis patients. *medRxiv* **11**, 15.21266154 (2021).

24	**human stool**	Extrap Soil DNA Kit Plus ver.2 (NIPPON STEEL Eco-Tech Corporation) FastDNA SPIN Kit for Feces (MP Biomedicals) ISOSPIN Fecal DNA Kit (Nippon Gene) MagAttract PowerMicrobiome RNA/DNA EP Kit MORA-EXTRACT kit (Kyokuto Pharmaceutical QIAmp PowerFecal Pro DNA kit (Qiagen) Quick DNA Fecal/Soil microbe Miniprep Kit (Zymo Research)	SRS	Accel NGS 2S Plus DNA Library Kit (Swift Biosciences) TruSeq DNA PCR-Free Library Prep Kit KAPA HTP Library Preparation Kit (Roche) KAPA HyperPrep Kit PCR-free (Roche) TruSeq Nano DNA Library Prep Kit NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) QIAseq FX DNA Library Kit NEBNext Ultra II FS DNA Library Prep Kit Nextera DNA Flex Library Prep Kit SMARTer ThruPLEX DNA-Seq Kit (Takara Bio)		Illumina NextSeq 500	2x150 bp	Demultiplex: BBmap v38.46 Qc: fastp v0.20.0 for human samples: BMTagger v3.101 Data analysis and visualisation: R v4.0.2 Data handling: dplyr v1.0.2 Visualisation : ggplot2 v3.3.2 Metagenome assembly: MEGAHIT v1.2.9. Statistics: Quast v5.0.0 Read annotate: kraken v2.0.8 For OTU taxonomy profiles: mOTUs2 v2.5.1	19. Tourlousse, D. M. et al. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. *Microbiome* **9,** 95 (2021).
25	**human stool**	QIAamp DNA Stool Minikit (Qiagen) PSP Spin Stool DNA Plus Kit (Invitek) MoBio PowerSoil DNA Isolation Kit (Mo Bio)	SRS		V1-V3 V1-V2	454 Titanium pyrosequenc 454 FLX pyrosequenc ing		QIIME	84. Wu, G.D., C. et al. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol* . **10**, 206 (2010).
26	**human stool & saliva, conjuctiva, bile, sputum, plaque and water**	DNeasy PowerSoil Pro (Qiagen) QiAamp DNA Microbiome Kit (Qiagen) ZymoBIOMICS DNA Miniprep Kit (Zymo Researc)	SRS LRS	MGIEasy Universal DNA Library Prep Set	WGS WGS	DNBSEQ-G400 ONT MinION	2x100		85. Rehner, J., et al. Systematic Cross-biospecimen Evaluation of DNA Extraction Kits for Long- and Short-read Multi-metagenomic Sequencing Studies. *Genomics Proteomics Bioinformatics* . **20**, 405-417 (2022).

27	**10 diverse sample types (human body sites and environment, including stool)**	MagAttract PowerSoil DNA Isolation Kit (Qiagen) MagAttract PowerSoil Pro DNA Isolation Kit (Qiagen) Norgen Stool DNA Isolation Kit (Norgen Biotek) MagMAX Microbiome Ultra Nucleic Acid Isolation Kit (Applied Biosystems) NucleoMag Food kit (Macherey-Nagel) ZymoBIOMICS 96 MagBead DNA Kit (Zymo Research)	SRS	NexteraXT	V4 ITS shallow WGS	Miseq Illumina HiSeq	2*150	QIIME2	21. Shaffer, J. P. et al. A comparison of six DNA extraction protocols for 16S, ITS and shotgun metagenomic sequencing of microbial communities. *Biotechniques* . **73**, 34-46 (2022).

b. bird stool

1	**chicken stool**	tissue kit genomic DNA template stool mini kit	Real-time PCR						60. Flekna, G., Schneeweiss, W., Smulders, F. J. M., Wagner, M. & Hein, I. Real-time PCR method with statistical analysis to compare the potential of DNA isolation
2	**chicken stool**	Kit for Feces Kit for Soil MiniPrep DNA Stool Stool Mini Kit DNA Isolation PowerSoil DNA Isolation Genomic DNA miniMAG	Real-time PCR						63. Josefsen, M, H., Andersen, S. C., Christensen, J. & Hoorfar, J. Microbial food safety: Potential of DNA extraction methods for use in diagnostic metagenomics. *J. Microbiol. Methods* **114**, 30-4 (2015).

c. other human tissues

1	**human colonic tissue**	DNA/RNA method of protocol method as	microarray			chip			86. Ó Cuív, P. et al. The effects from DNA extraction methods on the evaluation of microbial diversity associated with human colonic tissue. *Microb. Ecol* . **61**, 353-62 (2011).
2	**Biopsies from six anatomic regions in**	Stool Mini Kit Mini Kit	SRS		V2	pyrosequenc		Greengenes	87. Momozawa, Y. et al. Characterization of bacteria in biopsies of colon and stools by high throughput sequencing of the V2
3	**pediatric bronchoalveolar lavage & microbial mock community**	trimethylammo (saline) Tissue Kit MoBio PowerSoil DNA Isolation Kit (MoBio)	SRS		V8-V9	pyrosequenc	~ 250	QIIME	88. Willner, D. et al. Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS One* . **7**, e34605 (2012).

4	**human lung tissue**	Mini kit and QIAamp orm:Isoamyl orm:Isoamyl step and Phenol:Chlorof orm:Isoamyl alcohol step and Bead-beating and QIAamp DNA Mini kit (QIAGEN)	SRS	Index	V3-V4 ITS	MiSeq	2x300	QIIME	89. Pérez-Brocal, V. et al. Optimized DNA extraction and purification method for characterization of bacterial and fungal communities in lung tissue samples. *Sci. Rep* . **10**, 17377 (2020).
5	**human biopsies (Colon)**	Microbiome Technique developed by the group	SRS qPCR SRS		V3-V4 WGS	NovaSeq Illumina NovaSeq	2x250 2x150	Uparse Mothur	90. Bruggeling, C. E. et al. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. Microbiologyopen. **10**, e1191 (2021).
6	**human urine samples**	BiOstic and Tissue PowerSoil UltraClean Maxwell RSC Purefood GMO and Authentication (Promega)	SRS	amplifeid	V4	MiSeq	2x150	DADA2	91. Karstens, L. et al. Benchmarking DNA isolation kits used in analyses of the urinary microbiome. Sci. Rep. **11**, 6186 (2021).
7	**human upper airways (nose, saliva, pharynx) & mock**	Mini Kit PowerViral Microbiome Stool DNA ZymoBIOMIC S DNA Kit (Zymo Research)	SRS	DNA sample	WGS	MiSeq	2x300	orX2	92. Mancabelli, L. et al. Guideline for the analysis of the microbial communities of the human upper airways. *J. Oral. Microbiol* . **14**, 2103282 (2022).
8	**human urine**	Bacteremia kit Microbiome kit (Beckman- and Tissue kit MagNA Pure Compact Kit (Roche) - all the optional lysis steps were performed	SRS	amplified	V1–V3 V3-V4 V4-V5 V6-V8	MiSeq	2x300	SILVA,	93. Vendrell, J. A. et al. Determination of the Optimal Bacterial DNA Extraction Method to Explore the Urinary Microbiota. *Int. J. Mol. Sci* . **23**, 1336 (2022).
9	**human urine**	Pre–Cell Lysis: Mechanical With Enzymatic Lysis (lytic enzyme solution (Qiagen) and	LRS		mNGS	MinION			94. Zhang, L. et al. Comparison Analysis of Different DNA Extraction Methods on Suitability for Long-Read Metagenomic Nanopore Sequencing. *Front. Cell. Infect. Microbiol* . **12**, 919903 (2022).

d. other tissues from animals

1	**cow and sheep rumen**	extraction method [32] Spin Kit with Spin Kit with matrix (Bio- based method chloroform chloroform, chloroform chloroform, chloroform Stool DNA Kit, Stool DNA Kit, DNA Stool beating plus MiniPrep	pyrosequencing qPRC		rRNA	Titanium		QIIME	95. Henderson, G. et al. Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One* . **8**, e74787 (2013).
2	**Chicken caeca and crop**	procedure	SRS	DNA Library Metagenomic	WGS V3-V4	MiSeq MiSeq	2x150 bp 2x150 bp	MG-RAST Silva SSU	96. Durazzi, F. et al. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci. Rep.* **11**, 3030

d. plant

1	**Arabidopsis thaliana, corn and soybean**	Plant PCR Kit DNA Isolation kit (Qiagen) Fungal/Bacteria	SRS	amplified	V4 V3-V4 V5-V7	MiSeq	2x250	QIIME2	97. Giangacomo, C., Mohseni, M., Kovar, L. & Wallace, J. G. Comparing DNA Extraction and 16S rRNA Gene Amplification Methods for Plant Associated

e. bacteria strains and cultures

1	Genbank-50 strains	chloroform	point detection specified, most		V2, V6			MEGALIGN	30. Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland,
2	single	LC DNA III	rtPCR		(not				24. de Boer, R. et al. Improved
3	**mock microbial community**	**genomic**	SRS	Nextera XT Mondrian Illumina's MALBAC	WGS?	HiSeq 2000	2*150		98. Bowers, R.M., et al. Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community. *BMC*
4	**bacterial cocktail mixes**	HMW (Macherey-S DNA PowerFecal Pro kit (Claremont on Moss et al.:	LRS		WGS	ONT			99. Gand, M., et al. Comparison of 6 DNA extraction methods for isolation of high yield of high molecular weight DNA suitable for shotgun metagenomics Nanopore sequencing to detect bacteria. *BMC Genomics* **24**, 438 (2023).

f. water and other nonliving source

1	**Lake Taihu's water (China)**	Water DNA Kit (OMEGA,	SRS	regions were amplified from omic Sequencing Library Preparation with some minor modifications.	V3 V4 V6	Ion Torrent PGM		Statistical significance was tested in the differential taxa among V3, V4, and V6 using SPSS18.0 software	100. Zhang, J. Evaluation of different 16S rRNA gene V regions for exploring bacterial diversity in a eutrophic freshwater lake. *Sci. Total. Environ* . **618**, 1254-1267 (2018).
2	**Lake Baikal's water (Russia)**	treatment,	SRS	Ultra II DNA	V2-V3 V3-V4	MiSeq			22. Bukin, Y. et al. The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci. Data* **6,** 190007 (2019).
3	**Equatorial Pacific & North Atlantic - water**	MOBIO PowerSoil DNA Isolation Kit (Mo Bio)	SRS	V regions were apmlified	V4 V6	Illumina MiSeq	2x250 & 2x150		23. Kerrigan, Z., Kirkpatrick, J. B. & D'Hondt, S. Influence of 16S rRNA Hypervariable Region on Estimates of Bacterial Diversity and Community Composition in Seawater and Marine Sediment

4	**Activated Sludge, Biofilm, and Anaerobic Digestate**	FastDNA Spin kit for Soil (MP Biomedicals) MicrobiomeTM Purification Kit (ThermoFisher Scientific FavorPrep Soil DNA Isolation Mini Kit (Favorgen Biotech	capillary electrophoresis						101. Florczyk, M., Cydzik-Kwiatkowska, A., Ziembinska-Buczynska, A. & Ciesielski A. Comparison of Three DNA Extraction Kits for Assessment of Bacterial Diversity in Activated Sludge, Biofilm, and Anaerobic Digestate. *Appl. Sci.* **12**, 9797 (2022).
5	**Soil**	not specified	SRS+LRS	TruSeq Nano SMRTbell Template Prep Kit	400 bp WGS, 20 000 bp	HiSeq 2000 PacBio Sequel	not specif	rk, BLAST, PB assembly:MetaFlye, IL: SPAdes	102. Xu, G. et al. Combined assembly of long and short sequencing reads improve the efficiency of exploring the soil metagenome. *BMC Genomics* **23**, 37 (2022).
6	**On-Site Wastewater**	DNA Isolation	SRS	Ultra™ II	WGS V4	HiSeq MiSeq	2x150 2x150		103. de Vries, J. et al. Comparative Analysis of Metagenomic

10.2. Supplementary Data 2

Supplementary Data 2. Applied wet lab kits and bioinformatic pipelines

a. DNA extraction kits used in our study. HMW: High-molecular weight DNA.

Company	Kit	Bead-beating	HMW
Qiagen	QIAamp Fast DNA Stool Mini Kit	No	No
Macherey-Nagel	NucleoSpin DNA Stool Mini kit	Yes	No
Invitrogen	PureLink™ Microbiome DNA Purification Kit	Yes	No
Zymo Reserach	Quick-DNA™ HMW MagBead Kit	No	Yes

b. Library preparation kits utilized in this project. mWGS: metagenomic whole-genome sequencing.

Company	Kit	16S rRNA region	mWGS
Zymo Research	Quick-16S NGS Library Prep Kit	Partial (V1-V2)	
Zymo Research	Quick-16S NGS Library Prep Kit	Partial (V3-V4)	
PerkinElmer	NEXTFLEX® 16S V1-V3 Amplicon-Seq Kit	Partial (V1-V3)	
Illumina	DNA Prep Kit		Yes
Oxford Nanopore Technologies	16S Barcoding Kit	Full-lenght (V1-V9)	
Pacific Biosciences	Full-Length 16S Library Preparation Using SMRTbell Express Template Prep Kit 2.0	Full-lenght (V1-V9)	

c. The applied bioinformatic tools

Program	Partial 16S rRNA gene	Full-length 16S rRNA gene	Shotgun
DADA2	Yes		
Emu	Yes	Yes	
EPI2ME		Yes	
Kaiju			Yes
sourmash			Yes
minitax	Yes	Yes	Yes

10.3. Supplementary Data 3

Supplementary Data 3.

a. The table summarizes the details of DNA purification from canine sample (the main tested dog), the obtained yield and quality of DNA

Sample	Initial amount (g)	Elution (ul)	gDNA cc. (ng/ul)		Length of the highest detected peak (bp)	Lenght of the longest detected DNA fragment (bp)	Average of the highest detected peak (bp)	Average of the longest detected DNA fragment (bp)
Qiagen	0,2	100	Q1	2,25	-	-	8359	9361
			Q2	1,89	13178	15062		
			Q3	1,96	-	-		
			Q4	1,16	3540	3660		
Invitrogen	0,2	100	I1	14,4	14293	26235	14469	26347
			I2	12	15162	26503		
			I3	20,8	12685	24945		
			I4	11,2	15734	27705		
Macherey-Nagel	0,2	100	MN1	20	6023	18277	7255	21325
			MN2	23	8235	23199		
			MN3	22	7867	21278		
			MN4	31	6896	22547		
Zymo Research	0,1	50	Z1	11,2	23514	35526	18286	33867
			Z2	13,2	16785	34863		
			Z3	13,9	20340	33513		
			Z4	11,3	12505	31568		

b. DNA purification from canine sample (control dogs for validation) and

Sample	Initial amount (g)	Elution (ul)	gDNA cc. (ng/ul)	
Qiagen	0,2	100	Female 1	1,85
			Male 1	4,96
			Female 2	41,2
			Female 3	26,8
			Male 2	43,4
			Male 3	19

c. Statistical Comparisons of DNA Fragment Lengths Across Different DNA Isolation Kits. This table shows the differences in DNA fragment lengths between pairs of DNA isolation kits, along with their corresponding p-values and significance levels. The "difference" column represents the mean difference in fragment lengths between kit pairs. The "p-value" column indicates the statistical significance of each comparison, with values less than 0.05 considered significant. The "Significance" column denotes whether the difference is statistically significant ("yes") or not ("no"). In all cases where significance is indicated, the 95% confidence intervals exclude zero, confirming the statistical significance of the differences observed. Conversely, no significant difference was detected between Macherey-Nagel and Invitrogen kits, as evidenced by a confidence interval that includes zero.

DNA isolation kit pairs	difference	p-value	Significance
Macherey-Nagel - Qiagen	-11,964.25	=0.005	yes
Macherey-Nagel - ZymoResearch	12,542.25	=0.001	yes
Macherey-Nagel - Invitrogen	-5,021.75	0.154	no
Zymo-Research - Invitrogen	7,520.50	= 0.025	yes
Zymo-Research - Qiagen	24,506.50	< 0.001	yes
Qiagen - Invitrogen	-16,986.00	< 0.001	yes

10.4. Supplementary Data 4

Supplementary Data 4. Sample-wise Shannon and Simpson indices.

