



Quantifying variation across 16S rRNA gene sequencing runs in human microbiome studies

Andrew J. Hoisington^{1,2,3,4} · Christopher E. Stamper^{1,2,3} · Joseph C. Ellis⁵ · Christopher A. Lowry^{1,2,3,6} · Lisa A. Brenner^{1,2,3,7,8}

Received: 21 February 2024 / Revised: 15 May 2024 / Accepted: 21 May 2024

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024

Abstract

Recent microbiome research has incorporated a higher number of samples through more participants in a study, longitudinal studies, and metanalysis between studies. Physical limitations in a sequencing machine can result in samples spread across sequencing runs. Here we present the results of sequencing nearly 1000 16S rRNA gene sequences in fecal (stabilized and swab) and oral (swab) samples from multiple human microbiome studies and positive controls that were conducted with identical standard operating procedures. Sequencing was performed in the same center across 18 different runs. The simplified mock community showed limitations in accuracy, while precision (e.g., technical variation) was robust for the mock community and actual human positive control samples. Technical variation was the lowest for stabilized fecal samples, followed by fecal swab samples, and then oral swab samples. The order of technical variation stability was inverse of DNA concentrations (e.g., highest in stabilized fecal samples), highlighting the importance of DNA concentration in reproducibility and urging caution when analyzing low biomass samples. Coefficients of variation at the genus level also followed the same trend for lower variation with higher DNA concentrations. Technical variation across both sample types and the two human sampling locations was significantly less than the observed biological variation. Overall, this research providing comparisons between technical and biological variation, highlights the importance of using positive controls, and provides semi-quantified data to better understand variation introduced by sequencing runs.

Key points

- *Mock community and positive control accuracy were lower than precision.*
- *Samples with lower DNA concentration had increased technical variation across sequencing runs.*
- *Biological variation was significantly higher than technical variation due to sequencing runs.*

Keywords Microbiome · Sequencing variance · Accuracy · Precision · Positive controls

✉ Andrew J. Hoisington
andrew.hoisington@va.gov

¹ Veterans Health Administration, Rocky Mountain Mental Illness Research Education and Clinical Center (MIRECC) for Veteran Suicide Prevention, Rocky Mountain Regional Veterans Affairs Medical Center (RMRVAMC), Aurora, CO, USA

² Department of Physical Medicine and Rehabilitation, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

³ Military and Veteran Microbiome: Consortium for Research and Education (MVM-CoRE), Aurora, CO, USA

⁴ Department of Systems Engineering and Management, Air Force Institute of Technology, Wright-Patterson Air Force Base, Dayton, OH, USA

⁵ Netellis LLC, Knoxville, TN, USA

⁶ Department of Integrative Physiology, Center for Neuroscience, and Center for Microbial Exploration, University of Colorado Boulder, Boulder, CO, USA

⁷ Department of Psychiatry, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

⁸ Department of Neurology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Introduction

Microbiome research has grown exponentially as technical advances in sequencing and novel findings linking microbiome diversity and community structure to physiology and behavior encourage exploration in the field. Partially due to the rapid growth, there are challenges comparing results across studies due to differences in study design, cohorts of interest, sequencing technology, and/or sampling and analysis procedures. Initial standardization of sequencing methods emerged from standard operating procedures developed for major studies (e.g., Human Microbiome Project (Turnbaugh et al. 2007), Earth Microbiome Project (Gilbert et al. 2014), American Gut Project (McDonald et al. 2018). More recently, concentrated efforts of multiple research consortia have joined together (e.g., Microbiome Quality Control Project (Sinha et al. 2015), International Human Microbiome Standards (Cardona et al. 2012; Santiago et al. 2014), with government agencies (e.g., National Institute for Biological Standards and Controls, National Institute of Standards and Technology) to propose standards for conducting microbiome research. Moreover, the reporting of methods and data in microbiome studies has been aided by the introduction of the Strengthening the Organization and Reporting of Microbiome Studies (STORM) checklist (Mirzayi et al. 2021).

Researchers using the 16S rRNA gene for identification of bacteria in human samples have several key decision points that can impact findings. Pre-sequencing decisions include sampling methods (Sinha et al. 2016; Vogtmann et al. 2017), sample storage (Cardona et al. 2012), extraction kit (Kennedy et al. 2014), and primer selection (Abellan-Schneyder et al. 2021). In addition, over a dozen sequencing platforms and bioinformatics pipelines for analysis of the gut microbiome composition when using 16S rRNA amplicon sequencing are available, which introduces other biases to findings (Allali et al. 2017). Even sequencing platforms developed by the same company (e.g., Illumina MiSeq and iSeq, San Diego, CA, USA) have non-uniform sequencing outputs (Salamon et al. 2022).

One aspect of variability in microbiome sequencing that is less studied is technical reproducibility when using the same sequencing machine for multiple sequencing runs. A physical limitation of sequencing machines is the number of samples in each run to maintain adequate sequencing depth per sample. Studies that require multiple sequencing runs are common due to decreases in sampling costs, increases in desired samples per study, and the use of longitudinal studies. The aim of this study was to investigate the influence of sequencing runs on accuracy and precision,

explore the use of positive controls, and increase understanding regarding the difference between technical (i.e., variation due to processes) and biological variation (i.e., variation due to different participants or same participant at a different time). To achieve these aims, we analyzed 995 16S rRNA gene sample sequencing results from multiple studies and positive controls with the same standard operating procedure.

