Supplementary methods

S1. DNA extraction Procedures

S1.1 Pre-extraction PBS wash and bead-beating

Before kit-specific protocols, approximately 7.5 mL fecal slurry was vortexed for 5 s and transferred to a sterile 50 mL tube, followed by addition of 7.5 mL sterile 1× PBS (pH 7.4). Six sterile 4 mm steel beads were added, and tubes were vortexed at 1800 rpm for 10 min using a MULTI-TX5 Digital Multi-tube Vortex to improve homogenization and mechanical lysis.

S1.2 QIAamp Fast DNA Stool Mini Kit (Fast Stool)

One milliliter of homogenized slurry was mixed with 1 mL InhibitEX Buffer, incubated at 95 °C for 5 min, and processed following the manufacturer's protocol with noted modifications, including overnight incubation with Proteinase K and Buffer AL. DNA was eluted in 200 µL Buffer ATE. One extract (sample ID 316, horse H2) was lost during processing

S1.3 DNeasy PowerFecal Pro kit (PowerFecal Pro)

One milliliter slurry was added to a PowerBead Pro Tube with 800 µL Solution CD1 and vortexed at 1800 rpm for 10 min, followed by overnight lysis at room temperature. Subsequent steps followed the manufacturer's protocol, including two centrifugation steps, addition of CD2 and CD3, column purification, and elution in 100 µL 10 mM Tris (pH 8.0). Extracts were stored at −20 °C until sequencing.

S1.4 Controls and collection samples

Three extraction replicates were produced for each PBS-washed slurry for both kits. Collection samples were created by pooling 2.5 mL slurry from each of the three biological replicates per horse. Negative controls consisted of 7.5 mL ethanol and were included approximately every 12 samples per extraction kit. Positive controls consisted of 250 µL undiluted ZymoBIOMICS Microbial Community Standard (Zymo Research, Cat. No. D6300, from Nordic Biosite, product code Biosite-D6300/ZymoBIOMICS-Microbial-Community-Standard).

S2. 16S rRNA Gene Amplification and Sequencing

S2.1 Primers and PCR

Amplicons targeted the V4 region using 515F (Parada-modified: GTGYCAGCMGCCGCGGTAA [1] and 806R (Apprill-modified: GGACTACNVGGGTWTCTAAT) [2]. These primers are modified versions of the original Caporaso et al. V4 primers [3], later refined by Apprill et al [2] and Parada et al [1–3]. They produce an expected amplicon size of 250-300 base pairs. PCR was conducted by combining the following components in a total volume of 25 µL: 5 µL of 2x Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Cat. No. F631S), 5 µL of GC enhancer (Thermo Fisher Scientific, Cat. No. F631S), 0.5 µL of reverse primer (10 µM), 1 µL of forward primer (10 µM), 5 µL of nuclease-free water, and 1 µL of DNA. The forward primers were tagged with barcodes numbered 101-196 according to the IonCode system, while the reverse primers remained untagged. The PCR program included an initial denaturation step at 98 °C for 30 seconds, followed by 35 cycles consisting of denaturation at 98 °C for 10 seconds, annealing at 50 °C for 10 seconds, and extension at 72 °C for 15 seconds, concluded by a final extension at 72 °C for 5 minutes.

Ten samples from each PCR tray were run on 2% agarose gels to verify the presence of the desired amplicon size, using the E-Gel PowerSnap system (Thermo Fisher Scientific, Waltham, MA, USA) and a 1 kb E-Gel ladder (Thermo Fisher Scientific, Waltham, MA, USA).

Following PCR, the libraries were normalized using the SequalPrep Normalization Kit (Thermo Fisher Scientific, Cat. No. A1051001) according to the manufacturer's instructions except eluting in 15 µl.