***In vivo* samples**

sample	host	DNA extraction method	Target (V-region)	platform	program	database	sample_name	Shannon	Simpson
Tot! Illumina V1-V2 minitax NCBI.genome.collection I1	dog ('Tot!')	Invitrogen (I)	V1-V2	Illumina	minitax	NCBI genome collection	I1	4,008	0,947
Tot! Illumina V1-V2 minitax NCBI.genome.collection I2	dog ('Tot!')	Invitrogen (I)	V1-V2	Illumina	minitax	NCBI genome collection	I2	3,904	0,951
Tot! Illumina V1-V2 minitax NCBI.genome.collection I3	dog ('Tot!')	Invitrogen (I)	V1-V2	Illumina	minitax	NCBI genome collection	I3	4,020	0,948
Tot! Illumina V1-V2 minitax NCBI.genome.collection I4	dog ('Tot!')	Invitrogen (I)	V1-V2	Illumina	minitax	NCBI genome collection	I4	3,887	0,945
Tot! Illumina V1-V2 minitax NCBI.genome.collection MN1	dog ('Tot!')	Macherey-Nagel (MN)	V1-V2	Illumina	minitax	NCBI genome collection	MN1	4,059	0,952
Tot! Illumina V1-V2 minitax NCBI.genome.collection MN2	dog ('Tot!')	Macherey-Nagel (MN)	V1-V2	Illumina	minitax	NCBI genome collection	MN2	3,965	0,952
Tot! Illumina V1-V2 minitax NCBI.genome.collection MN3	dog ('Tot!')	Macherey-Nagel (MN)	V1-V2	Illumina	minitax	NCBI genome collection	MN3	4,045	0,950
Tot! Illumina V1-V2 minitax NCBI.genome.collection MN4	dog ('Tot!')	Macherey-Nagel (MN)	V1-V2	Illumina	minitax	NCBI genome collection	MN4	3,962	0,953
Tot! Illumina V1-V2 minitax NCBI.genome.collection Z1	dog ('Tot!')	Zymo (Z)	V1-V2	Illumina	minitax	NCBI genome collection	Z1	4,133	0,960
Tot! Illumina V1-V2 minitax NCBI.genome.collection Z2	dog ('Tot!')	Zymo (Z)	V1-V2	Illumina	minitax	NCBI genome collection	Z2	3,886	0,947
Tot! Illumina V1-V2 minitax NCBI.genome.collection Z3	dog ('Tot!')	Zymo (Z)	V1-V2	Illumina	minitax	NCBI genome collection	Z3	3,898	0,950
Tot! Illumina V1-V2 minitax NCBI.genome.collection Z4	dog ('Tot!')	Zymo (Z)	V1-V2	Illumina	minitax	NCBI genome collection	Z4	3,884	0,944
Tot! Illumina V1-V3 minitax NCBI.genome.collection I1	dog ('Tot!')	Invitrogen (I)	V1-V3	Illumina	minitax	NCBI genome collection	I1	3,910	0,930
Tot! Illumina V1-V3 minitax NCBI.genome.collection I2	dog ('Tot!')	Invitrogen (I)	V1-V3	Illumina	minitax	NCBI genome collection	I2	3,620	0,917
Tot! Illumina V1-V3 minitax NCBI.genome.collection I3	dog ('Tot!')	Invitrogen (I)	V1-V3	Illumina	minitax	NCBI genome collection	I3	4,478	0,969
Tot! Illumina V1-V3 minitax NCBI.genome.collection I4	dog ('Tot!')	Invitrogen (I)	V1-V3	Illumina	minitax	NCBI genome collection	I4	3,915	0,945
Tot! Illumina V1-V3 minitax NCBI.genome.collection MN1	dog ('Tot!')	Macherey-Nagel (MN)	V1-V3	Illumina	minitax	NCBI genome collection	MN1	4,113	0,948
Tot! Illumina V1-V3 minitax NCBI.genome.collection MN2	dog ('Tot!')	Macherey-Nagel (MN)	V1-V3	Illumina	minitax	NCBI genome collection	MN2	3,855	0,923
Tot! Illumina V1-V3 minitax NCBI.genome.collection MN3	dog ('Tot!')	Macherey-Nagel (MN)	V1-V3	Illumina	minitax	NCBI genome collection	MN3	3,458	0,925
Tot! Illumina V1-V3 minitax NCBI.genome.collection MN4	dog ('Tot!')	Macherey-Nagel (MN)	V1-V3	Illumina	minitax	NCBI genome collection	MN4	4,214	0,956
Tot! Illumina V1-V3 minitax NCBI.genome.collection Q1	dog ('Tot!')	Qiagen (Q)	V1-V3	Illumina	minitax	NCBI genome collection	Q1	3,924	0,932
Tot! Illumina V1-V3 minitax NCBI.genome.collection Q2	dog ('Tot!')	Qiagen (Q)	V1-V3	Illumina	minitax	NCBI genome collection	Q2	4,032	0,948
Tot! Illumina V1-V3 minitax NCBI.genome.collection Q3	dog ('Tot!')	Qiagen (Q)	V1-V3	Illumina	minitax	NCBI genome collection	Q3	3,712	0,905
Tot! Illumina V1-V3 minitax NCBI.genome.collection Q4	dog ('Tot!')	Qiagen (Q)	V1-V3	Illumina	minitax	NCBI genome collection	Q4	3,034	0,922
Tot! Illumina V1-V3 minitax NCBI.genome.collection Z1	dog ('Tot!')	Zymo (Z)	V1-V3	Illumina	minitax	NCBI genome collection	Z1	4,159	0,950
Tot! Illumina V1-V3 minitax NCBI.genome.collection Z2	dog ('Tot!')	Zymo (Z)	V1-V3	Illumina	minitax	NCBI genome collection	Z2	3,756	0,919
Tot! Illumina V1-V3 minitax NCBI.genome.collection Z3	dog ('Tot!')	Zymo (Z)	V1-V3	Illumina	minitax	NCBI genome collection	Z3	4,327	0,957
Tot! Illumina V1-V3 minitax NCBI.genome.collection Z4	dog ('Tot!')	Zymo (Z)	V1-V3	Illumina	minitax	NCBI genome collection	Z4	3,143	0,844
Tot! Illumina V3-V4 minitax NCBI.genome.collection I1	dog ('Tot!')	Invitrogen (I)	V3-V4	Illumina	minitax	NCBI genome collection	I1	3,505	0,909
Tot! Illumina V3-V4 minitax NCBI.genome.collection I2	dog ('Tot!')	Invitrogen (I)	V3-V4	Illumina	minitax	NCBI genome collection	I2	3,548	0,921
Tot! Illumina V3-V4 minitax NCBI.genome.collection I3	dog ('Tot!')	Invitrogen (I)	V3-V4	Illumina	minitax	NCBI genome collection	I3	3,492	0,916
Tot! Illumina V3-V4 minitax NCBI.genome.collection I4	dog ('Tot!')	Invitrogen (I)	V3-V4	Illumina	minitax	NCBI genome collection	I4	3,504	0,912
Tot! Illumina V3-V4 minitax NCBI.genome.collection MN1	dog ('Tot!')	Macherey-Nagel (MN)	V3-V4	Illumina	minitax	NCBI genome collection	MN1	3,502	0,913
Tot! Illumina V3-V4 minitax NCBI.genome.collection MN2	dog ('Tot!')	Macherey-Nagel (MN)	V3-V4	Illumina	minitax	NCBI genome collection	MN2	3,545	0,919
Tot! Illumina V3-V4 minitax NCBI.genome.collection MN3	dog ('Tot!')	Macherey-Nagel (MN)	V3-V4	Illumina	minitax	NCBI genome collection	MN3	3,570	0,920
Tot! Illumina V3-V4 minitax NCBI.genome.collection MN4	dog ('Tot!')	Macherey-Nagel (MN)	V3-V4	Illumina	minitax	NCBI genome collection	MN4	3,608	0,922
Tot! Illumina V3-V4 minitax NCBI.genome.collection Z1	dog ('Tot!')	Zymo (Z)	V3-V4	Illumina	minitax	NCBI genome collection	Z1	3,576	0,925
Tot! Illumina V3-V4 minitax NCBI.genome.collection Z2	dog ('Tot!')	Zymo (Z)	V3-V4	Illumina	minitax	NCBI genome collection	Z2	3,677	0,934
Tot! Illumina V3-V4 minitax NCBI.genome.collection Z3	dog ('Tot!')	Zymo (Z)	V3-V4	Illumina	minitax	NCBI genome collection	Z3	3,522	0,925
Tot! Illumina V3-V4 minitax NCBI.genome.collection Z4	dog ('Tot!')	Zymo (Z)	V3-V4	Illumina	minitax	NCBI genome collection	Z4	3,857	0,947
Tot! Illumina mWGS minitax NCBI.genome.collection I1	dog ('Tot!')	Invitrogen (I)	mWGS	Illumina	minitax	NCBI genome collection	I1	4,195	0,938
Tot! Illumina mWGS minitax NCBI.genome.collection I2	dog ('Tot!')	Invitrogen (I)	mWGS	Illumina	minitax	NCBI genome collection	I2	4,363	0,951
Tot! Illumina mWGS minitax NCBI.genome.collection I3	dog ('Tot!')	Invitrogen (I)	mWGS	Illumina	minitax	NCBI genome collection	I3	4,260	0,940
Tot! Illumina mWGS minitax NCBI.genome.collection I4	dog ('Tot!')	Invitrogen (I)	mWGS	Illumina	minitax	NCBI genome collection	I4	4,202	0,937
Tot! Illumina mWGS minitax NCBI.genome.collection MN1	dog ('Tot!')	Macherey-Nagel (MN)	mWGS	Illumina	minitax	NCBI genome collection	MN1	4,403	0,955
Tot! Illumina mWGS minitax NCBI.genome.collection MN2	dog ('Tot!')	Macherey-Nagel (MN)	mWGS	Illumina	minitax	NCBI genome collection	MN2	4,414	0,957
Tot! Illumina mWGS minitax NCBI.genome.collection MN3	dog ('Tot!')	Macherey-Nagel (MN)	mWGS	Illumina	minitax	NCBI genome collection	MN3	4,439	0,953
Tot! Illumina mWGS minitax NCBI.genome.collection MN4	dog ('Tot!')	Macherey-Nagel (MN)	mWGS	Illumina	minitax	NCBI genome collection	MN4	4,390	0,953
Tot! Illumina mWGS minitax NCBI.genome.collection Q1	dog ('Tot!')	Qiagen (Q)	mWGS	Illumina	minitax	NCBI genome collection	Q1	3,414	0,882
Tot! Illumina mWGS minitax NCBI.genome.collection Q2	dog ('Tot!')	Qiagen (Q)	mWGS	Illumina	minitax	NCBI genome collection	Q2	3,225	0,866
Tot! Illumina mWGS minitax NCBI.genome.collection Q3	dog ('Tot!')	Qiagen (Q)	mWGS	Illumina	minitax	NCBI genome collection	Q3	3,194	0,871
Tot! Illumina mWGS minitax NCBI.genome.collection Q4	dog ('Tot!')	Qiagen (Q)	mWGS	Illumina	minitax	NCBI genome collection	Q4	3,479	0,884
Tot! Illumina mWGS minitax NCBI.genome.collection Z1	dog ('Tot!')	Zymo (Z)	mWGS	Illumina	minitax	NCBI genome collection	Z1	4,228	0,946
Tot! Illumina mWGS minitax NCBI.genome.collection Z2	dog ('Tot!')	Zymo (Z)	mWGS	Illumina	minitax	NCBI genome collection	Z2	4,273	0,948
Tot! Illumina mWGS minitax NCBI.genome.collection Z3	dog ('Tot!')	Zymo (Z)	mWGS	Illumina	minitax	NCBI genome collection	Z3	4,272	0,951
Tot! Illumina mWGS minitax NCBI.genome.collection Z4	dog ('Tot!')	Zymo (Z)	mWGS	Illumina	minitax	NCBI genome collection	Z4	4,283	0,947
Tot! Illumina mWGS minitax NCBI.genome.collection I1	dog ('Tot!')	Invitrogen (I)	mWGS	Illumina	minitax	NCBI genome collection	I1	4,018	0,928
Tot! Illumina mWGS minitax NCBI.genome.collection I2	dog ('Tot!')	Invitrogen (I)	mWGS	Illumina	minitax	NCBI genome collection	I2	4,160	0,942
Tot! Illumina mWGS minitax NCBI.genome.collection I3	dog ('Tot!')	Invitrogen (I)	mWGS	Illumina	minitax	NCBI genome collection	I3	4,050	0,929
Tot! Illumina mWGS minitax NCBI.genome.collection I4	dog ('Tot!')	Invitrogen (I)	mWGS	Illumina	minitax	NCBI genome collection	I4	4,011	0,926
Tot! Illumina mWGS minitax NCBI.genome.collection MN1	dog ('Tot!')	Macherey-Nagel (MN)	mWGS	Illumina	minitax	NCBI genome collection	MN1	4,218	0,948
Tot! Illumina mWGS minitax NCBI.genome.collection MN2	dog ('Tot!')	Macherey-Nagel (MN)	mWGS	Illumina	minitax	NCBI genome collection	MN2	4,219	0,949
Tot! Illumina mWGS minitax NCBI.genome.collection MN3	dog ('Tot!')	Macherey-Nagel (MN)	mWGS	Illumina	minitax	NCBI genome collection	MN3	4,230	0,945
Tot! Illumina mWGS minitax NCBI.genome.collection MN4	dog ('Tot!')	Macherey-Nagel (MN)	mWGS	Illumina	minitax	NCBI genome collection	MN4	4,196	0,946
Tot! Illumina mWGS minitax NCBI.genome.collection Q1	dog ('Tot!')	Qiagen (Q)	mWGS	Illumina	minitax	NCBI genome collection	Q1	3,169	0,837
Tot! Illumina mWGS minitax NCBI.genome.collection Q2	dog ('Tot!')	Qiagen (Q)	mWGS	Illumina	minitax	NCBI genome collection	Q2	2,990	0,820
Tot! Illumina mWGS minitax NCBI.genome.collection Q3	dog ('Tot!')	Qiagen (Q)	mWGS	Illumina	minitax	NCBI genome collection	Q3	3,007	0,827
Tot! Illumina mWGS minitax NCBI.genome.collection Q4	dog ('Tot!')	Qiagen (Q)	mWGS	Illumina	minitax	NCBI genome collection	Q4	3,245	0,841
Tot! Illumina mWGS minitax NCBI.genome.collection Z1	dog ('Tot!')	Zymo (Z)	mWGS	Illumina	minitax	NCBI genome collection	Z1	3,956	0,925
Tot! Illumina mWGS minitax NCBI.genome.collection Z2	dog ('Tot!')	Zymo (Z)	mWGS	Illumina	minitax	NCBI genome collection	Z2	4,044	0,935
Tot! Illumina mWGS minitax NCBI.genome.collection Z3	dog ('Tot!')	Zymo (Z)	mWGS	Illumina	minitax	NCBI genome collection	Z3	4,045	0,937
Tot! Illumina mWGS minitax NCBI.genome.collection Z4	dog ('Tot!')	Zymo (Z)	mWGS	Illumina	minitax	NCBI genome collection	Z4	4,104	0,938
Tot! ONT V1-V9 minitax NCBI.genome.collection I1	dog ('Tot!')	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I1	2,765	0,881
Tot! ONT V1-V9 minitax NCBI.genome.collection I2	dog ('Tot!')	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I2	2,838	0,892
Tot! ONT V1-V9 minitax NCBI.genome.collection I3	dog ('Tot!')	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I3	2,910	0,897
Tot! ONT V1-V9 minitax NCBI.genome.collection I4	dog ('Tot!')	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I4	2,865	0,889
Tot! ONT V1-V9 minitax NCBI.genome.collection MN1	dog ('Tot!')	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN1	2,935	0,907
Tot! ONT V1-V9 minitax NCBI.genome.collection MN2	dog ('Tot!')	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN2	3,045	0,908
Tot! ONT V1-V9 minitax NCBI.genome.collection MN3	dog ('Tot!')	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN3	2,987	0,903
Tot! ONT V1-V9 minitax NCBI.genome.collection MN4	dog ('Tot!')	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN4	3,046	0,912
Tot! ONT V1-V9 minitax NCBI.genome.collection Q1	dog ('Tot!')	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q1	1,386	0,740
Tot! ONT V1-V9 minitax NCBI.genome.collection Q2	dog ('Tot!')	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q2	2,210	0,875
Tot! ONT V1-V9 minitax NCBI.genome.collection Q3	dog ('Tot!')	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q3	1,735	0,815
Tot! ONT V1-V9 minitax NCBI.genome.collection Q4	dog ('Tot!')	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q4	2,282	0,886
Tot! ONT V1-V9 minitax NCBI.genome.collection Z1	dog ('Tot!')	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z1	3,204	0,925
Tot! ONT V1-V9 minitax NCBI.genome.collection Z2	dog ('Tot!')	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z2	2,999	0,914
Tot! ONT V1-V9 minitax NCBI.genome.collection Z3	dog ('Tot!')	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z3	3,119	0,926
Tot! ONT V1-V9 minitax NCBI.genome.collection Z4	dog ('Tot!')	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z4	3,016	0,914
Tot! PacBio V1-V9 minitax NCBI.genome.collection I1	dog ('Tot!')	Invitrogen (I)	V1-V9	PacBio	minitax	NCBI genome collection	I1	2,409	0,799
Tot! PacBio V1-V9 minitax NCBI.genome.collection I2	dog ('Tot!')	Invitrogen (I)	V1-V9	PacBio	minitax	NCBI genome collection	I2	2,581	0,835
Tot! PacBio V1-V9 minitax NCBI.genome.collection I3	dog ('Tot!')	Invitrogen (I)	V1-V9	PacBio	minitax	NCBI genome collection	I3	2,492	0,830
Tot! PacBio V1-V9 minitax NCBI.genome.collection I4	dog ('Tot!')	Invitrogen (I)	V1-V9	PacBio	minitax	NCBI genome collection	I4	2,477	0,809
Tot! PacBio V1-V9 minitax NCBI.genome.collection MN1	dog ('Tot!')	Macherey-Nagel (MN)	V1-V9	PacBio	minitax	NCBI genome collection	MN1	2,697	0,885
Tot! PacBio V1-V9 minitax NCBI.genome.collection MN2	dog ('Tot!')	Macherey-Nagel (MN)	V1-V9	PacBio	minitax	NCBI genome collection	MN2	2,739	0,864
Tot! PacBio V1-V9 minitax NCBI.genome.collection MN3	dog ('Tot!')	Macherey-Nagel (MN)	V1-V9	PacBio	minitax	NCBI genome collection	MN3	2,673	0,855
Tot! PacBio V1-V9 minitax NCBI.genome.collection MN4	dog ('Tot!')	Macherey-Nagel (MN)	V1-V9	PacBio	minitax	NCBI genome collection	MN4	2,727	0,868
Tot! PacBio V1-V9 minitax NCBI.genome.collection Z1	dog ('Tot!')	Zymo (Z)	V1-V9	PacBio	minitax	NCBI genome collection	Z1	2,846	0,895
Tot! PacBio V1-V9 minitax NCBI.genome.collection Z2	dog ('Tot!')	Zymo (Z)	V1-V9	PacBio	minitax	NCBI genome collection	Z2	3,032	0,928
Tot! PacBio V1-V9 minitax NCBI.genome.collection Z3	dog ('Tot!')	Zymo (Z)	V1-V9	PacBio	minitax	NCBI genome collection	Z3	2,851	0,916
Tot! PacBio V1-V9 minitax NCBI.genome.collection Z4	dog ('Tot!')	Zymo (Z)	V1-V9	PacBio	minitax	NCBI genome collection	Z4	3,001	0,925