Materials and methods

The Military and Veterans Microbiome Consortium for Research and Education (MVM-CoRE, <https://www.mirecc.va.gov/visn19/mvm/>) housed within the Rocky Mountain Mental Illness Research Education and Clinical Center (MIRECC) for Veteran Suicide Prevention has been studying the gut, oral, and skin microbiome using 16S rRNA gene sequencing from 2016 to the present using Illumina MiSeq machines. Sequencing runs for this manuscript were from a longitudinal study of United States Veterans (study: US-VMP) and a longitudinal study of non-Veterans seeking Emergency Department (ED) care for a recent mild traumatic brain injury (TBI) (study: ED-TBI). Longitudinal studies enabled comparisons between samples from the same and other participants. Microbiome sample collection was the same procedure as outlined in the US-VMP study (Brenner et al. 2018), with the addition of the OmniGene Gut kit (Cat. No. OMR-200, DNA Genotek, Ottawa, Canada). Briefly, oral microbiome samples were self-collected with double-tipped polyurethane swabs (BD BBL™ CultureSwab™ EZ II, Cat. No. B220144, Fisher Scientific, Pittsburgh, PA, USA) from the buccal mucosa. Participants provided two fecal microbiome samples from the same bowel movement. One fecal microbiome sample was collected with a sterile dual tipped swab using the “first wipe” method and another sample used the OmniGene Gut kit. Fecal samples were collected in-person during a study visit and immediately frozen or at the participant’s residence and shipped to the Rocky Mountain MIRECC via standard ground shipping. Positive controls were the same DNA extracted from a pooled sample of two individuals using the same sample collection procedures as outlined for the US-VMP and the ED-TBI studies. Specifically, we collected fecal swab (“Positive Control Fecal Swab”), OmniGene (“Positive Control Fecal Omni”), and oral swab (“Positive Control Oral Swab”) from the same two participants and the same bowel movements for fecal samples. DNA was also extracted from a mock community microbial standards commercial kit (ZymoBiomics Microbial Community Standard, Cat. No. D6300, Zymo Research, Irvine, CA, USA). Identical procedures were followed for the

sample collection and DNA extraction between the studies, positive controls, and the mock community.

Sample DNA was extracted from microbiome samples using the PowerSoil DNA extraction kit (Cat. No. 12955-4, Qiagen, Valencia, CA, USA) and quantified via Quant-IT dsDNA Assay Kit in triplicate (Cat. No. Q33120, Invitrogen, Waltham, MA, USA). DNA was extracted with 100 μ L of C6 in the final step of the PowerSoil kit. Mock and positive controls were vortexed and pipette mixed prior to being aliquoted to ensure samples each had one freeze-thaw cycle. Marker genes in isolated DNA were polymerase chain reaction (PCR)-amplified using GoTaq Master Mix (Cat. No. M5133, Promega, Madison, WI, USA) and 515 F (5'-GTGCCAGCMGCCGCGGTAA-3'), 806 R (5'-GGA CTACHVGGGTWCTAAT-3') primer pair (Integrated DNA Technologies, Coralville, IA, USA) targeting the V4 hypervariable region of the 16S rRNA gene modified with a unique 12-base sequence identifier for each sample and the Illumina adapter (Caporaso et al. 2012). The thermal cycling program consisted of an initial step at 94 °C for 3 min followed by 35 cycles (94 °C for 45 s, 55 °C for 1 min, and 72 °C for 1.5 min), and a final extension at 72 °C for 10 min. Products from the duplicate PCR reactions were pooled and successful amplification was visualized on an agarose gel. PCR products were cleaned, normalized, and sequenced at a university sequencing center on an Illumina MiSeq using V2 chemistry and 300 cycle, 2 \times 150-bp paired end sequencing. All sequencing was conducted between October 2022 and August 2023. The sequencing center was not involved in the study design or manuscript preparation. Demultiplexed single-end sequences were deposited in the NCBI Sequence Read Archive (BioProject accession ID: PRJNA1101562).

Sequencing data were initially processed using the Quantitative Insights Into Microbial Ecology program (QIIME2 v. 2023.5) (Bolyen et al. 2019). The Deblur algorithm (Amir et al. 2017) was used to denoise demultiplexed sequences. Quality-filtered sequences were assigned taxonomic classification based on the Silva database (v. 138) (Quast et al. 2012). Mock community and positive controls samples were rarefied to 8,600 sequences per sample and participant samples were rarefied at 11,000 sequences per sample.

Statistical analyses were performed with QIIME2 and the open-source statistical software R v. 4.2.2 (The R Core Team 2020) (<https://www.R-project.org>). All statistical tests were conducted with a two-tailed alpha level of 0.05. The alpha diversity metrics assessed were Observed Amplicon Sequencing Variants (ASVs), Shannon Diversity Index, and Pielou's Evenness. Beta diversity was performed using the vegan package (Oksanen et al. 2008) for unweighted UniFrac and weighted UniFrac (Lozupone et al. 2006). Statistical differences for sequencing runs were calculated through pairwise permutational multivariate analysis of variance (PERMANOVA) with 10,000 permutations with the

“adonis2” function. Microbial measures of taxonomic relative abundance were aggregated at the genus level. Intraclass correlation coefficient (ICC) is a measure of reliability or reproducibility that can be used to quantify the biological variability. To calculate ICC, genus or alpha diversity indices were first normalized with the bestNormalize function (Peterson 2021). Repeatability estimation of ICC was with a generalized linear mixed-effect model fitted by restricting maximum likelihood in a Gaussian datatype with 1000 bootstraps and permutations (Stoffel et al. 2017). ICC values range from 0 (i.e., no stability) to 1 (i.e., perfect stability). Values of ICC above 0.5 were considered high microbiome stability (Bobak et al. 2018). ICC was only calculated for mock community based on the number of repeated samples within the runs. Stability was also evaluated through assessment of percent coefficient of variation (%CV) for alpha diversity and for genera that had a mean relative abundance over 1% (i.e., “most abundant genera”).

Results

Accuracy and precision in simplified (mock) microbial communities

The mock community with a DNA concentration of 11.5 ng/ μ L was sequenced 31 times in 18 sequencing runs. Eight ASVs, corresponding with the expected number of mock community taxa, were observed across all samples, representing 97.1% relative abundance (i.e., Total False Positive Relative Abundance (Amos et al. 2020) = 2.9%). The mock community was expected to have evenly distributed taxa with 12.5% relative abundance per taxa. We observed an increased relative abundance of *Escherichia-Shigella* (mean 20.5% \pm standard deviation 4.4%), *Enterobacteriaceae* (18.2% \pm 2.8%), and *Staphylococcus* (16.4% \pm 2.9%), while underrepresented taxa included *Pseudomonas* (5.4% \pm 1.6%) and *Lactobacillus* (1.7% \pm 0.4%) (Fig. 1A, Supplemental Table S1). Despite the variance in relative abundance to theoretical values, all eight taxa had ICC values in the high stability range (Fig. 1B). Observed ASVs (58.9 \pm 23.9) were higher in the mock community than the expected number of eight, yet lower than the positive controls revealing this sample type is less diverse (Supplemental Fig. S1).