S2. 2 Ion Torrent workflow

Libraries followed the vendor's amplicon workflow using the Ion Plus Fragment Library Kit (Thermo Fisher Scientific, Cat. No. 4471252) and Ion Xpress™ Barcode Adapters 1-16 Kit (Thermo Fisher Scientific, Cat. No. 4471250), according to the manufacturer's protocol for barcoded amplicon libraries (400 bp). After PCR, amplicon quality was verified on 2% E-Gel agarose (E-Gel PowerSnap system) to confirm the expected V4 fragment size (~350 bp). Library concentrations were measured using Qubit dsDNA HS fluorometry, and plate-wise pools were prepared accordingly. Final library quantification was performed by TaqMan™ qPCR to determine molarity, and 100 pM dilutions were prepared for template loading based on qPCR results. Template preparation was performed on the Ion Chef™ system, following the manufacturer's amplicon workflow. Sequencing was performed on an Ion S5 system using the Ion 530™ Chip Kit, following the manufacturer's protocol for barcoded amplicon libraries (400 bp).

S2.3 Illumina workflow

Normalized samples were sequenced at the Norwegian Sequencing Centre (Oslo, Norway) using Illumina technology. Six samples were either manually diluted due to limited capacity on the normalization plate or left undiluted as their concentrations were already within the desired range (1–2 ng/µL). Libraries were quantified using a Qubit fluorometer and purified with 1.2× Ampure XP beads, then eluted in 25 µL of nuclease-free water. For each library, 1 µL from every sample on the normalization plate was pooled. The resulting libraries were subjected to a second purification with 1.2× Ampure XP beads and eluted in 25 µL of water, followed by Qubit quantification. Amplicon size was verified by agarose gel electrophoresis. Libraries were then submitted to the Norwegian Sequencing Centre for Illumina NovaSeq 6000 sequencing (2 × 250 bp paired-end).

Kommentert [AK1]: Jeg sjekket serienummeret på kjøringen, og lengden på reads for å være sikker. Alt peker mot NovaSeq

S3. Bioinformatics and statistical analysis

S3.1 Demultiplexing (cutadapt)

Sequencing reads from Ion Torrent and Illumina were demultiplexed and adapter/primer sequences removed with Cutadapt v4.4 [4]. Demultiplexed FASTQ files were then processed separately with the DADA22 v 1.26.0 [5] in R v4.2.2.

S3.2 Dada22 for Ion Torrent

Ion Torrent reads were filtered with filterAndTrim (minLen = 150, maxLen = 750, maxEE = 4, minQ = 3, rm.phix = FALSE, compress = FALSE, multithread = CPUS). Error models were learned with learnErrors (nbases = 1e8, multithread = TRUE, randomize = TRUE, trimLeft = 15), and dereplicated reads were denoised with dada2 using HOMOPOLYMER_GAP_PENALTY = –1 and BAND_SIZE = 32 to address homopolymer-associated indel errors. Chimeras were removed with removeBimeraDenovo (method = “consensus”, multithread = TRUE).

S3.3 Dada22 for Illumina

For Illumina paired-end reads, quality profiles were inspected and reads were filtered with filterAndTrim (maxN = 0, maxEE = c(2,2), truncQ = 2, minLen = 50, rm.phix = TRUE, compress = TRUE, multithread = TRUE). Forward and reverse error models were estimated with learnErrors using a custom loess-based fitting function (loessErrfun), which fits loess on log10-transformed substitution error rates versus quality score (weights = log10(total observations), degree = 1, span = 0.95), enforces bounds (1e-7 to 0.25) and monotonicity, and reconstructs the full substitution matrix (see https://github.com/benjjneb/dada22/issues/1307). Dereplicated reads were denoised with dada2 in pseudo-pooling mode (pool = “pseudo”, multithread = TRUE). Paired reads were merged with mergePairs (trimOverhang = TRUE), sequence tables were built with makeSequenceTable, and chimeras were removed with removeBimeraDenovo (method = “consensus”, multithread = TRUE).