***In vitro* samples**

sample	source	DNA extraction method	Target (V-region)	platform	program	database	sample_name	Shannon	Simpson
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection I1	Zymo D6300	Invitrogen (I)	V1-V2	Illumina	minitax	NCBI genome collection	I1	3,772	0,952
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection I2	Zymo D6300	Invitrogen (I)	V1-V2	Illumina	minitax	NCBI genome collection	I2	3,706	0,947
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection I3	Zymo D6300	Invitrogen (I)	V1-V2	Illumina	minitax	NCBI genome collection	I3	3,611	0,943
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection MN1	Zymo D6300	Macherey-Nagel (MN)	V1-V2	Illumina	minitax	NCBI genome collection	MN1	4,012	0,953
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection MN2	Zymo D6300	Macherey-Nagel (MN)	V1-V2	Illumina	minitax	NCBI genome collection	MN2	3,990	0,955
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection MN3	Zymo D6300	Macherey-Nagel (MN)	V1-V2	Illumina	minitax	NCBI genome collection	MN3	3,959	0,955
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection Q1	Zymo D6300	Qiagen (Q)	V1-V2	Illumina	minitax	NCBI genome collection	Q1	3,781	0,937
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection Q2	Zymo D6300	Qiagen (Q)	V1-V2	Illumina	minitax	NCBI genome collection	Q2	3,934	0,954
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection Q3	Zymo D6300	Qiagen (Q)	V1-V2	Illumina	minitax	NCBI genome collection	Q3	3,742	0,932
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection I1	Zymo D6300	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I1	2,902	0,920
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection I2	Zymo D6300	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I2	2,894	0,918
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection I3	Zymo D6300	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I3	2,889	0,918
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection I4	Zymo D6300	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I4	2,965	0,920
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection I5	Zymo D6300	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I5	2,876	0,918
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection MN1	Zymo D6300	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN1	2,902	0,897
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection MN2	Zymo D6300	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN2	2,881	0,899
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection MN3	Zymo D6300	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN3	2,776	0,885
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection MN4	Zymo D6300	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN4	2,716	0,875
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Q1	Zymo D6300	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q1	2,583	0,862
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Q2	Zymo D6300	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q2	2,518	0,853
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Q3	Zymo D6300	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q3	2,535	0,857
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Q4	Zymo D6300	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q4	2,642	0,887
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Q5	Zymo D6300	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q5	2,600	0,864
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Z1	Zymo D6300	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z1	2,998	0,910
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Z2	Zymo D6300	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z2	2,838	0,901
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Z3	Zymo D6300	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z3	2,714	0,877
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Z4	Zymo D6300	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z4	2,726	0,882
ZymoD6300 ONT V1-V9 minitax NCBI.genome.collection Z5	Zymo D6300	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z5	2,725	0,884
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection Z1	Zymo D6300	Zymo (Z)	V1-V2	Illumina	minitax	NCBI genome collection	Z1	3,902	0,956
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection Z2	Zymo D6300	Zymo (Z)	V1-V2	Illumina	minitax	NCBI genome collection	Z2	3,864	0,954
ZymoD6300 Illumina V1-V2 minitax NCBI.genome.collection Z3	Zymo D6300	Zymo (Z)	V1-V2	Illumina	minitax	NCBI genome collection	Z3	3,997	0,957
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection I1	Zymo D6331	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I1	1,999	0,815
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection I2	Zymo D6331	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I2	1,999	0,833
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection I3	Zymo D6331	Invitrogen (I)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	I3	1,842	0,781
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection MN1	Zymo D6331	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN1	2,275	0,837
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection MN2	Zymo D6331	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN2	2,288	0,839
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection MN3	Zymo D6331	Macherey-Nagel (MN)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	MN3	2,407	0,859
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection Q1	Zymo D6331	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q1	2,282	0,831
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection Q2	Zymo D6331	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q2	2,235	0,839
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection Q3	Zymo D6331	Qiagen (Q)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Q3	2,199	0,819
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection Z1	Zymo D6331	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z1	2,236	0,821
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection Z2	Zymo D6331	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z2	2,241	0,835
ZymoD6331 ONT V1-V9 minitax NCBI.genome.collection Z3	Zymo D6331	Zymo (Z)	V1-V9	MinION (ONT)	minitax	NCBI genome collection	Z3	2,230	0,817

10.5. Supplementary Data 5

Supplementary Data 5a. The statistical data of all sequencing performed during the project. a. Data obtained from a single sample of a 13-year-old dog.