Precision in complex (positive control) microbial communities

The Positive Control Fecal Omni sample yielded a DNA concentration of 45.7 ng/ μ L and was sequenced seven times across six runs. The most prevalent genera were *Bacteroides* (24.5% \pm 5.9%), *Blautia* (9.3% \pm 0.6%), *Faecalibacterium* (7.3% \pm 2.2%), and *Prevotella* (6.1% \pm 5.0%) (Fig. 2A).

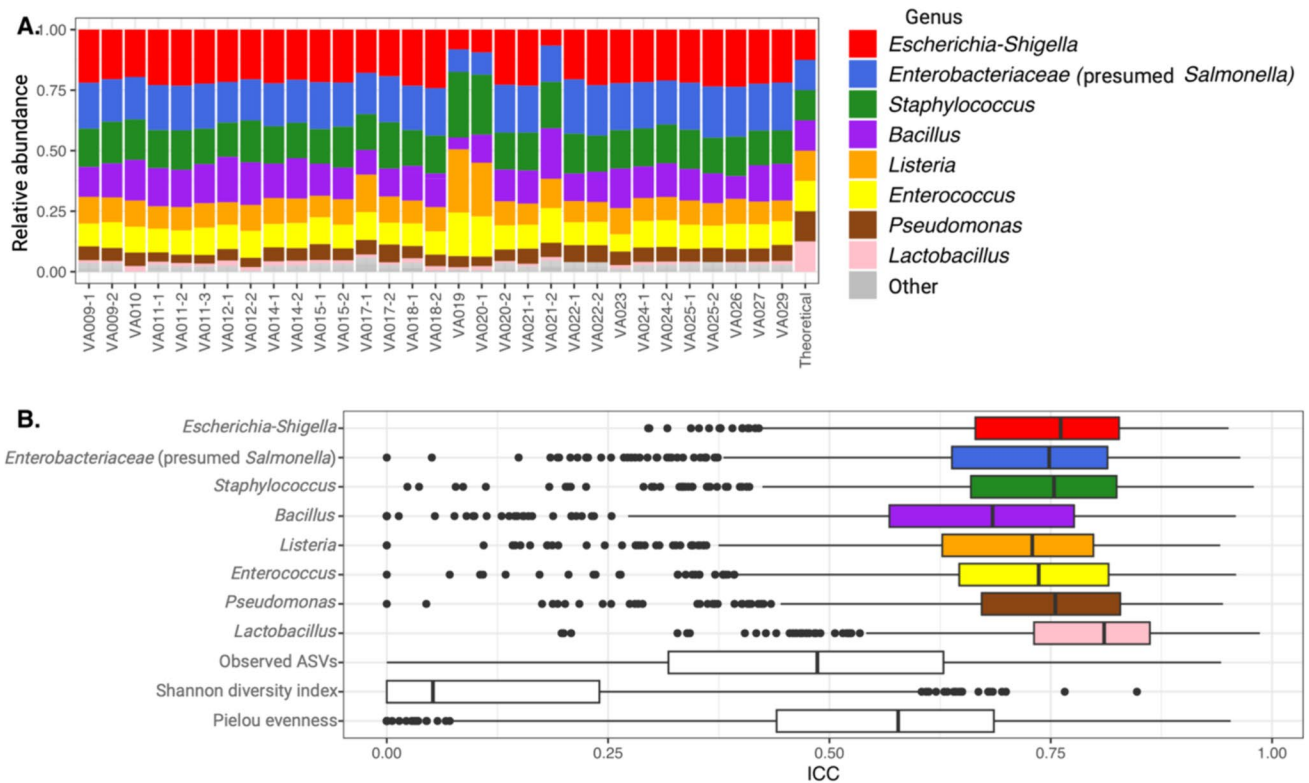


Fig. 1 Mock community relative abundance (%) for all samples and theoretical values based on even distribution of the eight genera (run is denoted by VA, data with multiple samples in the same run are denoted by a dash) (**A**) and intraclass correlation coefficients for

eight genera and three diversity indexes (coefficient of variations for most abundant taxa and alpha diversity measures in Supplemental Table S1) (**B**)

The mean measured alpha diversity values in the Positive Control Fecal Omni samples were 221 ± 22.0 for Observed ASVs, 4.2 ± 0.1 for Shannon diversity index, and 0.78 ± 0.02 for Pielou's evenness. The mean %CV for the most abundant genera (1% or higher, $n = 20$) was 40.0% (range 6.1–81.1%), significantly lower compared to the ED-TBI Omni participant samples (t -test, $p < 0.001$; mean 141.8%, range 59.3–257.8%) (Supplemental Table S2).

Seven runs were conducted with eleven samples from the Positive Control Fecal Swab sample at a DNA concentration of $35.5 \text{ ng}/\mu\text{L}$. The most abundant genera observed were *Bacteroides* ($12.8\% \pm 8.6\%$), *Blautia* ($12.5\% \pm 3.7\%$), *Faecalibacterium* ($10.0\% \pm 2.1\%$), and *Agathobacter* ($5.9\% \pm 2.4\%$) (Fig. 2B). The samples had a mean of 219 observed ASVs per sample (± 25.9), Shannon diversity index of $4.22 (\pm 0.06)$, and Pielou's evenness of $0.79 (\pm 0.03)$. The mean %CV for the most abundant genera ($n = 22$) was 36.0% (range 13.6–68.5%), significantly lower than the US-VMP Fecal Swab participant samples (t -test, $p < 0.001$; mean 215.8%, range 98.8–523.2%) (Supplemental Table S3).