S3.4 Merging dataset and assigning taxonomy

Per-sample ASV tables retaining size information were exported from both datasets. Identical sequences were clustered across datasets with VSEARCH cluster_size (--id 1.00) v 2.25.0; [6] and chimeras were further screened with VSEARCH chimera_denovo. Per-sample reads were mapped back to centroid sequences using the retained size data to build a combined OTU/centroid abundance matrix. Taxonomy was assigned with assignTaxonomy using the SILVA reference set silva_nr99_v138.2 [7]. The OTU table and taxonomy were merged with metadata in Phyloseq [8].

Kommentert [AK2]: Her er det formelt riktig å bruke ASV siden det er et ubesudlet produkt fra DADA2.

S3.5 Statistical modelling in R

All analyses were conducted in R (v4.2.2) using phyloseq [8] for data handling and ordination, vegan [9] for multivariate tests, ggplot2[10] for visualization, and dplyr [11], tidyr [12] and broom [13] for data wrangling and summaries. Bray–Curtis dissimilarities were computed from relative abundances; ordination used PCoA via phyloseq::ordinate(). Bray–Curtis distances were computed from the phyloseq object (physeq_method_illumina) and tested by PERMANOVA using adonis2 (vegan) with 999 permutations and marginal effects.

Two PERMANOVA designs were tailored to match the dataset structure. Model 1 (platform effect) included Library 1 samples sequenced on both Illumina and Ion Torrent; adonis2 was run with 9,999 permutations and permutations constrained within horse (strata = horse_name). Model 2 (storage effect) included Illumina samples across libraries with storage levels (Immediate, 4 °C, RT); library_tag was included as a fixed effect, and permutations were constrained within library (strata = library_tag). Homogeneity of multivariate dispersion was assessed with betadisper.

Phylum-level differences between platforms were tested using paired Wilcoxon tests on matched samples (paired by sample_number2), with Benjamini–Hochberg FDR correction across phyla and reporting median differences (Illumina − Ion Torrent). Differential abundance was assessed with DESeq2 (design ~ horse_name + extraction_kit), and adjusted p-values (padj) were reported using Benjamini–Hochberg correction. As a sensitivity analysis, samples were rarefied to 10,000 reads (phyloseq::rarefy_even_depth), followed by re-running ordination and PERMANOVA. Visualizations used the Okabe–Ito palette, shape coding by horse identity, and 68% confidence ellipses to illustrate group dispersion without implying statistical significance.

References

1. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology. 2016;18:1403–14. https://doi.org/10.1111/1462-2920.13023.

2. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol. 2015;75:129–37. https://doi.org/10.3354/ame01753.

3. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA. 2011;108 supplement_1:4516–22. https://doi.org/10.1073/pnas.1000080107.

4. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j. 2011;17:10. https://doi.org/10.14806/ej.17.1.200.

5. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3. https://doi.org/10.1038/nmeth.3869.

6. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:e2584. https://doi.org/10.7717/peerj.2584.

7. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research. 2012;41:D590–6. https://doi.org/10.1093/nar/gks1219.

8. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE. 2013;8:e61217. https://doi.org/10.1371/journal.pone.0061217.

9. Dixon P. VEGAN, a package of R functions for community ecology. J Vegetation Science. 2003;14:927–30. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

10. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer New York; 2009. https://doi.org/10.1007/978-0-387-98141-3.

11. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: A Grammar of Data Manipulation. 2014;:1.2.0. https://doi.org/10.32614/CRAN.package.dplyr.

12. Wickham H, Vaughan D, Girlich M. tidyr: Tidy Messy Data. 2014;:1.3.2. https://doi.org/10.32614/CRAN.package.tidyr.

13. Robinson D, Hayes A, Couch S, Hvitfeldt E. broom: Convert Statistical Objects into Tidy Tibbles. 2014;:1.0.12. https://doi.org/10.32614/CRAN.package.broom.