Source	Breed (name of the dog)	Age at sampling	Sex	name	sample	kit	barcode	date	seq_type	cycle	read_count	length min	length_median	length_mean	length_max	quality_min	quality_median	quality_mean	quality_max	avg_base_quality_min	avg_base_quality_median	avg_base_quality_mean	avg_base_quality_max
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_invitrogen_I1_R1	I1	Invitrogen	NA	2E+07	Illumina	500	41317	35	251	250,575308	251	2	32,02	31,4189837	37,58	0,06	0,13	0,125325895	0,19
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_invitrogen_I1_R2	I1	Invitrogen	NA	2E+07	Illumina	500	41317	35	251	250,599051	251	2	30,73	30,3160573	36,41	0,06	0,12	0,120963526	0,19
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_invitrogen_I2_R1	I2	Invitrogen	NA	2E+07	Illumina	500	27839	208	251	250,611624	251	20,47	33,83	33,0019347	37,68	0,08	0,14	0,131639786	0,17
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_invitrogen_I2_R2	I2	Invitrogen	NA	2E+07	Illumina	500	27839	208	251	250,653149	251	16,49	31,78	31,1826035	36,29	0,07	0,13	0,124394554	0,17
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_invitrogen_I3_R1	I3	Invitrogen	NA	2E+07	Illumina	500	36413	35	251	250,547305	251	2	32,44	31,7104202	37,56	0,06	0,13	0,126534479	0,17
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_invitrogen_I3_R2	I3	Invitrogen	NA	2E+07	Illumina	500	36413	35	251	250,654272	251	2	31,08	30,5634057	36,17	0,06	0,12	0,12191882	0,17
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_invitrogen_I4_R1	I4	Invitrogen	NA	2E+07	Illumina	500	36849	35	251	250,585091	251	2	33,53	32,6693875	37,66	0,06	0,13	0,130310185	0,22
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_invitrogen_I4_R2	I4	Invitrogen	NA	2E+07	Illumina	500	36849	35	251	250,627778	251	2	31,43	30,8354916	36,16	0,06	0,13	0,122993026	0,21
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	45884	35	251	250,602302	251	2	31,83	31,2777853	37,57	0,06	0,13	0,124758086	0,18
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	45884	35	251	250,651534	251	2	30,71	30,283365	36,54	0,06	0,12	0,120815971	0,17
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_macherey_nagel_MN2_R1	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	33105	191	251	250,607914	251	21,1	33,7	32,8188026	37,66	0,08	0,13	0,130921311	0,19
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_macherey_nagel_MN2_R2	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	33105	191	251	250,674098	251	16,25	31,69	31,0811944	36,49	0,06	0,13	0,124005135	0,19
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	53549	192	251	250,594652	251	19,54	32,54	31,7971934	37,59	0,08	0,13	0,126852042	0,19
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	53549	192	251	250,656894	251	15,96	31,07	30,5571215	36,22	0,06	0,12	0,121899382	0,19
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_macherey_nagel_MN4_R1	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	34775	35	251	250,616305	251	2	33,61	32,7972	37,64	0,06	0,13	0,130792811	0,25
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_macherey_nagel_MN4_R2	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	34775	35	251	250,666542	251	2	31,55	30,9765577	36,15	0,06	0,13	0,123575557	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_zymo_Z1_R1	Z1	Zymo_Research	NA	2E+07	Illumina	500	33134	35	251	250,647643	251	2	31,93	31,3771778	37,62	0,06	0,13	0,125148488	0,26
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_zymo_Z1_R2	Z1	Zymo_Research	NA	2E+07	Illumina	500	33134	35	251	250,619243	251	2	30,74	30,3028493	36,14	0,06	0,12	0,120937707	0,27
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_zymo_Z2_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	30560	35	251	250,660013	251	2	33,68	32,8363871	37,74	0,06	0,13	0,131001309	0,68
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_zymo_Z2_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	30560	35	251	250,684522	251	2	31,53	30,9265795	37,14	0,06	0,13	0,123382199	0,73
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_zymo_Z3_R1	Z3	Zymo_Research	NA	2E+07	Illumina	500	47335	35	251	250,656259	251	2	32,88	32,085891	37,96	0,06	0,13	0,127972959	0,68
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_zymo_Z3_R2	Z3	Zymo_Research	NA	2E+07	Illumina	500	47335	35	251	250,599641	251	2	31,19	30,6839518	36,76	0,06	0,12	0,122431182	0,7
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_zymo_Z4_R1	Z4	Zymo_Research	NA	2E+07	Illumina	500	55117	35	251	250,665584	251	2	33,78	32,9386955	37,72	0,06	0,13	0,131415171	0,7
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_2_20220525_zymo_Z4_R2	Z4	Zymo_Research	NA	2E+07	Illumina	500	55117	35	251	250,645409	251	2	31,58	31,0221404	37,12	0,06	0,13	0,12377125	0,73
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_invitrogen_I1_R1	I1	Invitrogen	NA	2E+07	Illumina	500	269541	238	251	250,717728	251	17,04	32,96	31,4086989	38,17	0,07	0,13	0,125454977	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_invitrogen_I1_R2	I1	Invitrogen	NA	2E+07	Illumina	500	269541	130	251	250,496589	251	14,14	29,23	28,335107	37,78	0,06	0,12	0,113171725	0,29
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_invitrogen_I2_R1	I2	Invitrogen	NA	2E+07	Illumina	500	19842	35	251	250,789437	251	2	33,265	31,6569398	38,12	0,06	0,13	0,126409132	0,18
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_invitrogen_I2_R2	I2	Invitrogen	NA	2E+07	Illumina	500	19842	35	251	250,468804	251	2	29,81	28,7658296	37,25	0,06	0,12	0,114893156	0,3
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	268776	145	251	250,70571	251	17,77	32,87	31,3906997	38,21	0,07	0,13	0,125383777	0,26
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	268776	35	250	250,444712	251	2	32,42	30,9435366	37,88	0,06	0,13	0,123522115	0,28
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_macherey_nagel_MN2_R1	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	255027	84	251	250,708196	251	16,26	33,02	31,4916547	38,18	0,06	0,13	0,125781702	0,43
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_macherey_nagel_MN2_R2	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	255027	68	251	250,489811	251	14,69	32,53	30,9647226	37,84	0,06	0,13	0,123594678	0,49
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_qiagen_Q1_R1	Q1	Qiagen	NA	2E+07	Illumina	500	223006	35	251	250,57542	251	2	34,04	32,1884823	38,24	0,06	0,14	0,128600755	0,83
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_qiagen_Q1_R2	Q1	Qiagen	NA	2E+07	Illumina	500	223006	35	251	250,870255	251	2	31,84	30,4923318	38,01	0,06	0,13	0,121512515	0,81
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_qiagen_Q2_R1	Q2	Qiagen	NA	2E+07	Illumina	500	195316	35	250	250,236627	251	2	34,29	32,4461968	38,19	0,06	0,14	0,129790801	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211204_qiagen_Q2_R2	Q2	Qiagen	NA	2E+07	Illumina	500	195316	35	251	250,869468	251	2	31,94	30,5987117	37,81	0,06	0,13	0,121945514	0,51
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_invitrogen_I3_R1	I3	Invitrogen	NA	2E+07	Illumina	500	57681	246	251	250,685564	251	16,76	27,15	27,0066611	32,24	0,07	0,11	0,1077197	0,13
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_invitrogen_I3_R2	I3	Invitrogen	NA	2E+07	Illumina	500	57681	182	250	250,446196	251	15,71	25,11	25,0670379	29,69	0,06	0,1	0,100124998	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_invitrogen_I4_R1	I4	Invitrogen	NA	2E+07	Illumina	500	921	35	251	250,094463	251	2	27,16	26,953051	32,61	0,06	0,11	0,107980456	0,53
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_invitrogen_I4_R2	I4	Invitrogen	NA	2E+07	Illumina	500	921	35	250	249,39848	251	2	24,87	24,7962866	30,69	0,06	0,1	0,099685125	0,5
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	309	35	251	248,122977	251	2	26,85	26,6554045	33,1	0,06	0,11	0,108770227	0,47
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	309	35	250	248,145631	251	2	25,12	24,8805825	32,49	0,06	0,1	0,101100324	0,46
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_macherey_nagel_MN4_R1	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	45965	214	251	250,689503	251	17,44	27,13	27,0061993	32,22	0,07	0,11	0,107716959	0,14
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_macherey_nagel_MN4_R2	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	45965	134	250	250,43866	251	15,12	25,12	25,0712129	31,55	0,06	0,1	0,100150767	0,23
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_qiagen_Q1_R1	Q1	Qiagen	NA	2E+07	Illumina	500	45486	35	251	250,446709	251	2	28,21	28,0444227	32,41	0,06	0,11	0,111969177	0,14
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_qiagen_Q1_R2	Q1	Qiagen	NA	2E+07	Illumina	500	45486	35	251	250,710526	251	2	26,18	26,0152836	30,08	0,06	0,1	0,103751924	0,13
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_qiagen_Q4_R1	Q4	Qiagen	NA	2E+07	Illumina	500	78	249	250	250,102564	251	22,77	27,77	27,3303846	31,25	0,09	0,11	0,108846154	0,12
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_qiagen_Q4_R2	Q4	Qiagen	NA	2E+07	Illumina	500	78	250	251	250,807692	251	18,53	25,49	25,468718	28,9	0,07	0,1	0,101794872	0,12
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_zymo_Z1_R1	Z1	Zymo_Research	NA	2E+07	Illumina	500	47946	239	251	250,598382	251	17,12	27,45	27,3075988	32,3	0,07	0,11	0,108966337	0,13
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_zymo_Z1_R2	Z1	Zymo_Research	NA	2E+07	Illumina	500	47946	239	251	250,647166	251	15,64	25,55	25,4374507	30,92	0,06	0,1	0,101503567	0,12
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_zymo_Z2_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	37384	246	251	250,778622	251	17,52	27,36	27,2305251	32,36	0,07	0,11	0,108573721	0,13
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20211216_zymo_Z2_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	37384	179	250	250,493848	251	16,43	25,45	25,3637011	29,85	0,07	0,1	0,101339343	0,16
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_invitrogen_I3_R1	I3	Invitrogen	NA	2E+07	Illumina	300	68160	138	151	150,82299	151	19,13	36,13	34,6935098	38,37	0,13	0,24	0,2300826	0,27
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_invitrogen_I3_R2	I3	Invitrogen	NA	2E+07	Illumina	300	68160	35	151	150,804401	151	2	36,3	35,027151	38,28	0,06	0,24	0,232336414	0,56
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_invitrogen_I4_R1	I4	Invitrogen	NA	2E+07	Illumina	300	1355	35	151	150,687085	151	2	36,25	34,8007675	38,27	0,06	0,24	0,230752768	0,26
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_invitrogen_I4_R2	I4	Invitrogen	NA	2E+07	Illumina	300	1355	35	151	150,757934	151	2	36,38	35,2181255	38,22	0,06	0,24	0,233557196	0,25
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	300	49333	147	151	150,840229	151	20,25	36,07	34,6258348	38,41	0,13	0,24	0,229569457	0,26
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	300	49333	60	151	150,794195	151	13,38	36,23	34,9726031	38,26	0,09	0,24	0,231982446	0,53
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_macherey_nagel_MN4_R1	MN4	Macherey-Nagel	NA	2E+07	Illumina	300	615105	71	151	150,827161	151	16,89	35,83	34,2828279	38,44	0,11	0,24	0,227344193	0,53
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_macherey_nagel_MN4_R2	MN4	Macherey-Nagel	NA	2E+07	Illumina	300	615105	71	151	150,82012	151	12,36	36,12	34,6655578	38,3	0,08	0,24	0,22991951	0,52
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_qiagen_Q2_R1	Q2	Qiagen	NA	2E+07	Illumina	300	62752	139	150	150,429994	151	19,33	36,03	34,7258567	38,45	0,13	0,24	0,230708344	0,26
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_qiagen_Q2_R2	Q2	Qiagen	NA	2E+07	Illumina	300	62752	87	150	150,372371	151	12,71	36,14	34,5663764	38,35	0,08	0,24	0,229831081	0,43
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_qiagen_Q2_R2	Q2	Qiagen	NA	2E+07	Illumina	300	62752	87	150	150,372371	151	12,71	36,14	34,5663764	38,35	0,08	0,24	0,229831081	0,43
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_qiagen_Q2_R2	Q2	Qiagen	NA	2E+07	Illumina	300	62752	87	150	150,372371	151	12,71	36,14	34,5663764	38,35	0,08	0,24	0,229831081	0,43
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_zymo_Z3_R1	Z3	Zymo_Research	NA	2E+07	Illumina	300	56050	143	151	150,768153	151	16,78	35,94	34,5201729	38,45	0,11	0,24	0,22896289	0,27
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_zymo_Z3_R2	Z3	Zymo_Research	NA	2E+07	Illumina	300	56050	129	151	150,761124	151	13,18	36,23	34,7165991	38,28	0,09	0,24	0,230323818	0,29
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_zymo_Z4_R1	Z4	Zymo_Research	NA	2E+07	Illumina	300	5345	35	151	150,815529	151	2	35,99	34,5905482	38,34	0,06	0,24	0,229391955	0,26
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220117_zymo_Z4_R2	Z4	Zymo_Research	NA	2E+07	Illumina	300	5345	35	151	150,849392	151	2	36,34	34,9613583	38,25	0,06	0,24	0,23175304	0,25
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_invitrogen_I3_R1	I3	Invitrogen	NA	2E+07	Illumina	300	128091	144	151	150,734579	151	15,63	29,23	28,8246167	35,58	0,1	0,19	0,19121562	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_invitrogen_I3_R2	I3	Invitrogen	NA	2E+07	Illumina	300	128091	35	151	150,927161	151	16,21	27,13	26,7955977	33,66	0,11	0,18	0,177519732	0,96
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	300	116707	145	151	150,74526	151	16,23	29,28	28,8700942	35,6	0,11	0,19	0,191517218	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	300	116707	35	151	150,923561	151	2	27,21	26,8553729	33,67	0,06	0,18	0,177924632	0,43
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_qiagen_Q1_R1	Q1	Qiagen	NA	2E+07	Illumina	300	63886	35	151	150,720299	151	2	29,81	29,3339038	35,98	0,06	0,2	0,194571737	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_qiagen_Q1_R2	Q1	Qiagen	NA	2E+07	Illumina	300	63886	35	151	150,846414	151	2	27,72	27,3501686	32,74	0,06	0,18	0,181264909	0,22
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_qiagen_Q2_R1	Q2	Qiagen	NA	2E+07	Illumina	300	104737	145	151	150,746193	151	17,09	29,89	29,420608	35,87	0,11	0,2	0,195161977	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_qiagen_Q2_R2	Q2	Qiagen	NA	2E+07	Illumina	300	104737	38	151	150,885742	151	16,1	27,91	27,5379518	33,04	0,11	0,18	0,182512866	0,87
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_qiagen_Q3_R1	Q3	Qiagen	NA	2E+07	Illumina	300	109123	35	151	150,845642	151	2	29,78	29,3123792	35,76	0,06	0,2	0,194296253	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_qiagen_Q3_R2	Q3	Qiagen	NA	2E+07	Illumina	300	109123	35	151	150,884222	151	2	27,64	27,2759473	33,39	0,06	0,18	0,180743473	0,22
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_zymo_Z3_R1	Z3	Zymo_Research	NA	2E+07	Illumina	300	100907	145	151	150,843004	151	17	29,55	29,1231935	35,5	0,11	0,2	0,193060442	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20220126_zymo_Z3_R2	Z3	Zymo_Research	NA	2E+07	Illumina	300	100907	91	151	150,939201	151	15,1	26,48	26,2083159	33,08	0,1	0,18	0,173606093	0,36
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_invitrogen_I4_R1	I4	Invitrogen	NA	2E+07	Illumina	500	183751	35	251	250,678908	251	2	34,69	32,9328836	38,47	0,06	0,14	0,131650767	0,94
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_invitrogen_I4_R2	I4	Invitrogen	NA	2E+07	Illumina	500	183751	35	251	250,596884	251	2	32,04	30,7288975	38,37	0,06	0,13	0,122720747	0,94
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	244414	35	251	250,683357	251	2	34,61	32,8996978	38,44	0,06	0,14	0,131510143	0,88
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	244414	35	251	250,59576	251	2	34,51	32,8923229	38,28	0,06	0,14	0,131512393	0,87
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_qiagen_Q3_R1	Q3	Qiagen	NA	2E+07	Illumina	500	188483	35	250	250,236276	251	2	35,68	33,8405625	38,4	0,06	0,14	0,135242754	0,16
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_qiagen_Q3_R2	Q3	Qiagen	NA	2E+07	Illumina	500	188483	35	251	250,844633	251	2	34,16	32,533652	38,34	0,06	0,14	0,129887735	0,55
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_qiagen_Q4_R1	Q4	Qiagen	NA	2E+07	Illumina	500	1202	35	250	249,550749	251	2	35,31	33,589975	38,11	0,06	0,14	0,134326123	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_qiagen_Q4_R2	Q4	Qiagen	NA	2E+07	Illumina	500	1202	35	251	250,063228	251	2	33,91	32,1837188	38,09	0,06	0,14	0,12874376	0,23
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_zymo_Z3_R1	Z3	Zymo_Research	NA	2E+07	Illumina	500	208620	160	251	250,61138	251	17,8	34,79	33,0940951	38,31	0,07	0,14	0,132212444	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_zymo_Z3_R2	Z3	Zymo_Research	NA	2E+07	Illumina	500	208620	35	251	250,707444	251	2	31,78	30,5400345	38,26	0,05	0,13	0,121871537	0,77
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_zymo_Z4_R1	Z4	Zymo_Research	NA	2E+07	Illumina	500	16029	35	251	250,669848	251	2	34,73	33,0177859	38,26	0,06	0,14	0,131949991	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v1_3_20221004_zymo_Z4_R2	Z4	Zymo_Research	NA	2E+07	Illumina	500	16029	35	251	250,384927	251	2	32,5	31,017439	38,13	0,06	0,13	0,123915403	0,38
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_invitrogen_I1_R1	I1	Invitrogen	NA	2E+07	Illumina	500	16540	39	251	250,596977	251	20,77	37,51	37,040269	38,28	0,08	0,15	0,148101572	0,96
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_invitrogen_I1_R2	I1	Invitrogen	NA	2E+07	Illumina	500	16540	53	251	250,914027	251	15,74	37,71	36,7983869	38,43	0,06	0,15	0,146293833	0,67
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_invitrogen_I2_R1	I2	Invitrogen	NA	2E+07	Illumina	500	11640	35	251	250,616151	251	2	37,55	37,0771899	38,24	0,06	0,15	0,148189003	0,7
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_invitrogen_I2_R2	I2	Invitrogen	NA	2E+07	Illumina	500	11640	35	251	250,876461	251	2	37,48	36,3248608	38,42	0,06	0,15	0,144768041	0,69
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_invitrogen_I3_R1	I3	Invitrogen	NA	2E+07	Illumina	500	9453	73	251	250,679996	251	19,4	37,48	37,0007225	38,44	0,08	0,15	0,147960436	0,46
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_invitrogen_I3_R2	I3	Invitrogen	NA	2E+07	Illumina	500	9453	73	251	250,91992	251	16,64	37,46	36,340639	38,41	0,07	0,15	0,144819634	0,5
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_invitrogen_I4_R1	I4	Invitrogen	NA	2E+07	Illumina	500	11754	102	251	250,641484	251	19,53	37,53	37,0779428	38,26	0,08	0,15	0,148192956	0,37
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_invitrogen_I4_R2	I4	Invitrogen	NA	2E+07	Illumina	500	11754	102	251	250,92326	251	16,13	37,7	36,7342505	38,42	0,06	0,15	0,146050706	0,37
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	19967	167	251	250,700319	251	20,04	37,47	37,0021375	38,2	0,08	0,15	0,147890649	0,23
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	19967	167	251	250,982025	251	16,04	37,79	36,8357011	38,51	0,06	0,15	0,146190894	0,23
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_macherey_nagel_MN2_R1	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	18104	42	251	250,725751	251	20,82	37,56	37,0653806	38,26	0,08	0,15	0,14798829	0,89
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_macherey_nagel_MN2_R2	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	18104	42	251	250,973045	251	16,37	37,81	36,919736	38,46	0,07	0,15	0,146513478	0,89
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	13703	247	251	250,729475	251	20,81	37,5	37,0602474	38,24	0,08	0,15	0,148074874	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	13703	246	251	250,982559	251	15,8	37,7	36,7650193	38,45	0,06	0,15	0,146132964	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_macherey_nagel_MN4_R1	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	10378	168	251	250,720659	251	20,31	37,54	37,0602872	38,48	0,08	0,15	0,148011177	0,23
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_macherey_nagel_MN4_R2	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	10378	168	251	250,9762	251	16,25	37,5	36,3666892	38,43	0,06	0,15	0,144773559	0,23
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_zymo_Z1_R1	Z1	Zymo_Research	NA	2E+07	Illumina	500	12654	53	251	250,779121	251	20,22	37,53	36,9943362	38,22	0,08	0,15	0,147810969	0,71
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_zymo_Z1_R2	Z1	Zymo_Research	NA	2E+07	Illumina	500	12654	53	251	250,88715	251	16,08	37,58	36,5096618	38,42	0,06	0,15	0,145383278	0,71
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_zymo_Z2_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	14569	52	251	250,814126	251	16,42	37,54	37,0557528	38,27	0,07	0,15	0,14804791	0,72
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_zymo_Z2_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	14569	52	251	250,886815	251	15,79	37,71	36,6701414	38,51	0,06	0,15	0,145821951	0,72
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_zymo_Z3_R1	Z3	Zymo_Research	NA	2E+07	Illumina	500	12589	39	251	250,840178	251	19,82	37,5	37,022078	38,22	0,08	0,15	0,147949797	0,95
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_zymo_Z3_R2	Z3	Zymo_Research	NA	2E+07	Illumina	500	12589	39	251	250,938438	251	15,75	37,85	37,0130908	38,47	0,06	0,15	0,146861546	0,95
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_zymo_Z4_R1	Z4	Zymo_Research	NA	2E+07	Illumina	500	16002	53	251	250,743095	251	20,25	37,56	37,1209624	38,25	0,08	0,15	0,148343332	0,71
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220531_zymo_Z4_R2	Z4	Zymo_Research	NA	2E+07	Illumina	500	16002	54	251	250,890951	251	15,72	37,85	37,0281734	38,47	0,06	0,15	0,146964754	0,7
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_invitrogen_I1_R1	I1	Invitrogen	NA	2E+07	Illumina	500	21337	44	251	250,870788	251	16,61	29,68	29,5382523	36,85	0,07	0,12	0,117826311	0,82
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_invitrogen_I1_R2	I1	Invitrogen	NA	2E+07	Illumina	500	21337	44	251	250,935699	251	15,26	30,1	29,9965684	37,88	0,06	0,12	0,119642874	0,81
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_invitrogen_I2_R1	I2	Invitrogen	NA	2E+07	Illumina	500	14412	73	251	250,893492	251	16,69	29,82	29,7062164	37,04	0,07	0,12	0,118474188	0,51
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_invitrogen_I2_R2	I2	Invitrogen	NA	2E+07	Illumina	500	14412	73	251	250,930058	251	14,52	29,69	29,5515265	37,73	0,06	0,12	0,117819872	0,52
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_invitrogen_I3_R1	I3	Invitrogen	NA	2E+07	Illumina	500	12860	149	251	250,934215	251	16,17	29,43	29,2950381	37,1	0,06	0,12	0,116782271	0,25
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_invitrogen_I3_R2	I3	Invitrogen	NA	2E+07	Illumina	500	12860	149	251	250,951633	251	15,3	29,56	29,4267761	36,07	0,06	0,12	0,117252722	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_invitrogen_I4_R1	I4	Invitrogen	NA	2E+07	Illumina	500	14439	37	251	250,843618	251	17,03	29,73	29,57781	36,78	0,07	0,12	0,118062193	0,98
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_invitrogen_I4_R2	I4	Invitrogen	NA	2E+07	Illumina	500	14439	37	251	250,894106	251	15,41	30,1	29,9852033	37,88	0,06	0,12	0,119699425	0,98
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	19716	53	251	250,899118	251	17,61	29,56	29,4638461	37,84	0,07	0,12	0,117501014	0,69
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	19716	53	251	250,955975	251	15,84	30,26	30,1783288	37,42	0,06	0,12	0,12028099	0,71
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_macherey_nagel_MN2_R1	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	22856	246	251	250,938659	251	16,67	29,82	29,7026991	35,42	0,07	0,12	0,118366293	0,14
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_macherey_nagel_MN2_R2	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	22856	246	251	250,976068	251	15,05	30,24	30,135676	35,95	0,06	0,12	0,120070004	0,14
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	20308	43	251	250,919047	251	17,37	29,58	29,4413492	35,78	0,07	0,12	0,117402501	0,8
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	20308	43	251	250,956323	251	16,26	30,01	29,8911907	36,14	0,07	0,12	0,11917077	0,8
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_macherey_nagel_MN4_R1	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	14494	38	251	250,910308	251	16,57	29,8	29,7000407	36,18	0,07	0,12	0,118447634	0,95
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_macherey_nagel_MN4_R2	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	14494	38	251	250,952877	251	15,43	29,68	29,5551766	35,83	0,06	0,12	0,117837036	0,91
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_zymo_Z1_R1	Z1	Zymo_Research	NA	2E+07	Illumina	500	15061	35	251	250,871323	251	2	29,58	29,4374384	37,81	0,06	0,12	0,117443729	0,71
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_zymo_Z1_R2	Z1	Zymo_Research	NA	2E+07	Illumina	500	15061	35	251	250,865812	251	2	29,87	29,7178939	37,5	0,06	0,12	0,118639533	0,71
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_zymo_Z2_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	16269	40	251	250,876329	251	17,63	29,75	29,6337347	37,35	0,07	0,12	0,118230991	0,93
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_zymo_Z2_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	16269	40	251	250,900486	251	15,46	30,18	30,0121126	37,63	0,06	0,12	0,119733235	0,91
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_zymo_Z3_R1	Z3	Zymo_Research	NA	2E+07	Illumina	500	17780	35	251	250,861867	251	2	29,61	29,4938538	37,66	0,06	0,12	0,117685039	0,66
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_zymo_Z3_R2	Z3	Zymo_Research	NA	2E+07	Illumina	500	17780	35	251	250,860236	251	2	30,34	30,22759	38,07	0,06	0,12	0,120552868	0,68
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_zymo_Z4_R1	Z4	Zymo_Research	NA	2E+07	Illumina	500	22861	42	251	250,896068	251	17,91	29,83	29,7150238	37,4	0,07	0,12	0,118538122	0,76
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220621_zymo_Z4_R2	Z4	Zymo_Research	NA	2E+07	Illumina	500	22861	53	251	250,92765	251	15,84	30,47	30,3611609	37,58	0,06	0,12	0,121071257	0,71
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_invitrogen_I1_R1	I1	Invitrogen	NA	2E+07	Illumina	500	15274	40	251	250,639518	251	20,16	37,62	37,1803398	38,23	0,08	0,15	0,148401859	0,99
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_invitrogen_I1_R2	I1	Invitrogen	NA	2E+07	Illumina	500	15274	40	251	250,943171	251	15,28	37,94	37,1704419	38,45	0,06	0,15	0,147243682	0,92
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_invitrogen_I2_R1	I2	Invitrogen	NA	2E+07	Illumina	500	11174	53	251	250,608019	251	21,1	37,66	37,2432415	38,3	0,08	0,15	0,148611956	0,71
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_invitrogen_I2_R2	I2	Invitrogen	NA	2E+07	Illumina	500	11174	53	251	250,862717	251	15,65	37,92	37,0409406	38,46	0,06	0,15	0,146810453	0,71
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_invitrogen_I3_R1	I3	Invitrogen	NA	2E+07	Illumina	500	8326	148	251	250,671511	251	19,94	37,61	37,15138	38,25	0,08	0,15	0,148275282	0,26
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_invitrogen_I3_R2	I3	Invitrogen	NA	2E+07	Illumina	500	8326	148	251	250,923373	251	15,26	37,94	37,0357795	38,45	0,06	0,15	0,146664665	0,26
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_invitrogen_I4_R1	I4	Invitrogen	NA	2E+07	Illumina	500	10532	73	251	250,659134	251	20,77	37,65	37,2389081	38,49	0,08	0,15	0,148558678	0,52
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_invitrogen_I4_R2	I4	Invitrogen	NA	2E+07	Illumina	500	10532	73	251	250,95414	251	14,39	37,96	37,2326596	38,48	0,06	0,15	0,147383213	0,52
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	15096	101	251	250,673887	251	20,71	37,56	37,1430087	38,21	0,08	0,15	0,148310811	0,38
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	15096	101	251	250,965488	251	15,55	37,91	37,0053306	38,45	0,06	0,15	0,146565315	0,38
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_macherey_nagel_MN2_R1	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	17042	247	251	250,730137	251	20,3	37,67	37,248009	38,26	0,08	0,15	0,148445605	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_macherey_nagel_MN2_R2	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	17042	248	251	250,982279	251	15,13	37,98	37,270176	38,51	0,06	0,15	0,1474516	0,16
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	11927	247	251	250,72382	251	20,43	37,63	37,2174042	38,24	0,08	0,15	0,148434644	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	11927	248	251	250,981051	251	15,4	37,96	37,2015293	38,49	0,06	0,15	0,147241553	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_macherey_nagel_MN4_R1	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	9258	72	251	250,694318	251	20,68	37,64	37,2801091	38,25	0,08	0,15	0,148736228	0,53
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_macherey_nagel_MN4_R2	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	9258	96	251	250,968784	251	15,33	37,92	37,0840765	38,45	0,06	0,15	0,146906459	0,39
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_zymo_Z1_R1	Z1	Zymo_Research	NA	2E+07	Illumina	500	11794	105	251	250,782601	251	19,37	37,63	37,1396651	38,24	0,08	0,15	0,148123622	0,36
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_zymo_Z1_R2	Z1	Zymo_Research	NA	2E+07	Illumina	500	11794	105	251	250,875106	251	14,63	37,98	37,1587816	38,48	0,06	0,15	0,147051043	0,36
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_zymo_Z2_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	13808	56	251	250,760429	251	20,31	37,65	37,2053665	38,3	0,08	0,15	0,148499421	0,64
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_zymo_Z2_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	13808	56	251	250,84929	251	15,3	37,96	37,126208	38,47	0,06	0,15	0,147072711	0,49
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_zymo_Z3_R1	Z3	Zymo_Research	NA	2E+07	Illumina	500	10820	58	251	250,774584	251	20,9	37,61	37,2050481	38,23	0,08	0,15	0,14853512	0,64
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_zymo_Z3_R2	Z3	Zymo_Research	NA	2E+07	Illumina	500	10820	58	251	250,874399	251	14,19	37,98	37,2560758	38,48	0,06	0,15	0,147537893	0,65
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_zymo_Z4_R1	Z4	Zymo_Research	NA	2E+07	Illumina	500	14300	40	251	250,725804	251	20,61	37,66	37,253914	38,25	0,08	0,15	0,148631469	0,92
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_v3_4_20220629_zymo_Z4_R2	Z4	Zymo_Research	NA	2E+07	Illumina	500	14300	40	251	250,871329	251	15,21	38	37,3210958	38,47	0,06	0,15	0,14773986	0,93
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_invitrogen_I1_R1	I1	Invitrogen	NA	2E+07	Illumina	500	770141	35	243	214,32046	251	2	38,08	37,0599083	38,91	0,06	0,15	0,190092295	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_invitrogen_I1_R2	I1	Invitrogen	NA	2E+07	Illumina	500	770141	35	247	215,120688	251	2	37,19	35,6693549	38,8	0,06	0,15	0,183676651	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_invitrogen_I2_R1	I2	Invitrogen	NA	2E+07	Illumina	500	2593556	35	249	219,062758	251	2	38,09	37,0902656	38,92	0,06	0,15	0,183879508	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_invitrogen_I2_R2	I2	Invitrogen	NA	2E+07	Illumina	500	2593556	35	250	219,962597	251	2	37	35,342801	38,8	0,05	0,15	0,17598683	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_invitrogen_I3_R1	I3	Invitrogen	NA	2E+07	Illumina	500	838178	35	251	231,220349	251	2	37,93	36,8651366	38,89	0,06	0,15	0,171135833	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_invitrogen_I3_R2	I3	Invitrogen	NA	2E+07	Illumina	500	838178	35	251	231,663266	251	2	36,71	35,0231973	38,82	0,05	0,15	0,163755968	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_invitrogen_I4_R1	I4	Invitrogen	NA	2E+07	Illumina	500	826156	35	250	225,705868	251	2	38,07	37,0418397	38,91	0,06	0,15	0,179236294	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_invitrogen_I4_R2	I4	Invitrogen	NA	2E+07	Illumina	500	826156	35	250	226,174997	251	2	37,23	35,6307884	38,81	0,05	0,15	0,173471378	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	2330316	35	223	204,829757	251	2	37,97	36,8776854	38,91	0,06	0,17	0,201321452	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	2330316	35	226	205,975914	251	2	37,55	35,7287814	38,83	0,05	0,16	0,195527868	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_macherey_nagel_MN2_R1	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	1560588	35	223	206,469636	251	2	38,09	37,0980788	38,9	0,06	0,17	0,198280398	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_macherey_nagel_MN2_R2	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	1560588	35	228	208,173137	251	2	36,88	35,059847	38,83	0,05	0,16	0,187731022	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	729000	35	251	229,918167	251	2	37,91	36,8121841	38,91	0,06	0,15	0,172641756	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	729000	35	251	230,441235	251	2	36,74	35,0365362	38,82	0,05	0,15	0,165380713	1,08