Eleven Positive Control Oral Swab samples at a DNA concentration of $8.7 \text{ ng}/\mu\text{L}$ were sequenced in five runs. The most abundant genera observed in the Positive Control Oral Swab

were *Streptococcus* ($38.4\% \pm 2.0\%$), *Haemophilus* ($24.2\% \pm 1.8\%$), and *Gemella* ($8.7\% \pm 0.1\%$) (Fig. 2C). Alpha diversity values—all of which were lower than Positive Control Fecal Swab and Positive Control Fecal Omni samples—were measured for Shannon diversity index (2.2 ± 0.1), Observed ASV (60.6 ± 13.0), and Pielou's evenness (0.5 ± 0.03). The mean %CV for the most abundant genera ($n = 9$) was 11.0% (range 5.3–17.2%), significantly lower than ED-TBI Oral Swab participant samples (t -test, $p = 0.023$; mean 218.6%, range 42.1–787.9%) (Supplemental Table S4).

All positive controls were analyzed to compare alpha and beta diversity trends. Beta diversity among the positive control types was significantly different using either Weighted UniFrac (PERMANOVA, $p = 0.001$) or Unweighted UniFrac (PERMANOVA, $p = 0.001$) (Fig. S2). All pairwise comparisons were significantly different for Weighted and Unweighted UniFrac (i.e., $p < 0.05$), including Positive Control Fecal Omni and Positive Control Fecal Swab samples. Three calculated alpha diversity metrics were significantly different among the positive control sample types (Kruskal–Wallis Rank Sum Test, $p < 0.001$) (Supplemental Fig. S1). Alpha diversities were not different between Positive Control Fecal Omni and Positive Control Fecal Swab samples (t -test: Observed ASVs,

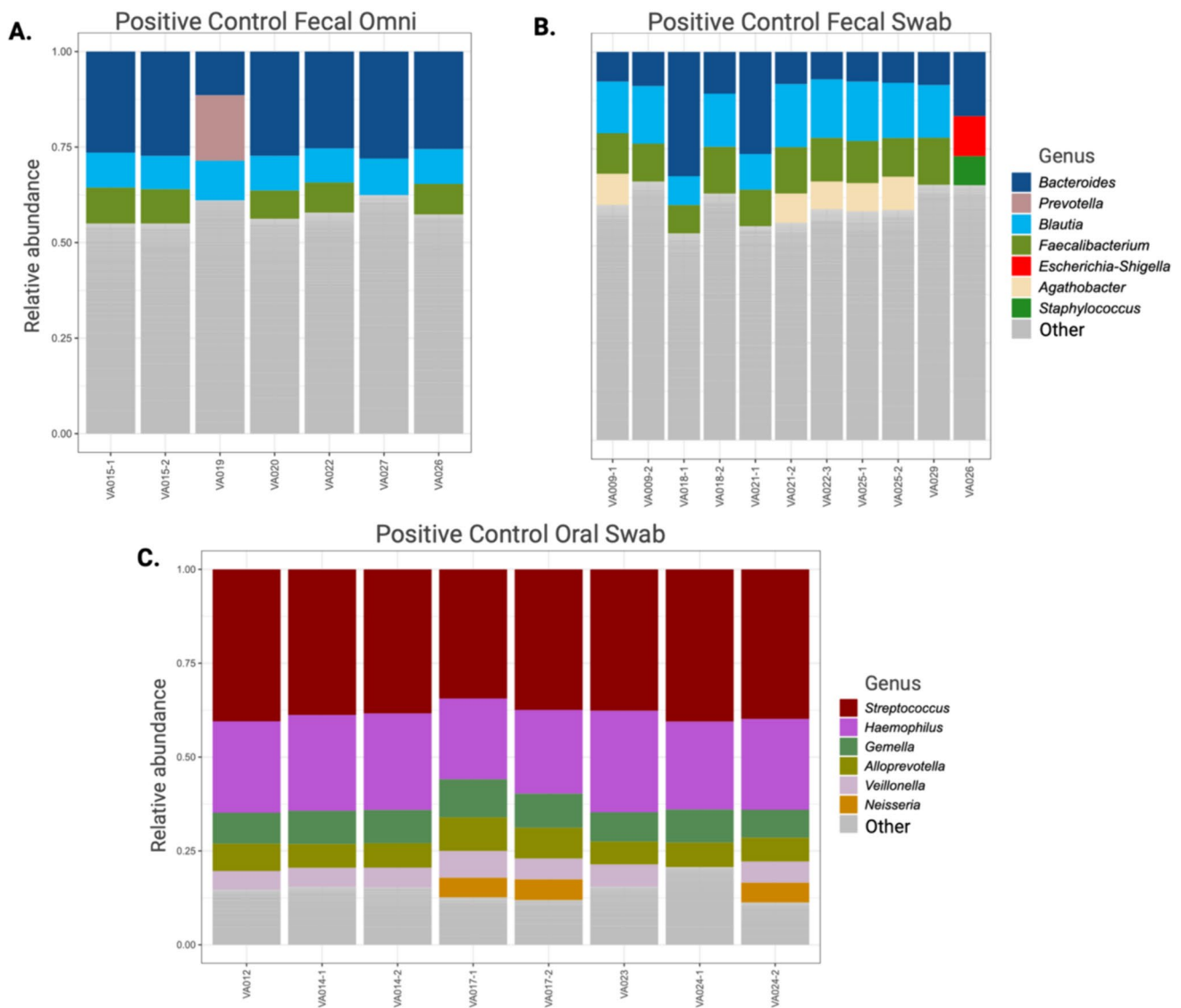


Fig. 2 Genera with the highest relative abundances for positive control samples from Positive Control Fecal Omni (A); Positive Control Fecal Swab (B); Positive Control Oral Swab (C) run denoted by VA

and data with multiple samples in the same run are denoted by a dash; additional details on taxonomic abundance provided in Supplemental Tables S2–4)

$p=0.75$, Shannon Diversity Index $p=0.26$, Pielou’s Evenness, $p=0.22$).

Precision in participant (longitudinal) microbial communities

Two longitudinal microbiome studies enabled comparison across sequencing runs with repeated samples from the same participants. The fecal omni ED-TBI participant samples included 154 paired samples (i.e., exact same DNA sequenced twice) across six sequencing runs. The fecal swab US-VMP participant samples included 63 paired samples across five sequencing runs. Finally, the Oral Swab participant samples included 252 paired samples across four

sequencing runs. Extracted DNA concentration was highest for fecal Omni samples (24.9 ± 19.1 ng/ μ L), followed by US-VMP Fecal Swab samples (12.0 ± 15.0 ng/ μ L) and Oral Swab samples (4.7 ± 6.8 ng/ μ L) (Supplemental Fig. S3). For all three sample types, participant microbial communities were significantly different based on Weighted UniFrac (PERMANOVA, $p < 0.001$). The microbial communities in Oral Swab samples were also significantly different based on run (PERMANOVA, $p = 0.008$), while ED-TBI Fecal Omni and US-VMP Fecal Swab communities were not significantly different (PERMANOVA, $p = 0.36$ and $p = 0.34$). Paired samples shared the most ASVs in the ED-TBI Fecal Omni samples, then US-VMP Fecal Swab samples, and finally ED-TBI Oral Swab samples (Fig. 3A, C, E). The

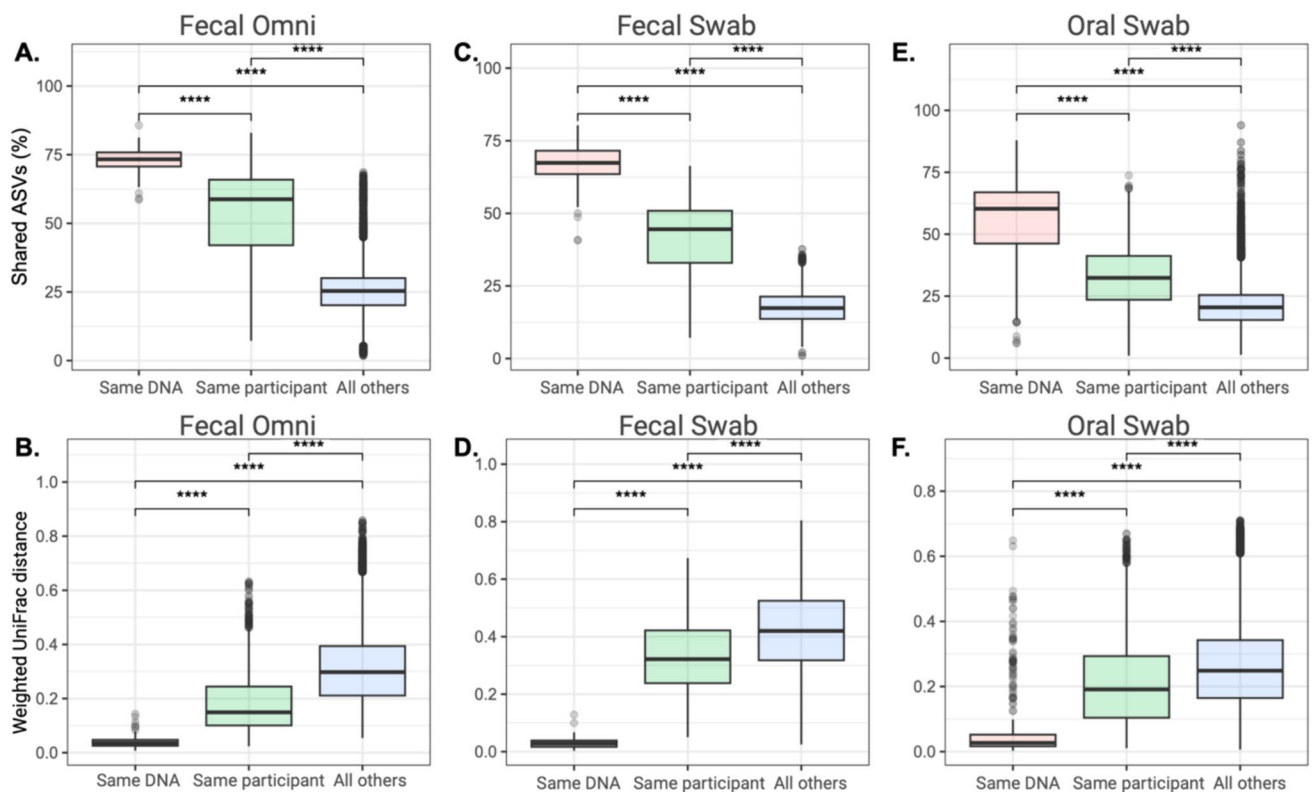


Fig. 3 Comparison based on either same participant sample in two different sequencing runs (Same DNA), same participant at different times from different sequencing runs, and remaining samples compared (all others) for as follows: Shared ASVs in ED-TBI Fecal Omni samples (A); Weighted UniFrac distance in ED-TBI Fecal

Omni samples (B); Shared ASVs in US-VMP Fecal Swab samples (C); Weighted UniFrac distance in US-VMP Fecal Swab samples (D); Shared ASVs in ED-TBI Oral Swab samples (E); Weighted UniFrac distance in ED-TBI Oral Swab samples (F). (**** $p < 0.001$)

same trend was observed in Weighted UniFrac distances (Fig. 3B, D, F). In all three sample types, the paired samples shared the most ASVs and had the most similar microbial community structure compared to samples from the same participant or other participants in the study.

Discussion

Accuracy is the agreement between a measured value and the items true value (Budowle et al. 2014) and was evaluated via an eight evenly distributed genera mock community standard. In the oversimplified mock community, genera were both over-represented (e.g., *Escherichia-Shigella* and *Salmonella*) and under-represented (e.g., *Pseudomonas* and *Lactobacillus*), potentially influenced by primer selection, extraction kit, sequencing machine, and bioinformatics pipeline (for more information see Karstens et al. (2021) or Abellan-Schneyder et al. (2021)). The limited accuracy observed when sequencing mock communities is unsurprising based on previous findings (Fouhy et al. 2016; Yeh et al. 2018); however, this concerning issue is outside the scope

of this manuscript. Advancements in mock communities (Mori et al. 2023) have increased the complexity (i.e., number of taxa) and added a focused target (i.e., taxa related to study area microbiome) that should improve accuracy in the future. Precision is the degree to which repeated measurements return the same results (Budowle et al. 2014). The simplified mock community had high stability across all eight genera and the microbial community. Precision was less stable in alpha diversity, perhaps due to the relatively small abundance of ASVs that had exaggerated impacts on these measures.