Supplementary Data 5a. The statistical data of all sequencing performed during the project. a. Data obtained from a single sample of a 13-year-old dog.

Source	Breed (name of the dog)	Age at sampling	Sex	name	sample	kit	barcode	date	seq_type	cycle	read_count	length min	length_median	length mean	length_max	quality_min	quality_median	quality mean	quality_max	avg base quality min	avg base quality median	avg base quality mean	avg base quality max
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_macherey_nagel_MN4_R1	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	796251	35	250	226,410353	251	2	37,99	36,9225171	38,9	0,06	0,15	0,178421704	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_macherey_nagel_MN4_R2	MN4	Macherey-Nagel	NA	2E+07	Illumina	500	796251	35	250	227,136144	251	2	36,63	34,7407552	38,8	0,05	0,15	0,169142469	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_qiagen_Q1_R1	Q1	Qiagen	NA	2E+07	Illumina	500	570785	35	242	201,051703	251	2	37,95	36,8356854	38,87	0,06	0,15	0,221314593	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_qiagen_Q1_R2	Q1	Qiagen	NA	2E+07	Illumina	500	570785	35	246	201,911772	251	2	37,01	35,4682405	38,8	0,06	0,15	0,214562033	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_qiagen_Q2_R1	Q2	Qiagen	NA	2E+07	Illumina	500	543153	35	249	202,946969	251	2	37,94	36,8058011	38,88	0,06	0,15	0,220982817	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_qiagen_Q2_R2	Q2	Qiagen	NA	2E+07	Illumina	500	543153	35	250	203,468163	251	2	37,36	35,9061142	38,78	0,05	0,15	0,217170944	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_qiagen_Q3_R1	Q3	Qiagen	NA	2E+07	Illumina	500	769924	35	189	178,031646	251	2	38,07	36,9258986	38,87	0,06	0,2	0,254572231	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_qiagen_Q3_R2	Q3	Qiagen	NA	2E+07	Illumina	500	769924	35	190	179,211596	251	2	37,49	35,9056714	38,8	0,06	0,19	0,248313509	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_qiagen_Q4_R1	Q4	Qiagen	NA	2E+07	Illumina	500	668908	35	213	186,117889	251	2	37,85	36,5234991	38,89	0,06	0,17	0,249886322	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_qiagen_Q4_R2	Q4	Qiagen	NA	2E+07	Illumina	500	668908	35	217	187,699847	251	2	36,72	34,7486706	38,78	0,06	0,16	0,239749741	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_zymo_Z1_R1	Z1	Zymo_Research	NA	2E+07	Illumina	500	544097	35	250	227,547586	251	2	37,79	36,6462413	38,88	0,06	0,15	0,174882898	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_zymo_Z1_R2	Z1	Zymo_Research	NA	2E+07	Illumina	500	544097	35	250	228,131943	251	2	36,6	34,9087756	38,79	0,06	0,15	0,167646339	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_zymo_Z2_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	773774	35	250	228,775487	251	2	37,77	36,6234226	38,87	0,06	0,15	0,172626452	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_zymo_Z2_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	773774	35	250	229,432137	251	2	36,35	34,6662314	38,78	0,05	0,15	0,164407282	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_zymo_Z3_R1	Z3	Zymo_Research	NA	2E+07	Illumina	500	718720	35	250	226,177314	251	2	37,83	36,6914707	38,86	0,06	0,15	0,175799463	1,07
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_zymo_Z3_R2	Z3	Zymo_Research	NA	2E+07	Illumina	500	718720	35	250	227,178765	251	2	35,93	34,1143898	38,78	0,06	0,15	0,164517976	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_zymo_Z4_R1	Z4	Zymo_Research	NA	2E+07	Illumina	500	445910	35	250	226,430367	251	2	37,86	36,723466	38,85	0,06	0,15	0,177614631	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220201_zymo_Z4_R2	Z4	Zymo_Research	NA	2E+07	Illumina	500	445910	35	250	226,955816	251	2	36,92	35,3023946	38,79	0,06	0,15	0,171789689	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_invitrogen_I1_R1	I1	Invitrogen	NA	2E+07	Illumina	300	984783	35	151	145,252943	151	2	38,17	37,404984	38,82	0,06	0,25	0,264163029	1,07
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_invitrogen_I1_R2	I1	Invitrogen	NA	2E+07	Illumina	300	984783	35	151	145,277727	151	2	38,05	37,1296393	38,78	0,06	0,25	0,262138288	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_invitrogen_I2_R1	I2	Invitrogen	NA	2E+07	Illumina	300	572258	35	151	146,090949	151	2	38,19	37,4352035	38,79	0,06	0,25	0,261758123	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_invitrogen_I2_R2	I2	Invitrogen	NA	2E+07	Illumina	300	572258	35	151	146,112974	151	2	38,11	37,1576508	38,79	0,06	0,25	0,259752245	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_invitrogen_I3_R1	I3	Invitrogen	NA	2E+07	Illumina	300	1116312	35	151	147,185378	151	2	38,11	37,5361076	38,78	0,06	0,25	0,258119065	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_invitrogen_I3_R2	I3	Invitrogen	NA	2E+07	Illumina	300	1116312	35	151	147,201972	151	2	38,07	37,0859147	38,79	0,06	0,25	0,256463578	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_invitrogen_I4_R1	I4	Invitrogen	NA	2E+07	Illumina	300	700918	35	151	145,753724	151	2	38,19	37,412051	38,79	0,06	0,25	0,263458265	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_invitrogen_I4_R2	I4	Invitrogen	NA	2E+07	Illumina	300	700918	35	151	145,784418	151	2	38,12	37,1237811	38,77	0,06	0,25	0,261355123	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	300	517938	35	151	143,513716	151	2	38,1	37,3016048	38,79	0,06	0,25	0,268746182	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	300	517938	35	151	143,547965	151	2	38,09	37,2016398	38,76	0,06	0,25	0,268048299	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_macherey_nagel_MN2_R1	MN2	Macherey-Nagel	NA	2E+07	Illumina	300	342199	35	151	144,546019	151	2	38,19	37,4712363	38,8	0,06	0,25	0,266151216	1,07
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_macherey_nagel_MN2_R2	MN2	Macherey-Nagel	NA	2E+07	Illumina	300	342199	35	151	144,615534	151	2	37,97	36,8280063	38,79	0,06	0,25	0,261368151	1,07
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	300	666621	35	151	146,68375	151	2	38,12	37,3402734	38,79	0,06	0,25	0,259766539	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	300	666621	35	151	146,712916	151	2	37,99	36,9184844	38,77	0,06	0,25	0,256810242	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_macherey_nagel_MN4_R1	MN4	Macherey-Nagel	NA	2E+07	Illumina	300	543061	35	151	145,615831	151	2	38,16	37,4140829	38,79	0,06	0,25	0,263879969	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_macherey_nagel_MN4_R2	MN4	Macherey-Nagel	NA	2E+07	Illumina	300	543061	35	151	145,661202	151	2	37,91	36,6216357	38,75	0,06	0,25	0,258302309	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_qiagen_Q1_R1	Q1	Qiagen	NA	2E+07	Illumina	300	621566	35	151	137,377423	151	2	38,06	37,0995716	38,79	0,06	0,25	0,290666591	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_qiagen_Q1_R2	Q1	Qiagen	NA	2E+07	Illumina	300	621566	35	151	137,433095	151	2	37,91	36,7977973	38,77	0,06	0,25	0,288370439	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_qiagen_Q2_R1	Q2	Qiagen	NA	2E+07	Illumina	300	483624	35	151	136,82437	151	2	38,06	37,0893635	38,77	0,06	0,25	0,292430214	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_qiagen_Q2_R2	Q2	Qiagen	NA	2E+07	Illumina	300	483624	35	151	136,876398	151	2	38,05	37,0488579	38,79	0,06	0,25	0,292262915	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_qiagen_Q3_R1	Q3	Qiagen	NA	2E+07	Illumina	300	644304	35	150	132,290486	151	2	38,13	37,1082854	38,79	0,06	0,25	0,307296012	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_qiagen_Q3_R2	Q3	Qiagen	NA	2E+07	Illumina	300	644304	35	150	132,38658	151	2	38,06	36,9191037	38,77	0,06	0,25	0,305575862	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_qiagen_Q4_R1	Q4	Qiagen	NA	2E+07	Illumina	300	508598	35	150	131,658571	151	2	37,99	36,8552529	38,77	0,06	0,25	0,310226584	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_qiagen_Q4_R2	Q4	Qiagen	NA	2E+07	Illumina	300	508598	35	150	131,780575	151	2	37,91	36,544896	38,76	0,06	0,25	0,307672818	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_zymo_Z1_R1	Z1	Zymo_Research	NA	2E+07	Illumina	300	1809879	35	151	146,316329	151	2	38,07	37,2674457	38,77	0,06	0,25	0,260167553	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_zymo_Z1_R2	Z1	Zymo_Research	NA	2E+07	Illumina	300	1809879	35	151	146,336012	151	2	37,99	36,9267599	38,78	0,06	0,25	0,257898931	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_zymo_Z2_R1	Z2	Zymo_Research	NA	2E+07	Illumina	300	990977	35	151	146,497223	151	2	38,03	37,1909601	38,78	0,06	0,25	0,259184714	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_zymo_Z2_R2	Z2	Zymo_Research	NA	2E+07	Illumina	300	990977	35	151	146,508985	151	2	38	36,9965083	38,79	0,06	0,25	0,257942132	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_zymo_Z3_R1	Z3	Zymo_Research	NA	2E+07	Illumina	300	982128	35	151	146,236466	151	2	38,09	37,2870714	38,79	0,06	0,25	0,260432887	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_zymo_Z3_R2	Z3	Zymo_Research	NA	2E+07	Illumina	300	982128	35	151	146,266339	151	2	37,78	36,423381	38,75	0,06	0,25	0,254554142	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_zymo_Z4_R1	Z4	Zymo_Research	NA	2E+07	Illumina	300	894390	35	151	145,708464	151	2	38,07	37,2376927	38,75	0,06	0,25	0,26231048	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_illumina_WGS_20220331_zymo_Z4_R2	Z4	Zymo_Research	NA	2E+07	Illumina	300	894390	35	151	145,733301	151	2	38,03	37,0416608	38,74	0,06	0,25	0,261008956	1,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_invitrogen_I1_bc09_pass	I1	Invitrogen	bc09	2E+07	ONT	NA	387216	129	1570	1552,55457	25769	13,64	19,94	19,9045601	40,44	0	0,01	0,011017003	0,17
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_invitrogen_I2_bc10_pass	I2	Invitrogen	bc10	2E+07	ONT	NA	3007719	93	1571	1554,13369	14688	13,17	19,92	19,8846835	53,7	0	0,01	0,010993474	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_invitrogen_I3_bc11_pass	I3	Invitrogen	bc11	2E+07	ONT	NA	738637	97	1570	1552,35546	7217	13,08	19,98	19,9372984	46,22	0	0,01	0,011037966	0,21
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_invitrogen_I4_bc12_pass	I4	Invitrogen	bc12	2E+07	ONT	NA	411406	90	1569	1551,95594	50873	13,67	19,91	19,8748847	48,73	0	0,01	0,010986641	0,23
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_macherey_nagel_MN1_bc05_pass	MN1	Macherey-Nagel	bc05	2E+07	ONT	NA	1677	179	1572	1539,78474	3936	14,41	19,84	19,8066369	28,38	0,01	0,01	0,011466905	0,11
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_macherey_nagel_MN2_bc06_pass	MN2	Macherey-Nagel	bc06	2E+07	ONT	NA	655079	91	1571	1553,53813	23079	13,89	19,92	19,8798951	43,37	0	0,01	0,010956526	0,24
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_macherey_nagel_MN3_bc07_pass	MN3	Macherey-Nagel	bc07	2E+07	ONT	NA	276858	129	1570	1551,09872	6604	13,72	19,92	19,8850109	35,74	0	0,01	0,011035224	0,17
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_macherey_nagel_MN4_bc08_pass	MN4	Macherey-Nagel	bc08	2E+07	ONT	NA	292982	125	1572	1553,61931	52060	13,67	19,92	19,8882089	35,26	0	0,01	0,010950775	0,17
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_qiagen_Q1_bc01_pass	Q1	Qiagen	bc01	2E+07	ONT	NA	4	1391	1533	1511	1587	17,39	19,425	19,5825	22,09	0,01	0,01	0,0125	0,02
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_qiagen_Q2_bc02_pass	Q2	Qiagen	bc02	2E+07	ONT	NA	12	1177	1593	1690,83333	3144	16,76	20,07	19,8891667	22,11	0,01	0,01	0,010833333	0,02
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_qiagen_Q3_bc03_pass	Q3	Qiagen	bc03	2E+07	ONT	NA	9	1523	1565	1590,55556	1863	16,72	18,58	19,2588889	22,52	0,01	0,01	0,01	0,01
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211209_qiagen_Q4_bc04_pass	Q4	Qiagen	bc04	2E+07	ONT	NA	17	1386	1561	1568,82353	1814	15,78	18,72	19,3694118	23,42	0,01	0,01	0,01	0,01
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211215_zymo_Z1_bc01_pass	Z1	Zymo_Research	bc01	2E+07	ONT	NA	972560	107	1580	1562,67905	8045	13,77	20,62	20,473072	38,31	0	0,01	0,011183855	0,18
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211215_zymo_Z2_bc02_pass	Z2	Zymo_Research	bc02	2E+07	ONT	NA	1348071	121	1586	1566,19385	5202	13,71	20,7	20,5441458	41,44	0	0,01	0,011183031	0,26
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211215_zymo_Z3_bc03_pass	Z3	Zymo_Research	bc03	2E+07	ONT	NA	1016927	139	1580	1559,44226	7254	13,7	20,58	20,4393906	37,99	0	0,01	0,011195832	0,14
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20211215_zymo_Z4_bc04_pass	Z4	Zymo_Research	bc04	2E+07	ONT	NA	1029868	128	1582	1563,08146	7393	13,28	20,6	20,4611359	36,87	0	0,01	0,011178666	0,16
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_invitrogen_I1_bc09_pass	I1	Invitrogen	bc09	2E+07	ONT	NA	87397	135	1572	1548,31088	7474	13,87	20,32	20,2127895	33,87	0	0,01	0,011172123	0,14
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_invitrogen_I2_bc02_pass	I2	Invitrogen	bc02	2E+07	ONT	NA	117483	150	1574	1542,85355	6593	13,8	20,45	20,3295909	36,42	0	0,01	0,011417567	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_invitrogen_I3_bc03_pass	I3	Invitrogen	bc03	2E+07	ONT	NA	56257	147	1572	1545,37572	7100	13,86	20,3	20,1875107	33,93	0	0,01	0,011269886	0,14
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_invitrogen_I4_bc04_pass	I4	Invitrogen	bc04	2E+07	ONT	NA	144428	131	1570	1545,46191	8307	13,97	20,35	20,235684	38,63	0	0,01	0,011246503	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_macherey_nagel_MN1_bc05_pass	MN1	Macherey-Nagel	bc05	2E+07	ONT	NA	2734	204	1574	1548,46086	3199	14,6	20,26	20,1509108	25,36	0,01	0,01	0,011203365	0,11
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_macherey_nagel_MN2_bc10_pass	MN2	Macherey-Nagel	bc10	2E+07	ONT	NA	109725	147	1572	1542,93416	3332	13,93	20,33	20,2177916	36,68	0,01	0,01	0,0112689	0,14
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_macherey_nagel_MN3_bc11_pass	MN3	Macherey-Nagel	bc11	2E+07	ONT	NA	147302	137	1575	1550,18619	3631	13,85	20,4	20,2792669	35,95	0	0,01	0,011168144	0,16
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_macherey_nagel_MN4_bc12_pass	MN4	Macherey-Nagel	bc12	2E+07	ONT	NA	74050	146	1573	1549,00989	4647	13,72	20,32	20,2050843	30,2	0	0,01	0,011136529	0,14
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_zymo_Z1_bc01_pass	Z1	Zymo_Research	bc01	2E+07	ONT	NA	45592	142	1576	1547,37053	9705	14,01	20,37	20,2522822	34,57	0	0,01	0,011281584	0,13
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_zymo_Z2_bc06_pass	Z2	Zymo_Research	bc06	2E+07	ONT	NA	36509	173	1581	1552,22701	3312	13,78	20,36	20,2278515	30,87	0,01	0,01	0,011156153	0,11
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_zymo_Z3_bc07_pass	Z3	Zymo_Research	bc07	2E+07	ONT	NA	20777	154	1576	1547,05564	5392	14,21	20,37	20,2447764	26,81	0	0,01	0,011369784	0,13
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_minion_v1_9_20220325_zymo_Z4_bc08_pass	Z4	Zymo_Research	bc08	2E+07	ONT	NA	45250	154	1580	1553,03483	6937	13,88	20,33	20,2062683	30,84	0	0,01	0,011116906	0,13
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_invitrogen_I1	I1	Invitrogen	NA	2E+07	PacBio	NA	23882	559	1486	1481,42358	3028	38,51	92,76	88,8126145	93	0,01	0,06	0,058095637	0,17
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_invitrogen_I2	I2	Invitrogen	NA	2E+07	PacBio	NA	18781	1445	1490	1482,52724	3025	37,32	92,74	88,8015494	93	0,03	0,06	0,058066131	0,06
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_invitrogen_I3	I3	Invitrogen	NA	2E+07	PacBio	NA	15535	1410	1486	1481,32932	3017	39,32	92,78	88,8292198	93	0,02	0,06	0,058058577	0,06
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_invitrogen_I4	I4	Invitrogen	NA	2E+07	PacBio	NA	40759	753	1486	1481,20194	3034	38,31	92,75	88,7931721	93	0,03	0,06	0,058057852	0,12
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_macherey_nagel_MN1	MN1	Macherey-Nagel	NA	2E+07	PacBio	NA	8054	1446	1491	1483,97715	3022	39,99	92,81	88,7828098	93	0,02	0,06	0,057974919	0,06
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_macherey_nagel_MN2	MN2	Macherey-Nagel	NA	2E+07	PacBio	NA	14272	1101	1491	1482,95116	3013	39,26	92,83	88,7946013	93	0,02	0,06	0,05799958	0,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_macherey_nagel_MN3	MN3	Macherey-Nagel	NA	2E+07	PacBio	NA	20782	846	1491	1482,85228	3022	40,05	92,8	88,8833808	93	0,03	0,06	0,058101241	0,11
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_macherey_nagel_MN4	MN4	Macherey-Nagel	NA	2E+07	PacBio	NA	55385	1368	1491	1482,73283	3052	39,58	92,78	88,8697232	93	0,02	0,06	0,058096055	0,07
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_zymo_Z1	Z1	Zymo_Research	NA	2E+07	PacBio	NA	37905	1151	1489	1485,12426	3029	37,94	92,8	88,9247221	93	0,02	0,06	0,058090489	0,08
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_zymo_Z2	Z2	Zymo_Research	NA	2E+07	PacBio	NA	20295	582	1491	1487,76945	2284	39,82	92,77	88,785863	93	0,03	0,06	0,058024637	0,15
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_zymo_Z3	Z3	Zymo_Research	NA	2E+07	PacBio	NA	17119	1446	1489	1484,02886	3012	38,13	92,8	88,8637456	93	0,03	0,06	0,05806706	0,06
dog (stool)	Pumi (Harsányi-Tekergő Bátor 'Toto')	13,5 years	male	toti_pacbio_v1_9_20220722_zymo_Z4	Z4	Zymo_Research	NA	2E+07	PacBio	NA	18349	1446	1491	1487,48459	3040	39,59	92,76	88,7620181	93	0,03	0,06	0,058004796	0,06