Precision was also assessed in more complex communities through sequencing the same DNA from positive control samples for fecal Omni, fecal swab, and oral swab samples. Assessment of fecal Omni and fecal swab positive controls had differing stability across sequencing runs with fecal Omni showing greater precision in taxonomic measures and alpha diversity in comparison to fecal swabs. Generally, the use of positive controls in microbiome studies has been low, recorded in under 10% of research up to 2018 (Hornung et al. 2019), yet the use of positive controls is expected to increase with the introduction of multiple

commercially available positive controls and literature on the subject. If precision (and therefore reliability) is of importance to a microbiome study or laboratory, the continual use of the same DNA across time is invaluable. We recommend the use of a pooled positive control sample from the same biogeographical region of interest and input DNA concentration in all sequencing runs to accurately assess sequencing precision and provide quality control metrics to decide acceptability of individual sequencing runs.

Fecal Omni samples and fecal swab samples taken from the same bowel movement and the same individuals had differing microbial communities, indicating either sample processing, DNA concentration, or another factor is important in stability across sequencing runs. We used participant samples across sequencing runs to assess the impact of DNA concentration on precision. The highest stability was observed for fecal Omni samples, followed by fecal swab samples, and then oral swab samples. DNA concentrations followed the same order with the highest concentrations in fecal Omni samples. The ability to sequence low biomass accurately is an important topic in microbiome research (Bender et al. 2018) with variability in precision introduced from exogenous bacterial DNA concentrations (Salter et al. 2014) that can be amplified through select bioinformatics pipelines (Caruso et al. 2019). Importantly in the present study for all sample types, the technical variation observed between sequencing runs was significantly lower than the biological variance of repeated samples from the same individual or other participants. Therefore, while variance due to sequencing runs should be assessed, the differences at a microbial community level appear to be a minor issue in clinical research.

Clinical studies of microbial communities have recently focused less on community measures (e.g., lower alpha diversity results in worse health outcomes) and more on individual taxa variations between timepoints or participant cohorts. Our results suggest care should be taken when reporting differences in taxa relative abundance from a 16S rRNA gene sequencing study across sampling runs. The most abundant taxa in our mock community and positive controls had %CVs across the sequencing runs from 5.3% to 81.1%, similar in magnitude that others have reported in microbiome quality control studies (Barlow et al. 2020; Bender et al. 2018). Although concerning, the observed %CVs for genus-level taxa are similar in magnitude to other historically used biological measurements across platforms for blood-based chemokines and cytokines (McKay et al. 2017). Microbiome analysis with ICC to determine stability is a recently used statistical approach in the field, first appearing in a 2017 manuscript on wild red squirrels (Ren et al. 2017). ICC revealed

that all eight genera in the mock community were stable. While the %CV at the genus level were variable across sequencing runs, their values were generally 75% less than %CV across studies again indicating biological variance dominates technical variance. However, technical variance at a genus level might still be an important factor when applying one of the many differentially abundant estimation tools that are often used in clinical research.

The present study has limitations including that the results were only obtained from one sequencing center over the period of 1 year. It is possible that the use of other sequencing centers or sequencing across a longer period of time could have different levels of consistency between sequencing runs compared to what we observed. Another limitation was the use of a simplified mock community with only eight bacterial genera. More recently developed mock communities have additional complexity and new protocols exist to assist laboratories in developing a study-specific mock community (Colovas et al. 2022). A strength of this study was the large sample size that included multiple positive controls, two sampling methods, and two human body locations. Additionally, the use of longitudinal samples enabled direct comparison of technical and biological variation.

In conclusion, this study investigated the technical variability introduced between sequencing runs on the resulting taxonomy, alpha diversity, and beta diversity. Specifically, we characterized the variability with a simplified mock community, positive control samples, and actual participant samples for two sampling methods (e.g., swab or commercially available stabilization kits) and two human sampling locations (i.e., fecal and oral). Based on our results, the following are recommendations for laboratories to understand and limit variation across sequencing runs: (1) use positive controls from the same biogeographic region in each sequencing run to assess variation; (2) consider more complex positive controls when feasible; (3) use a standardized stabilizing agent in sample collection; and (4) normalize DNA concentrations pre-amplification. Given the number of sequencing studies that exist to date, the development of bioinformatics tools to adequately adjust results post sequencing is an important knowledge gap. These results provide a context for technical variability in microbiome studies that span multiple sequencing runs, between studies from the same laboratory, and between laboratories that use identical standard operating procedures. Additionally, results provide a context to establish meaningful biological variances that are not attributed to technical variance that can be used to verify adequate sequencing run quality or adjust power estimates for sample size calculations.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00253-024-13198-z>.

Author contribution AJH, JCE, CAL, and LAB conceived and designed the study. AJH and CES conducted the experiments and performed the analysis. AJH drafted the initial version of the manuscript. Edits with significant changes were introduced by all the authors. All authors read and approved the manuscript.

Funding This project was in part supported by the Department of Veterans Affairs (VA) Rocky Mountain Mental Illness Research Center (MIRECC) for Suicide Prevention (study: US-VMP) and the MIND-SOURCE Brain Injury Network, Contract IHEA #140755 (study: ED-TBI).

Data availability Demultiplexed single-end sequences were deposited in the NCBI Sequence Read Archive (BioProject accession ID: PRJNA1101562).

Declarations

Ethics approval and consent to participate All procedures performed in studies involving human participants were in accordance with the ethical standards of the institution and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Competing interests Dr. Brenner reports grants from the VA, DOD, NIH, and the State of Colorado, editorial support from Wolters Kluwer, and royalties from the American Psychological Association, Oxford University Press, and the Rand Corporation. In addition, she consults with sports leagues via her university affiliation. Dr. Lowry reports grants from the VA, NIH, NSF, and Institute for Cannabis Research.

Disclaimer The views, opinions, and/or findings contained in this article are those of the author(s) and should not be construed as an official Department of Defense or VA position, policy, or decision unless so designated by other documentation.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abellan-Schneyder I, Machado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, List M, Neuhaus K (2021) Primer, pipelines, parameters: Issues in 16S rRNA gene sequencing. *mSphere* 6(1):10. <https://doi.org/10.1128/msphere.01202-20>
- Allali I, Arnold JW, Roach J, Cadenas MB, Butz N, Hassan HM, Koci M, Ballou A, Mendoza M, Ali R, Azcarate-Peril MA (2017) A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiol* 17(1):194. <https://doi.org/10.1186/s12866-017-1101-8>
- Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A (2017) Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems* 2(2):e00191–e00116
- Amos GCA, Logan A, Anwar S, Fritzsche M, Mate R, Bleazard T, Rijpkema S (2020) Developing standards for the microbiome field. *Microbiome* 8(1):98. <https://doi.org/10.1186/s40168-020-00856-3>
- Barlow JT, Bogatyrev SR, Ismagilov RF (2020) A quantitative sequencing framework for absolute abundance measurements of mucosal and luminal microbial communities. *Nat Commun* 11(1):2590. <https://doi.org/10.1038/s41467-020-16224-6>
- Bender JM, Li F, Adisetiyo H, Lee D, Zabih S, Hung L, Wilkinson TA, Pannaraj PS, She RC, Bard JD, Tobin NH, Aldrovandi GM (2018) Quantification of variation and the impact of biomass in targeted 16S rRNA gene sequencing studies. *Microbiome* 6(1):155. <https://doi.org/10.1186/s40168-018-0543-z>
- Bobak CA, Barr PJ, O'Malley AJ (2018) Estimation of an inter-rater intra-class correlation coefficient that overcomes common assumption violations in the assessment of health measurement scales. *BMC Med Res Methodol* 18(1):1–11
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciulek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37(8):852–857. <https://doi.org/10.1038/s41587-019-0209-9>
- Brenner LA, Hoisington AJ, Stearns-Yoder KA, Stamper CE, Heinze JD, Postolache TT, Hadidi DA, Hoffmire CA, Stanislawski MA, Lowry CA (2018) Military-related exposures, social determinants of health, and dysbiosis: the United States-Veteran Microbiome Project (US-VMP). *Front Cell Infect Microbiol* 400(8). <https://doi.org/10.3389/fcimb.2018.00400>
- Budowle B, Connell ND, Bielecka-Oder A, Colwell RR, Corbett CR, Fletcher J, Forsman M, Kadavy DR, Markotic A, Morse SA, Murch RS, Sajantila A, Schmedes SE, Kerns KL, Turner SD, Minot S (2014) Validation of high throughput sequencing and microbial forensics applications. *Invest Genet* 5(1):9. <https://doi.org/10.1186/2041-2223-5-9>
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6(8):1621–1624
- Cardona S, Eck A, Cassellas M, Gallart M, Alastrue C, Dore J, Azpiroz F, Roca J, Guarner F, Manichanh C (2012) Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC Microbiol* 12(1):158. <https://doi.org/10.1186/1471-2180-12-158>
- Caruso V, Song X, Asquith M, Karstens L (2019) Performance of microbiome sequence inference methods in environments with varying biomass. *mSystems* 4(1). <https://doi.org/10.1128/msystems.00163-18>