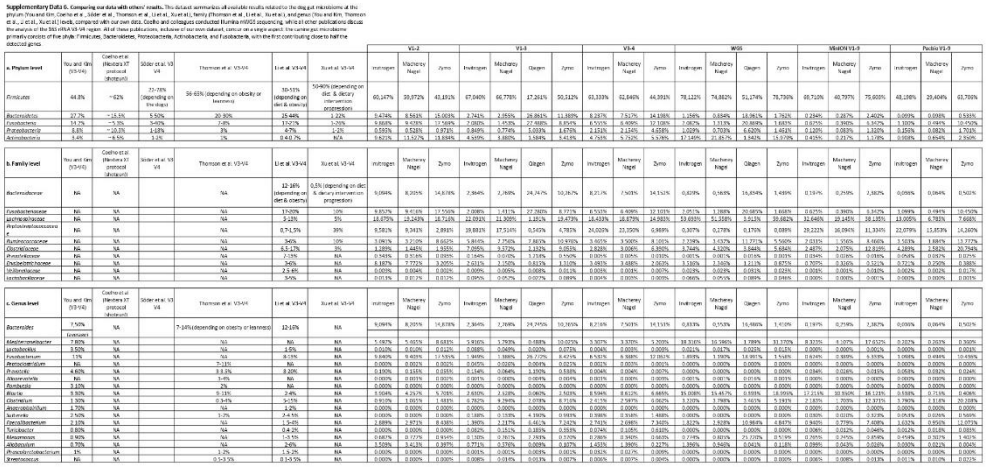
b. The statistical data of the ONT sequencing performed on samples from the six control dogs.

source	breed (name of the dog)	Age at sampling	Sex	name	sample	kit	barcode	date	seq type	cycle	read count	length min	length median	length mean	length max	quality min	quality median	quality mean	quality max	avg base quality min	avg_base_quality_median	avg_base_quality_mean	avg_base_quality_max
dog (stool)	Pumi (Duna-menti Dumás Blöff)	26 weeks	male	Bloff01_DMD_minion_v1_9_20221011_macherey_nagel_bc09	Bloff01	Macherey-Nagel	bc09	2E+07	ONT	NA	241411	134	1573	1554,32103	4781	11,11	18,9	18,6316832	34,92	0	0,01	0,010643136	0,14
dog (stool)	Pumi (Duna-menti Dumás Borcsa)	26 weeks	femal	Borcsa01_DMD_minion_v1_9_20221011_macherey_nagel_bc05	Borcsa0	Macherey-Nagel	bc05	2E+07	ONT	NA	246197	148	1563	1504,87333	4854	11,48	19,04	18,75747	38,72	0	0,01	0,011678859	0,14
dog (stool)	Pumi (Duna-menti Dumás Boszi)	26 weeks	femal	Boszi01_DMD_minion_v1_9_20221011_macherey_nagel_bc01	Boszi01	Macherey-Nagel	bc01	2E+07	ONT	NA	201090	136	1569	1534,89081	4803	11,55	19,01	18,7349359	42,2	0	0,01	0,011011786	0,14
dog (stool)	Pumi (Duna-menti Dumás Briós)	21 weeks	male	Brios01_DMD_minion_v1_9_20221011_macherey_nagel_bc13	Brios01	Macherey-Nagel	bc13	2E+07	ONT	NA	774014	134	1559	1460,12811	4905	10,83	19,01	18,7289659	46,18	0	0,01	0,012819309	0,14
dog (stool)	Pumi (Vöröskői Kondacsipkedő Lidérc 'Loki')	6,7 years	male	Loki01_DMD_minion_v1_9_20221011_macherey_nagel_bc17	Loki01	Macherey-Nagel	bc17	2E+07	ONT	NA	278320	144	1576	1548,70895	4777	11,44	18,88	18,6234317	45,41	0	0,01	0,010837094	0,14
dog (stool)	Pumi (Galla-Hegyi Suba 'Sugó')	7,2 years	femal	Sugo01_DMD_minion_v1_9_20221011_macherey_nagel_bc21	Sugo01	Macherey-Nagel	bc21	2E+07	ONT	NA	309374	127	1572	1541,95891	6006	11,72	19,17	18,8897526	49,85	0	0,01	0,010949983	0,14

c. The statistics of the sequencing of libraries made from the microbial mix.