- Colovas J, Bintarti AF, Mechan Llontop ME, Grady KL, Shade A (2022) Do-it-yourself mock community standard for multi-step assessment of microbiome protocols. *Curr Protocols* 2(9):e533. <https://doi.org/10.1002/cpz1.533>
- Fouhy F, Clooney AG, Stanton C, Claesson MJ, Cotter PD (2016) 16S rRNA gene sequencing of mock microbial populations-impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiol* 16(1):123. <https://doi.org/10.1186/s12866-016-0738-z>
- Gilbert JA, Jansson JK, Knight R (2014) The Earth Microbiome Project: successes and aspirations. *BMC Biol* 12:1–4
- Hornung BVH, Zwittink RD, Kuijper EJ (2019) Issues and current standards of controls in microbiome research. *FEMS Microbiol Ecol* 95(5). <https://doi.org/10.1093/femsec/fiz045>
- Karstens L, Siddiqui NY, Zaza T, Barstad A, Amundsen CL, Sysoeva TA (2021) Benchmarking DNA isolation kits used in analyses of the urinary microbiome. *Sci Rep* 11(1):6186. <https://doi.org/10.1038/s41598-021-85482-1>
- Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, Louis P, Thomson JM (2014) The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS ONE* 9(2):e88982. <https://doi.org/10.1371/journal.pone.0088982>
- Lozupone C, Hamady M, Knight R (2006) UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* 7:371. <https://doi.org/10.1186/1471-2105-7-371>
- McDonald D, Hyde ER, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, DeRight Goldasich L, Dorrestein PC, Dunn RR, Fahimipour AK, Gaffney J, Gilbert JA, Gogul G, Green JL, Hugenholtz P, Humphrey G, Huttenhower C, Jackson MA, Janssen S, Jeste DV, Jiang L, Kelley ST, Knights D, Kosciulek T, Ladau J, Leach J, Marotz C, Meleshko D, Melnik AV, Metcalf JL, Mohimani H, Montassier E, Navas-Molina J, Nguyen TT, Peddada S, Pevzner P, Pollard KS, Rahnavard G, Robbins-Pianka A, Sangwan N, Shorestein J, Smarr L, Song SJ, Spector T, Swafford AD, Thackray VG, Thompson LR, Tripathi A, Vazquez-Baeza Y, Vrbancac A, Wischmeyer P, Wolfe E, Zhu Q, Knight R (2018) American gut: an open platform for citizen-science microbiome research. *mSystems* 3(3):e00031–e00018. <https://doi.org/10.1128/mSystems.00031-18>
- McKay HS, Margolick JB, Martínez-Maza O, Lopez J, Phair J, Rappocciolo G, Denny TN, Magpantay LI, Jacobson LP, Bream JH (2017) Multiplex assay reliability and long-term intra-individual variation of serologic inflammatory biomarkers. *Cytokine* 90:185–192. <https://doi.org/10.1016/j.cyto.2016.09.018>
- Mirzayi C, Renson A, Furlanello C, Sansone S-A, Zohra F, Elsafoury S, Geistlinger L, Kasselmann LJ, Eckenrode K, van de Wijgert J, Loughman A, Marques FZ, MacIntyre DA, Arumugam M, Azhar R, Beghini F, Bergstrom K, Bhatt A, Bisanz JE, Braun J, Bravo HC, Buck GA, Bushman F, Casero D, Clarke G, Collado MC, Cotter PD, Cryan JF, Demmer RT, Devkota S, Elinav E, Escobar JS, Fettweis J, Finn RD, Fodor AA, Forslund S, Franke A, Furlanello C, Gilbert J, Grice E, Haibe-Kains B, Handley S, Herd P, Holmes S, Jacobs JP, Karstens L, Knight R, Knights D, Koren O, Kwon DS, Langille M, Lindsay B, McGovern D, McHardy AC, McWeeney S, Mueller NT, Nezi L, Olm M, Palm N, Pasolli E, Raes J, Redinbo MR, Rühlemann M, Balfour Sartor R, Schloss PD, Schriml L, Segal E, Shardell M, Sharpton T, Smirnova E, Sokol H, Sonnenburg JL, Srinivasan S, Thingholm LB, Turnbaugh PJ, Upadhyay V, Walls RL, Wilmes P, Yamada T, Zeller G, Zhang M, Zhao N, Zhao L, Bao W, Culhane A, Devanarayan V, Dopazo J, Fan X, Fischer M, Jones W, Kusko R, Mason CE, Mercer TR, Sansone S-A, Scherer A, Shi L, Thakkar S, Tong W, Wolfinger R, Hunter C, Segata N, Huttenhower C, Dowd JB, Jones HE, Waldron L, Genomic Standards C, Massive A, Quality Control S (2021) Reporting guidelines for human microbiome research: the STORMS checklist. *Nat Med* 27(11):1885–1892. <https://doi.org/10.1038/s41591-021-01552-x>
- Mori H, Kato T, Ozawa H, Sakamoto M, Murakami T, Taylor TD, Toyoda A, Ohkuma M, Kurokawa K, Ohno H (2023) Assessment of metagenomic workflows using a newly constructed human gut microbiome mock community. *DNA Res* 30(3). <https://doi.org/10.1093/dnares/dsad010>
- Oksanen J, Kindt R, Legendre P, O'Hara B, Simpson GL, Solymos P, Stevens MHH, Wagner H (2008) *vegan: community ecology package*. vol R package version 2.6–4, <https://cran.r-project.org/web/packages/vegan/index.html>. Accessed 26 Oct 2023
- Peterson RA (2021) Finding optimal normalizing transformations via bestNormalizer. *R J* 13(1):310–329. <https://doi.org/10.32614/RJ-2021-041>
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplins J, Glöckner FO (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41(D1):D590–D596. <https://doi.org/10.1093/nar/gks1219>
- Ren T, Boutin S, Humphries MM, Dantzer B, Gorrell JC, Coltman DW, McAdam AG, Wu M (2017) Seasonal, spatial, and maternal effects on gut microbiome in wild red squirrels. *Microbiome* 5:1–14
- Salamon D, Zapała B, Krawczyk A, Potasiewicz A, Nikiforuk A, Stój A, Gosiewski T (2022) Comparison of iSeq and MiSeq as the two platforms for 16S rRNA sequencing in the study of the gut of rat microbiome. *Appl Microbiol Biotechnol* 106(22):7671–7681. <https://doi.org/10.1007/s00253-022-12251-z>
- Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12(1):87. <https://doi.org/10.1186/s12915-014-0087-z>
- Santiago A, Panda S, Mengels G, Martinez X, Azpiroz F, Dore J, Guarner F, Manichanh C (2014) Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol* 14(1):112. <https://doi.org/10.1186/1471-2180-14-112>
- Sinha R, Abnet CC, White O, Knight R, Huttenhower C (2015) The microbiome quality control project: baseline study design and future directions. *Genome Biol* 16(1):276. <https://doi.org/10.1186/s13059-015-0841-8>
- Sinha R, Chen J, Amir A, Vogtmann E, Shi J, Inman KS, Flores R, Sampson J, Knight R, Chia N (2016) Collecting fecal samples for microbiome analyses in epidemiology studies. *Cancer Epidemiol Biomarkers Prev* 25(2):407–416. <https://doi.org/10.1158/1055-9965.Epi-15-0951>
- Stoffel MA, Nakagawa S, Schielzeth H (2017) rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol Evol* 8(11):1639–1644. <https://doi.org/10.1111/2041-210X.12797>
- The R Core Team (2020) *R: A language and environment for statistical computing*. 4.0.0 edn
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett C, Knight R, Gordon JI (2007) The Human Microbiome Project: exploring the microbial part of ourselves in a changing world. *Nature* 449(7164):804–810. <https://doi.org/10.1038/nature06244>
- Vogtmann E, Chen J, Amir A, Shi J, Abnet CC, Nelson H, Knight R, Chia N, Sinha R (2017) Comparison of collection methods for fecal samples in microbiome studies. *Am J Epidemiol* 185(2):115–123. <https://doi.org/10.1093/aje/kww177>
- Yeh Y-C, Needham DM, Sieradzki ET, Fuhrman JA (2018) Taxon disappearance from microbiome analysis reinforces the value of mock communities as a standard in every sequencing run. *mSystems* 3(3). <https://doi.org/10.1128/mSystems.00023-18>