source	name of the kit	name	sample	kit	barcode	date	seq type	cycle	read count	length min	length median	length mean	length max	quality min	quality median	quality mean	quality max	avg base quality min	avg_base_quality_median	avg_base_quality_mean	avg_base_quality_max
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_invitrogen_I1I_R1	I1	Invitrogen	NA	2E+07	Illumina	500	44200	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_invitrogen_I1I_R2	I1	Invitrogen	NA	2E+07	Illumina	500	44200	240	250	250,006041	251	14,6	24,73	24,3855224	32,78	0,06	0,1	0,097403167	0,13
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_invitrogen_I1_R1	I1	Invitrogen	NA	2E+07	Illumina	500	56568	240	251	250,868601	251	17,09	25,51	25,2280703	30,63	0,07	0,1	0,100536169	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_invitrogen_I1_R2	I1	Invitrogen	NA	2E+07	Illumina	500	56568	244	250	250,012339	251	15,12	21,98	21,7508176	29,33	0,06	0,09	0,086982393	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_invitrogen_I2_R1	I2	Invitrogen	NA	2E+07	Illumina	500	59093	245	251	250,848239	251	17,98	25,46	25,2035739	30,31	0,07	0,1	0,10045403	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_invitrogen_I2_R2	I2	Invitrogen	NA	2E+07	Illumina	500	59093	244	250	250,006295	251	14,65	24,21	23,8836769	35,42	0,06	0,1	0,095556496	0,14
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	71089	245	251	250,906005	251	17,45	25,59	25,3431248	30,82	0,07	0,1	0,101037432	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	71089	244	250	249,999831	251	14,58	23,67	23,3457313	30,36	0,06	0,09	0,093463264	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_macherey_nagel_MN2I_R1	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	61673	245	251	250,884747	251	17,72	25,92	25,6531965	30,63	0,07	0,1	0,102268416	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_macherey_nagel_MN2I_R2	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	61673	241	250	250,000032	251	14,63	23,91	23,5949803	29,3	0,06	0,1	0,094444408	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	71146	242	251	250,918773	251	18,18	25,92	25,6620711	30,42	0,07	0,1	0,102284317	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	71146	238	250	250,001912	251	15,13	24,24	23,8870399	34,43	0,06	0,1	0,09553847	0,14
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_qiagen_Q1_R1	Q1	Qiagen	NA	2E+07	Illumina	500	55853	241	251	250,942152	251	18,01	25,8	25,5092142	30,24	0,07	0,1	0,101677976	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_qiagen_Q1_R2	Q1	Qiagen	NA	2E+07	Illumina	500	55853	243	250	249,999445	251	14,78	24,95	24,620465	35,09	0,06	0,1	0,098344762	0,14
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_qiagen_Q2I_R1	Q2	Qiagen	NA	2E+07	Illumina	500	35525	246	251	250,938804	251	18,27	26,11	25,7869393	30,55	0,07	0,1	0,10279437	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_qiagen_Q2I_R2	Q2	Qiagen	NA	2E+07	Illumina	500	35525	244	250	249,997945	251	15,11	24,76	24,372916	32,99	0,06	0,1	0,097576214	0,13
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_qiagen_Q2_R1	Q2	Qiagen	NA	2E+07	Illumina	500	66969	245	251	250,879497	251	17,99	25,88	25,6323314	30,49	0,07	0,1	0,102176529	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_qiagen_Q2_R2	Q2	Qiagen	NA	2E+07	Illumina	500	66969	244	250	250,000209	251	13,82	23,86	23,5324289	30,9	0,06	0,1	0,094206424	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_zymo_Z1_R1	Z1	Zymo_Research	NA	2E+07	Illumina	500	59420	245	251	250,909037	251	16,11	25,75	25,460791	30,24	0,06	0,1	0,1014793	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_zymo_Z1_R2	Z1	Zymo_Research	NA	2E+07	Illumina	500	59420	244	250	250,003921	251	14,81	24,32	23,9591831	36,06	0,06	0,1	0,095844833	0,14
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_zymo_Z2I_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	60036	245	251	250,880205	251	17,49	25,7	25,421744	30,47	0,07	0,1	0,101332201	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_zymo_Z2I_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	60036	244	250	250,002082	251	14,91	24,71	24,3640161	30,77	0,06	0,1	0,097372243	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_zymo_Z2_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	37680	245	251	250,909926	251	17,58	26,01	25,7361269	30,27	0,07	0,1	0,102610934	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221104_zymo_Z2_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	37680	244	250	249,999602	251	14,56	24,46	24,1448726	33,73	0,06	0,1	0,096527866	0,13
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_invitrogen_I1I_R1	I1	Invitrogen	NA	2E+07	Illumina	500	34332	248	251	250,375626	251	16,44	36,92	35,4463262	38,07	0,07	0,15	0,141849878	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_invitrogen_I1I_R2	I1	Invitrogen	NA	2E+07	Illumina	500	34332	244	250	249,823139	251	15,12	37,57	36,3613416	38,23	0,06	0,15	0,145090586	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_invitrogen_I1_R1	I1	Invitrogen	NA	2E+07	Illumina	500	43292	247	251	250,48099	251	19,31	36,88	35,4262587	38,09	0,08	0,15	0,14170863	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_invitrogen_I1_R2	I1	Invitrogen	NA	2E+07	Illumina	500	43292	248	250	249,885037	251	15,19	36,67	35,1111695	38,18	0,06	0,15	0,140866211	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_invitrogen_I2_R1	I2	Invitrogen	NA	2E+07	Illumina	500	45377	247	251	250,382837	251	18,82	36,9	35,4378401	38,11	0,07	0,15	0,141789673	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_invitrogen_I2_R2	I2	Invitrogen	NA	2E+07	Illumina	500	45377	247	250	249,869051	251	14,82	37,52	36,291694	38,22	0,06	0,15	0,144883531	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_macherey_nagel_MN1_R1	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	54337	247	251	250,855274	251	18,4	37,34	36,0577976	38,1	0,07	0,15	0,143894032	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_macherey_nagel_MN1_R2	MN1	Macherey-Nagel	NA	2E+07	Illumina	500	54337	244	250	250,232696	251	15,36	37,45	36,2218821	38,22	0,06	0,15	0,144653367	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_macherey_nagel_MN2I_R1	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	53513	248	251	250,893035	251	19,02	37,44	36,1978637	38,07	0,08	0,15	0,144409396	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_macherey_nagel_MN2I_R2	MN2	Macherey-Nagel	NA	2E+07	Illumina	500	53513	248	250	250,154	251	14,9	37,51	36,2449145	38,26	0,06	0,15	0,144646908	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_macherey_nagel_MN3_R1	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	56916	247	251	250,863641	251	19,84	37,45	36,1781506	38,1	0,08	0,15	0,144297386	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_macherey_nagel_MN3_R2	MN3	Macherey-Nagel	NA	2E+07	Illumina	500	56916	244	250	250,133056	251	15,17	37,51	36,1888792	38,26	0,06	0,15	0,144463947	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_qiagen_Q1_R1	Q1	Qiagen	NA	2E+07	Illumina	500	43471	247	251	250,959283	251	18,47	37,49	36,3204129	38,11	0,07	0,15	0,144803432	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_qiagen_Q1_R2	Q1	Qiagen	NA	2E+07	Illumina	500	43471	244	250	250,285452	251	14,7	37,62	36,4852414	38,26	0,06	0,15	0,145439948	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_qiagen_Q2I_R1	Q2	Qiagen	NA	2E+07	Illumina	500	28308	248	251	250,967748	251	20,87	37,61	36,5385004	38,12	0,08	0,15	0,145546842	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_qiagen_Q2I_R2	Q2	Qiagen	NA	2E+07	Illumina	500	28308	248	250	250,343719	251	15	37,65	36,4902632	38,24	0,06	0,15	0,145376925	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_qiagen_Q2_R1	Q2	Qiagen	NA	2E+07	Illumina	500	53932	38	251	250,906085	251	18,55	37,47	36,2615907	38,08	0,07	0,15	0,144660498	0,98
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_qiagen_Q2_R2	Q2	Qiagen	NA	2E+07	Illumina	500	53932	38	250	250,156327	251	13,84	37,49	36,2376539	38,22	0,06	0,15	0,144693132	0,96
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_zymo_Z1_R1	Z1	Zymo_Research	NA	2E+07	Illumina	500	49391	247	251	250,701444	251	16,89	37,27	35,8944178	38,1	0,07	0,15	0,143318418	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_zymo_Z1_R2	Z1	Zymo_Research	NA	2E+07	Illumina	500	49391	244	250	250,043793	251	15,23	37,49	36,1774356	38,25	0,06	0,15	0,144447369	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_zymo_Z2I_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	45782	248	251	250,8817	251	20,04	37,33	36,0312739	38,08	0,08	0,15	0,143784457	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_zymo_Z2I_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	45782	248	250	250,117754	251	14,2	37,54	36,3511751	38,25	0,06	0,15	0,145067712	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_zymo_Z2_R1	Z2	Zymo_Research	NA	2E+07	Illumina	500	32672	248	251	250,724474	251	18,81	37,39	36,0998494	38,08	0,08	0,15	0,144058827	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_illumina_v1_v2_20221110_zymo_Z2_R2	Z2	Zymo_Research	NA	2E+07	Illumina	500	32672	244	250	250,045788	251	14,99	37,57	36,3243325	38,29	0,06	0,15	0,144899608	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_invitrogen_I1_bc07_pass	I1	Invitrogen	bc07	2E+07	ONT	NA	98785	147	1611	1592,11923	6865	13,85	21,95	21,6483725	38,01	0	0,01	0,011673331	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_invitrogen_I2_bc08_pass	I2	Invitrogen	bc08	2E+07	ONT	NA	108616	153	1611	1595,4249	4820	13,96	21,91	21,6163925	35,83	0	0,01	0,011501528	0,15
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_invitrogen_I3_bc09_pass	I3	Invitrogen	bc09	2E+07	ONT	NA	112478	167	1611	1596,25592	7558	13,94	21,96	21,6635366	38,88	0	0,01	0,011554793	0,13
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_macherey_nagel_MN1_bc04_pass	MN1	Macherey-Nagel	bc04	2E+07	ONT	NA	129460	152	1602	1566,51404	4849	14,08	21,81	21,5095046	38,03	0	0,01	0,011851151	0,13
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_macherey_nagel_MN2_bc05_pass	MN2	Macherey-Nagel	bc05	2E+07	ONT	NA	2	1146	1865	1865	2584	21,99	22,685	22,685	23,38	0,01	0,015	0,015	0,02
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_macherey_nagel_MN3_bc06_pass	MN3	Macherey-Nagel	bc06	2E+07	ONT	NA	129552	140	1604	1585,48299	4890	13,93	21,84	21,5379351	33,95	0	0,01	0,011461575	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_qiagen_Q1_bc10_pass	Q1	Qiagen	bc10	2E+07	ONT	NA	124598	166	1603	1579,1413	4202	14,13	21,8	21,4897658	55,8	0	0,01	0,011504599	0,13
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_qiagen_Q2_bc11_pass	Q2	Qiagen	bc11	2E+07	ONT	NA	80928	168	1603	1588,20835	6922	14,23	21,85	21,5312819	40,31	0	0,01	0,011376409	0,13
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_qiagen_Q3_bc12_pass	Q3	Qiagen	bc12	2E+07	ONT	NA	112156	157	1603	1586,50192	5860	14,11	21,77	21,47023	38,17	0	0,01	0,011367381	0,12
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_zymo_Z1_bc01_pass	Z1	Zymo_Research	bc01	2E+07	ONT	NA	118856	129	1604	1557,07405	3590	13,67	21,85	21,542172	34,64	0,01	0,01	0,012152268	0,13
bacteria mix	Microbial Community Standard (Zymo Research)	BACMIX_minion_v1_9_20220720_zymo_Z2_bc02_pass	Z2	Zymo_Research	bc02	2E+07	ONT	NA	94064	151	1607	1590,18969	6589	14,2	21,97	21,66464	37,45	0	0,01	0,01158534	0,14
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

10.6. Supplementary Data 6



Supplementary Data 6. Comparing our data with others' results. This dataset summarizes all available results related to the dog gut microbiome at the phylum (Yasuo and Ilori, Coelho et al., Sidder et al., Thomson et al., Li et al., Xu et al.), family (Thomson et al., Li et al., Xu et al.), and genus (Xu and Ilori, Thomson et al., Li et al., Xu et al.) levels, compared with our own data. Coelho and colleagues conducted Illumina HiSeq sequencing, while all other publications discuss the analysis of the 16S rRNA V3-V4 region. All of these publications, inclusive of our own dataset, concur on a single aspect: the ruminal gut microbiome primarily consists of five phyla: Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, and Fusobacteria, with the first contributing close to half the microbial genes.

a. Phylum level	You and Ilori (V3-V4)	Coelho et al. (Illumina V7 protocol (shotgun))	Sidder et al. V3 V4	Thomson et al. V3-V4	Li et al. V3-V4	Xu et al. V3-V4	V1-2			V1-3				V3-4			W45				MetON V1-3				Packo V1-3		
							Invitrogen	Macherey Nagel	Zymo	Invitrogen	Macherey Nagel	Qiagen	Zymo	Invitrogen	Macherey Nagel	Zymo	Invitrogen	Macherey Nagel	Qiagen	Zymo	Invitrogen	Macherey Nagel	Zymo	Invitrogen	Macherey Nagel	Zymo
Firmicutes	44.8%	~62%	72-78% (depending on the dogs)	56-65% (depending on obesity or leanness)	70-93% (depending on diet & obesity)	50-80% (depending on diet & dietary intervention progression)	60,147%	39,072%	43,191%	67,040%	66,778%	47,261%	50,132%	63,333%	62,644%	44,391%	78,122%	74,882%	51,274%	79,730%	85,710%	40,797%	75,603%	48,198%	29,404%	61,790%
Bacteroidetes	37.7%	~16.5%	5-9%	20-30%	15-44%	1-22%	0,434%	8,561%	15,003%	2,741%	2,955%	26,841%	11,889%	8,267%	7,512%	14,108%	1,268%	0,854%	16,961%	1,762%	0,234%	0,287%	1,402%	0,09%	0,038%	0,583%
Proteobacteria	9.1%	~10.1%	1-18%	3%	6-7%	1-2%	0,592%	0,528%	0,971%	0,840%	0,774%	1,033%	1,476%	2,152%	2,134%	4,058%	1,028%	0,703%	6,620%	1,461%	0,120%	0,082%	1,321%	0,156%	0,082%	1,701%
Actinobacteria	3.4%	~4.5%	3-5%	1%	0.4-0.7%	N/A	0,421%	13,102%	13,934%	4,107%	3,885%	3,584%	3,413%	4,726%	5,752%	3,238%	17,149%	21,817%	3,341%	11,078%	0,435%	0,213%	1,178%	0,058%	0,454%	2,260%

b. Family level	You and Ilori (V3-V4)	Coelho et al. (Illumina V7 protocol (shotgun))	Sidder et al. V3 V4	Thomson et al. V3-V4	Li et al. V3-V4	Xu et al. V3-V4	V1-2			V1-3				V3-4			W45				MetON V1-3				Packo V1-3		
							Invitrogen	Macherey Nagel	Zymo	Invitrogen	Macherey Nagel	Qiagen	Zymo	Invitrogen	Macherey Nagel	Zymo	Invitrogen	Macherey Nagel	Qiagen	Zymo	Invitrogen	Macherey Nagel	Zymo	Invitrogen	Macherey Nagel	Zymo
Bacteroidaceae	NA	NA		NA	12-34% (depending on diet & obesity)	0,5% (depending on diet & dietary intervention progression)	9,094%	8,205%	14,878%	2,364%	2,488%	24,747%	10,287%	8,21%	7,503%	14,152%	0,829%	0,569%	15,854%	1,499%	0,159%	0,259%	1,382%	0,09%	0,064%	0,502%
Fusobacteriaceae	NA	NA		NA	1-10%	5%	9,857%	9,418%	17,556%	7,008%	1,411%	27,260%	8,771%	8,553%	8,409%	12,103%	2,051%	1,388%	20,685%	1,668%	0,675%	0,366%	3,342%	0,09%	0,194%	0,959%
Prevotellaceae	NA	NA		NA	1-10%	5%	10,475%	15,248%	16,716%	22,092%	11,300%	1,181%	13,473%	15,433%	16,879%	14,061%	53,693%	51,358%	3,612%	39,682%	12,646%	10,145%	38,135%	13,05%	6,78%	7,668%
Porphyromonadaceae	NA	NA		NA	0.5-1.5%	39%	0,581%	0,343%	2,893%	19,881%	17,514%	0,543%	4,783%	26,023%	23,750%	0,969%	0,307%	0,178%	0,170%	0,08%	39,222%	16,094%	31,334%	22,07%	15,853%	14,36%
Ruminococcaceae	NA	NA		NA	34%	10%	3,961%	3,210%	8,602%	1,841%	2,750%	7,881%	10,910%	3,602%	3,500%	8,101%	7,59%	3,157%	11,771%	5,560%	2,531%	1,350%	8,466%	2,55%	1,88%	13,777%
Clostridiaceae	NA	NA		NA	6.5-17%	3%	1,28%	1,445%	3,935%	7,06%	3,173%	1,332%	0,953%	2,82%	3,06%	0,39%	0,744%	4,332%	1,644%	5,464%	2,487%	2,07%	11,916%	4,28%	1,58%	20,764%
Veillonellaceae	NA	NA		NA	1-15%	NA	0,544%	0,534%	0,195%	0,154%	0,014%	1,214%	0,556%	0,057%	0,025%	0,032%	0,081%	0,407%	0,059%	0,03%	0,154%	0,018%	0,08%	0,04%	0,01%	0,025%
Enterobacteriaceae	NA	NA		NA	3-6%	NA	0,187%	0,772%	0,305%	0,018%	0,150%	0,813%	0,110%	0,282%	0,68%	0,263%	0,110%	0,355%	0,212%	0,473%	0,207%	0,310%	0,512%	0,71%	0,22%	0,38%
Erysipelotrichaceae	NA	NA		NA	2-3.4%	NA	0,003%	0,004%	0,002%	0,009%	0,004%	0,004%	0,011%	0,005%	0,003%	0,007%	0,018%	0,018%	0,012%	0,024%	0,003%	0,001%	0,010%	0,002%	0,002%	0,017%
Lachnospiraceae	NA	NA		NA	3-5%	NA	0,011%	0,017%	0,012%	0,005%	0,002%	0,007%	0,008%	0,005%	0,005%	0,003%	0,006%	0,005%	0,008%	0,006%	0,003%	0,003%	0,007%	0,002%	0,002%	0,005%

c. Genus level	You and Ilori (V3-V4)	Coelho et al. (Illumina V7 protocol (shotgun))	Sidder et al. V3 V4	Thomson et al. V3-V4	Li et al. V3-V4	Xu et al. V3-V4	V1-2			V1-3				V3-4			W45				MetON V1-3				Packo V1-3		
							Invitrogen	Macherey Nagel	Zymo	Invitrogen	Macherey Nagel	Qiagen	Zymo	Invitrogen	Macherey Nagel	Zymo	Invitrogen	Macherey Nagel	Qiagen	Zymo	Invitrogen	Macherey Nagel	Zymo	Invitrogen	Macherey Nagel	Zymo
Bacteroides	7.50% (Lemoore)	NA		7-34% (depending on obesity or leanness)	12-18%	NA	9,094%	8,205%	14,878%	2,364%	2,488%	24,747%	10,287%	8,21%	7,503%	14,152%	0,83%	0,57%	15,85%	1,49%	0,16%	0,26%	1,38%	0,09%	0,06%	0,50%
Alistipes/unclassified Rikenellaceae	7.80%	NA		NA	NA	NA	1,487%	5,401%	8,683%	5,916%	1,793%	0,488%	10,072%	5,307%	3,370%	5,203%	18,316%	10,196%	3,799%	11,770%	8,331%	4,107%	17,612%	8,32%	2,25%	0,30%
Lactobacillus	3.50%	NA		NA	1.6%	NA	0,012%	0,618%	0,412%	0,088%	0,045%	0,003%	0,073%	0,004%	0,003%	0,003%	0,011%	0,017%	0,023%	0,413%	0,000%	0,000%	0,001%	0,00%	0,00%	0,01%
Fusobacterium	1.1%	NA		NA	0.13%	NA	9,840%	9,403%	17,535%	5,905%	1,388%	26,772%	8,421%	8,53%	8,388%	12,061%	1,898%	1,298%	18,981%	1,548%	0,634%	0,309%	3,338%	0,08%	0,08%	0,496%
Peptoclostridium	NA	NA		5-13%	NA	NA	0,000%	0,002%	0,002%	0,005%	0,002%	0,003%	0,023%	0,002%	0,000%	0,001%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,00%	0,00%	0,000%
Prevotella	4.60%	NA		8-13%	8.20%	NA	0,190%	0,181%	0,035%	0,134%	0,064%	1,190%	0,588%	0,04%	0,024%	0,002%	0,000%	0,000%	0,000%	0,000%	0,034%	0,016%	0,018%	0,05%	0,02%	0,024%
Weissella	NA	NA		3-4%	NA	NA	0,000%	0,001%	0,002%	0,001%	0,000%	0,000%	0,000%	0,001%	0,000%	0,000%	0,001%	0,001%	0,001%	0,000%	0,000%	0,000%	0,000%	0,00%	0,00%	0,000%
Romboutsia	3.10%	NA		2%	NA	NA	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,00%	0,00%	0,000%
Blautia	3.30%	NA		0-18%	2-4%	NA	0,564%	4,257%	5,708%	2,810%	2,128%	0,040%	2,503%	8,594%	8,812%	6,466%	35,038%	19,417%	0,933%	18,995%	17,213%	10,364%	38,125%	0,53%	0,71%	0,406%
Clostridium	3.30%	NA		0-34%	3-15%	NA	0,911%	1,051%	1,483%	0,792%	0,298%	1,078%	0,713%	2,417%	2,587%	0,062%	3,210%	3,788%	5,461%	5,183%	2,183%	1,293%	12,371%	3,78%	2,11%	20,188%
Anaerostipes/Lachnospira	1.70%	NA		NA	1-2%	NA	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,00%	0,00%	0,000%
Sutterella	2.50%	NA		1-2%	2-4.5%	NA	0,000%	0,000%	0,000%	0,188%	0,038%	4,190%	0,981%	0,388%	0,368%	1,468%	0,000%	0,000%	0,000%	0,000%	0,000%	0,033%	0,528%	0,04%	0,04%	0,569%
Faecalibacterium	2.10%	NA		1-3%	1-3.4%	NA	2,888%	2,971%	0,438%	1,390%	2,217%	0,461%	7,242%	2,741%	2,088%	2,266%	1,812%	1,928%	10,984%	4,847%	0,340%	0,779%	7,408%	1,63%	0,95%	12,073%
Turicibacter	0.80%	NA		NA	0.4-2%	NA	0,000%	0,000%	0,000%	0,082%	0,151%	0,185%	0,35%	0,074%	0,105%	0,618%	0,000%	0,000%	0,000%	0,000%	0,006%	0,032%	0,048%	0,02%	0,01%	0,085%
Alloprevotella	0.90%	NA		NA	1-5.5%	NA	0,667%	0,172%	0,414%	0,130%	0,061%	1,244%	0,150%	0,098%	0,340%	0,660%	0,274%	0,401%	25,102%	0,539%	0,250%	0,245%	0,854%	0,45%	0,31%	3,401%
Akkermansia	0.70%	NA		0-4%	2-4%	NA	1,508%	3,413%	0,387%	0,771%	0,376%	0,008%	0,107%	1,435%	1,390%	0,227%	1,356%	0,940%	0,041%	0,148%	0,096%	0,043%	0,016%	0,00%	0,02%	0,004%
Phascolarctobacterium	1%	NA		1-2%	1.5-2%	NA	0,000%	0,000%	0,000%	0,001%	0,001%	0,003%	0,001%	0,012%	0,017%	0,000%	0,002%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,00%	0,00%	0,000%
Streptococcus	NA	NA		0.5-53%	NA	NA	0,000%	0,000%	0,000%	0,000%	0,016%	0,013%	0,007%	0,000%	0,007%	0,004%	0,000%	0,000%	0,000%	0,000%	0,008%	0,008%	0,018%	0,01%	0,01%	0,027%

10.7. Supplementary Data 7

Supplementary Data 7. a. DNA purification from MCS. The table shows the yield (quantity measured by Qubit) and the interval of the largest peaks detected by TapeStation.

Samples	Initial volume (µl)	Elution (µl)	gDNA cc. (ng/µl)		Largest peak	
					Highest average (bp)	Largest average (bp)
Qiagen	75	50	1	0,44	26271	>60000
			2	0,292		
			3	0,252		
			4	0,552		
			5	0,286		
Invitrogen	75	50	1	0,286	15170	>60000
			2	0,334		
			3	0,342		
			4	0,222		
			5	0,252		
Macherey-Nagel	75	50	1	9,86	5419	54740
			2	11		
			3	9,92		
			4	10,9		
			5	9,14		
Zymo Research	75	50	1	40,4	9849	>60000
			2	42,6		
			3	34,6		
			4	37		
			5	40		

b. DNA purification from GMS. The table shows the yield (quantity measured by Qubit) and the interval of the largest peaks detected by TapeStation.

Samples	Initial volume (µl)	Elution (µl)	gDNA cc. (ng/µl)		Largest peak	
					Highest average (bp)	Largest average (bp)
Qiagen	75	50	1	1,17	6895	>60000
			2	1,29		
			3	0,978		
Invitrogen	75	50	1	0,06	4746	58897
			2	0,07		
			3	0,075		
Macherey-Nagel	75	50	1	11,8	5340	55538
			2	12,7		
			3	0,038 *		
Zymo Research	75	50	1	4,36	5539	>60000
			2	4,14		
			3	3,76		

10.8. Supplementary Data 8

Supplementary Data 8. Sequencing platform-specific CIGAR scoring schemes for calculating calculating CIGAR scores in minitax's alignement processing.

Platform	Match Score	Mismatch Score	Insertion Score	Deletion Score	Gap Opening Penalty	Gap Extension Penalty	Description
Illumina	2	-4	-3	-3	-4	-2	Optimized for high-accuracy, short reads. Higher penalties for mismatches and indels to reflect the platform's low error rate.
ONT	1	-3	-2	-2	-2	-1	Adjusted for longer reads with higher error rates. More lenient penalties to accommodate frequent indels and mismatches.
PacBio	2	-3	-3	-3	-3	-2	Balanced settings for long, high-fidelity reads (e.g., HiFi mode). Moderate penalties for indels to support accurate alignment in repetitive regions.

10.9. Supplementary Data 9

Supplementary Data 9. PCR amplification conditions for Zymo Research Quick-16S library preparation (V1-V2 and V3-V4 amplicon sequencing) – PCR I

PCR steps	Temperature	Time	Cycle number
Initial denaturation	95°C	10 min	1
Denaturation	95°C	30 sec	20 (dog) or 12 (MCS
Annealing	55°C	30 sec	
Extension	72°C	3 min	

10.10. Supplementary Data 10

Supplementary Data 10. Index barcode (ZA7-ZA5) combinations used for library preparation using the Zymo Research Quick-16S NGS Library Prep Kit.

Abbreviations: I: Invitrogen; MN: Macherey-Nagel; Z: Zymo Research

a, Barcodes used for canine stool samples				
Sample	**Index Barcode ZA7**	**Sequence ZA7**	**Index Barcode ZA5**	**Sequence ZA5**
I1	ZA701	ACCTGGAT	ZA501	TTCTAGAC
I2	ZA702	GTGCCATA	ZA502	CCGATCTT
I3	ZA703	TGAATCCG	ZA503	TAAGATCC
I4	ZA704	CATGATGC	ZA504	AGGTCATT
MN1	ZA705	AATGTCCT	ZA501	TTCTAGAC
MN2	ZA706	ATAGGCTC	ZA502	CCGATCTT
MN3	ZA701	ACCTGGAT	ZA503	TAAGATCC
MN4	ZA702	GTGCCATA	ZA504	AGGTCATT
Z1	ZA703	TGAATCCG	ZA501	TTCTAGAC
Z2	ZA704	CATGATGC	ZA502	CCGATCTT
Z3	ZA705	AATGTCCT	ZA503	TAAGATCC
Z4	ZA706	ATAGGCTC	ZA504	AGGTCATT
b, Barcode sequences applied for MCS samples				
Sample	**Index Barcode ZA7**	**Sequence ZA7**	**Index Barcode ZA5**	**Sequence ZA5**
I1	ZA707	TTGCGGAG	ZA505	GTGTGTCA
I2	ZA705	AATGTCCT	ZA504	AGGTCATT
I3	ZA701	ACCTGGAT	ZA501	TTCTAGAC
Q1	ZA705	AATGTCCT	ZA503	TAAGATCC
Q2	ZA702	GTGCCATA	ZA502	CCGATCTT
Q3	ZA706	ATAGGCTC	ZA504	AGGTCATT
MN1	ZA703	TGAATCCG	ZA501	TTCTAGAC
MN3	ZA708	GCCTTCCA	ZA508	CCACAGGT
MN4	ZA704	CATGATGC	ZA502	CCGATCTT
Z1	ZA706	ATAGGCTC	ZA505	GTGTGTCA
Z2	ZA709	GTCAGTCT	ZA508	CCACAGGT
Z3	ZA705	AATGTCCT	ZA505	GTGTGTCA

10.11. Supplementary Data 11

Supplementary Data 11. PCR amplification conditions for Zymo Research Quick-16S library preparation (V1-V2 and V3-V4 amplicon sequencing) – PCR II (barcoded PCR)

PCR steps	Temperature	Time	Cycle number
Initial denaturation	95°C	10 min	1
Denaturation	95°C	30 sec	5
Annealing	55°C	30 sec	
Extension	72°C	3 min	

10.12. Supplementary Data 12

Supplementary Data 12. Summary table for showing the amount of DNA used for PerkinElmer library preparation, the amount of libraries used for sequencing. The table also shows the barcode ID-s as well as the index sequences.

Sample	Genomic DNA cc (ng/ul)	Library cc (ng/ul)	Barcode	primer index
Q1	22,2	3,24	11	AAGCGTACGTCC
Q2	26,4	4,02	13	TCGGGAAGGTCC
Q3	25,5	8,56	37	TCGGGAAGGTCC
Q4	36	0,16	38	GAGGCATCGGCC
MN1	2,5	29,2	14	TTATCAGTCCTT
MN2	2,2	34,8	15	GTCATCGCGTCC
MN3	2,27	42,8	39	AATAATTGGTCC
MN4	1,6	40,6	40	GTCGTCAACCGG
I1	3,47	34,8	16	CCGTCTCTCCGG
I2	4,1	5,22	18	ACGCTCTTCCGG
I3	2,4	41,8	41	AAATCTCAGGCC
I4	4,46	45,4	42	GTGCGCGGCCGG
Z1	4,46	25,8	13	AAGCGTACGTCC
Z2	3,78	13,8	14	TTATCAGTCCTT
Z3	3,59	34,6	16	CCGTCTCTCCGG
Z4	4,42	4,1	18	ACGCTCTTCCGG

10.13. Supplementary Data 13

Supplementary Data 13. PCR Conditions for PerkinElmer NEXTFLEX® 16S V1-V3 Amplicon-Seq Kit for Illumina – PCR I

PCR steps	Temperature	Time	Cycle number
Initial denaturation	98°C	4 min	1
Denaturation	98°C	30 sec	8
Annealing	60°C	30 sec	
Extension	72°C	30sec	
Final extension	72°C	4 min	1

10.14. Supplementary Data 14

Supplementary Data 14. PCR Conditions for PerkinElmer NEXTFLEX® 16S V1-V3 Amplicon-Seq Kit for Illumina – PCR II

PCR steps	Temperature	Time	Cycle number
Initial denaturation	98°C	4 min	1
Denaturation	98°C	30 sec	varying according to the amount of the initial amount of DNA
Annealing	60°C	30 sec	*
Extension	72°C	30sec	*
Final extension	72°C	4 min	1

10.15. Supplementary Data 15

Supplementary Data 15. Summary table for showing the barcode IDs and sequences for ONT 16S library preparation. IDs used for library preparation

a, from the main subject of the study			
Replicate #1		**Replicate #2**	
Sample	**Barcode**	**Sample**	**Barcode**
I1	bc09	I1	bc09
I2	bc10	I2	bc02
I3	bc11	I3	bc03
I4	bc12	I4	bc04
MN1	bc05	MN1	bc05
MN2	bc06	MN2	bc10
MN3	bc07	MN3	bc11
MN4	bc08	MN4	bc12
Z1	bc01	Z1	bc01
Z2	bc02	Z2	bc06
Z3	bc03	Z3	bc07
Z4	bc04	Z4	bc08
Q1	bc01		
Q2	bc02		
Q3	bc03		
Q4	bc04		

b, from the six additional dogs	
Sample	**Barcode**
Female 1	bc21
Male 1	bc17
Female 2	bc1
Female 3	bc5
Male 2	bc9
Male 3	bc13

c, form MCS	
Sample	**Barcode**
I1	bc7
I2	bc8
I3	bc9
I4	bc6
I5	bc7
MN1	bc4
MN2	bc5
MN3	bc6
MN4	bc3
MN5	bc4
Z1	bc1
Z2	bc2
Z3	bc3
Z4	bc1
Z5	bc2
Q1	bc10
Q2	bc11
Q3	bc12
Q4	bc8
Q5	bc9

d, barcode sequences		
ONT's Barcode ID	**ID used in the study**	**Barcode sequence**
16S01	bc01	AAGAAAGTTGTCGGTGTCTTTGTG
16S02	bc02	TCGATTCCGTTTGTAGTCGTCTGT
16S03	bc03	GAGTCTTGTGTCCCAGTTACCAGG
16S04	bc04	TTCGGATTCTATCGTGTTTCCCTA
16S05	bc05	CTTGTCCAGGGTTTGTGTAACCTT
16S06	bc06	TTCTCGCAAAGGCAGAAAGTAGTC
16S07	bc07	GTGTTACCGTGGGAATGAATCCTT
16S08	bc08	TTCAGGGAACAAACCAAGTTACGT
16S09	bc09	AACTAGGCACAGCGAGTCTTGGTT
16S10	bc10	AAGCGTTGAAACCTTTGTCCTCTC
16S11	bc11	GTTTCATCTATCGGAGGGAATGGA
16S12	bc12	CAGGTAGAAAGAAGCAGAATCGGA
16S13	bc13	AGAACGACTTCCATACTCGTGTGA
16S17	bc17	ACCCTCCAGGAAAGTACCTCTGAT
16S21	bc21	GAGCCTCTCATTGTCCGTTCTCTA

10.16. Supplementary Data 16

Supplementary Data 16. PCR Conditions for preparation of ONT V1-V9 libraries

PCR step	Temperature	Time	No. of cycles
Initial denaturation	95 °C	1 min	1
Denaturation	95 °C	20 secs	25
Annealing	55 °C	30 secs	25
Extension	65 °C	2 mins	25
Final extension	65 °C	5 mins	1
Hold	4 °C	∞	

10.17. Supplementary Data 17

Supplementary Data 17. Summary table of primers used for PacBio 16S library preparation.

	Forward primer	Forward primer sequence	Reverse primer	Reverse primer sequence
I1	16S_Fw_1007	TCTGTATCTCTATGTG	16S_Rev_1056	ATGTGCGTGTGTGTCT
I2	16S_Fw_1008	ACAGTCGAGCGCTGCG	16S_Rev_1056	ATGTGCGTGTGTGTCT
I3	16S_Fw_1012	ACACTAGATCGCGTGT	16S_Rev_1056	ATGTGCGTGTGTGTCT
I4	16S_Fw_1015	CGCATGACACGTGTGT	16S_Rev_1056	ATGTGCGTGTGTGTCT
MN1	16S_Fw_1020	CACGACACGACGATGT	16S_Rev_1056	ATGTGCGTGTGTGTCT
MN2	16S_Fw_1022	CACTCACGTGTGATAT	16S_Rev_1056	ATGTGCGTGTGTGTCT
MN3	16S_Fw_1024	CATGTAGAGCAGAGAG	16S_Rev_1056	ATGTGCGTGTGTGTCT
MN4	16S_Fw_1005	CACTCGACTCTCGCGT	16S_Rev_1057	CTCTCAGACGCTCGTC
Z1	16S_Fw_1007	TCTGTATCTCTATGTG	16S_Rev_1057	CTCTCAGACGCTCGTC
Z2	16S_Fw_1008	ACAGTCGAGCGCTGCG	16S_Rev_1057	CTCTCAGACGCTCGTC
Z3	16S_Fw_1012	ACACTAGATCGCGTGT	16S_Rev_1057	CTCTCAGACGCTCGTC
Z4	16S_Fw_1015	CGCATGACACGTGTGT	16S_Rev_1057	CTCTCAGACGCTCGTC

10.18. Supplementary Data 18

Supplementary Data 18. PCR conditions for preparation of PacBio V1-V9 libraries

PCR steps	Temperature	Time	Cycle number
Initial denaturation	95°C	3 min	1
Denaturation	95°C	30 sec	25
Annealing	57°C	30 sec	
Extension	72°C	60sec	

10.19. Supplementary Data 19

Supplementary Data 19. Summary table of primers from Illumina DNA Prep Kit used for WGS library preparation.

Sample	Index Barcode i7	Barcode Sequence i7	Index Barcode i5	Barcode Sequence i5
Q1	H705	GGACTCCT	H505	GTAAGGAG
Q2	H705	GGACTCCT	H503	TATCCTCT
Q3	H706	TAGGCATG	H505	GTAAGGAG
Q4	H714	GCTCATGA	H517	GCGTAAGA
I1	H711	AAGAGGCA	H503	TATCCTCT
I2	H714	GCTCATGA	H505	GTAAGGAG
I3	H714	GCTCATGA	H506	ACTGCATA
I4	H711	AAGAGGCA	H505	GTAAGGAG
MN1	H707	CTCTCTAC	H506	ACTGCATA
MN2	H710	CGAGGCTG	H517	GCGTAAGA
MN3	H706	TAGGCATG	H506	ACTGCATA
MN4	H705	GGACTCCT	H517	GCGTAAGA
Z1	H711	AAGAGGCA	H506	ACTGCATA
Z2	H714	GCTCATGA	H503	TATCCTCT
Z3	H706	TAGGCATG	H517	GCGTAAGA
Z4	H707	CTCTCTAC	H503	TATCCTCT

10.20. Supplementary Data 20

Supplementary Data 20. PCR Conditions for preparation of Illumina WGS libraries

PCR steps	Temperature	Time	Cycle number
Pre-heating	68°C	3 min	1
Initial denaturation	98°C	3 min	1
Denaturation	98°C	45 sec	5
Annealing	62°C	30 sec	
Extension	68°C	2 min	
Final extension	68°C	1 min	1

10.21. Supplementary Data 21

Supplementary Data 21.

Genus	EPI2ME value	NCBI value
Clostridium	0.0137855341364841	0.0135811686783089
Kineothrix	0.00103065605367019	N/A
Collinsella	0.0341797461460624	0.0359682063346239
Terrisporobacter	0.00232185350281726	N/A
Paeniclostridium	0.00827780884220112	N/A
Romboutsia	0.00209802830920327	N/A
Blautia	0.921588848109184	1.0123767434625
Enterocloster	0.00636597606432633	0.0260434289128637
Campylobacter	0.00104069475650225	N/A
Erysipelatoclostridium	0.00866333190917401	N/A
Faecalicatena	0.0128830781652849	0.0202894786939917
Mediterraneibacter	0.00822035492654341	0.438629625707591
Phocaeicola	0.0169983976390598	0.0183188779189124
Megamonas	0.00813308189787587	0.0088669435189919
Roseburia	0.00140255975128613	0.00950004122302111
Cellulosilyticum	0.00263681138534562	N/A
Bacteroides	0.00183680964869668	N/A
Lacrimispora	0.0167242255407207	0.0271792333362727
Fusobacterium	0.0399184899712927	0.0409118529388183
Peptacetobacter	0.0045492181048344	1.73183878618779
Catenibacterium	N/A	0.0462073780062417
Eisenbergiella	N/A	0.0336467399273302
Faecalimonas	N/A	0.296900378156428
Lachnoclostridium	N/A	0.0100645523848315
Intestinibacter	N/A	0.0140135036154961
Tyzzerella	N/A	0.0400242743675637
Coprococcus	N/A	0.0135261646295627