

# Comparative Investigation of the Suitability of DNA Preparation Methods for Microbiome Analysis of Complex Clinical Samples

**Brendon Coenraad Mann**

North-West University Potchefstroom Campus: North-West University

**Cara Wepener**

North-West University Potchefstroom Campus: North-West University

**Johannes Jacobus Bezuidenhout** (✉ [10926542@nwu.ac.za](mailto:10926542@nwu.ac.za))

North-West University <https://orcid.org/0000-0003-3798-2848>

**Urban Vermeulen**

Austell Pharmaceuticals

**Anne Grobler**

Austell Pharmaceuticals

---

## Research Article

**Keywords:** DNA preparation, Mechanical lysis, Bead beating, 16S rRNA gene sequencing, Microbiome

**Posted Date:** February 18th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-207952/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Efficient DNA preparation is essential for accurate and reproducible microbial data acquisition. In this study an in-house method (NWU lysis system), involving a lysis micro tube (LMT) was employed to determine its suitability for safe, rapid and accurate characterisation of the bacterial microbiome associated with clinical samples in comparison to two available commercial DNA extraction kits. During the experimental setup, it was confirmed that the LMT is suitable for downstream bacterial microbiome analyses with the incorporation of short read sequencing; but due to the fragmented nature of the DNA its suitability is highly dependent on the sequencing technology applied downstream. The study also served to demonstrate that shearing of DNA can have a significant impact on downstream microbiome analyses, based on the sequencing technology used; and that care must be taken especially if long read sequencing is to be used employed.

## Introduction

Since the onset of next generations sequencing (NGS) the field of clinical genetics has been rapidly changing (Vrijenhoek et al., 2015). The swift progression of NGS technology and its use in clinical laboratories has allowed for remarkable progress in the genetic diagnosis of both inherited disorders and infectious diseases, as well as the provision of methods for studying accompanying risk factors such as the microbiome (Di Reste *et al.*, 2018). Novel interactions between the human microbiome and health and disease are continuously emerging, and important host–microbiome interactions are increasingly used as targets for both diagnostics and personalised therapeutics (Behrouzi et al., 2019). The majority of NGS platforms still carry high initial infrastructure costs, have extensive computing requirements and are not of a compact and portable nature. These limitations have resulted in the introduction of platforms such as the MinION developed by Oxford nanopore technologies (ONT), as well as rapidly reducing the cost associated with current technologies such as the Illumina Miseq (Besser et al., 2018). Taking the rapid development, improvement and reduction in costs of various NGS platforms into account, many clinical laboratories and research entities are preparing and/or are already in the progress of adapting their diagnostic process to incorporate or focus mainly on NGS approaches Vrijenhoek et al., 2015.

Sequencing of the 16S gene is a commonly employed method to evaluate the microbial community in clinical samples (Pollock et al., 2018). Previous studies have emphasized that effective DNA preparation is crucial for the accurate characterisation of the complex bacterial populations often present in clinical samples, and that lysis efficiency is a major limitation, especially with samples containing hard to lyse bacteria (Lim et al., 2018). Several methods are currently available for effective DNA preparation prior to 16S sequencing, in the form of various commercial kits. These approaches to cellular lysis of available commercial DNA preparation methods differ, and the lysis method may depend on heat, chemical, enzymatic, or mechanical lysis. Additionally, the specific DNA isolation techniques used also differ, such as spin column-based methods or the use of magnetic beads; while others have opted for direct DNA amplification of crude lysates, forgoing conventional DNA purification (Videvall et al., 2017; Pollock et al., 2018; Fiedorová et al., 2019)

The NWU-TB test is a molecular-based sensitive and specific TB diagnostic tool. The NWU-TB system consists of three sequential steps: cell lysis in a LMT (Supplementary Fig. 2), using an automated lyser device (NWU lysis system: Supplementary Fig. 3) completely inactivating TB within 7 minutes, followed by multiplex-PCR within 25 min. The lysis step of the NWU-TB test is a highly efficient and cost effective process that is based on the use of heat, chemical and mechanical means to achieve cellular lysis. In addition to its lysis efficacy, the lysis process is also extremely safe, with lysis occurring within the confines of a lysis micro tube, offering the potential for this system to be used with decreased biosafety measures (Mutingwende et al., 2015).

Authors of recent studies have emphasized that the inclusion of mechanical lysis in conjunction with other lysis methods is essential to minimize possible biases due to some microbial cells being more resistant to lysis than others (Lim et al., 2018; Pollock et al., 2018)

The decision was thus made to evaluate DNA obtained by means of the NWU lysis method as preferential tool for downstream microbiome analyses on two sample types: a clinical sputum sample evaluated with long read sequencing technology (ONT MinION), representing a lower complexity sample and a more complex gut sample evaluated with short read sequencing technology (Illumina Miseq), focussing on 16S amplicon-based sequencing in comparison to DNA obtained with the use of various commercially available kits.

## Materials And Methods

### Samples and DNA preparation

DNA obtained by means of the NWU lysis method (method H) was compared to DNA obtained by means of two available commercial kits on both the *caecum* contents from an available mouse study (representing a gut sample), and a previously collected less complex clinical sputum sample; thus representing two samples types of varying complexity. The two commercial kits selected for comparison included the QIAamp DNA microbiome kit (Qiagen, Germany) and the GenElute™ Stool DNA Isolation Kit (Sigma, USA), and were labelled as method Q and S respectively. These kits were selected on the basis that both were bead-beating kits, and thus included both chemical and mechanical lysis to ensure that difficult to lyse bacteria are properly lysed during the DNA preparation process.

### Sputum sample

For the evaluation of a low complexity clinical sample, a sputum sample of sufficient volume was collected from a concurrently running study. The sample was initially collected, snap frozen and stored at -80°C until processing. The selected sample was divided into 12 x 250 µl aliquots before further processing. Aliquots were extracted in triplicate using either method Q or S; while lysate was prepared with the NWU in-house cell lysis method (H) with and without additional purification (carried out with the spin column technology of method Q and labelled as HQ).

Extractions with method Q and S were done according to the manufacturers' instructions. Freeze–thaw cycles may compromise bacterial integrity, and the benzonase treatment used during the host DNA removal Protocol of method Q may lead to a loss of exposed bacterial DNA. The decision was thus made to omit the host DNA removal step from method Q considering the samples have been previously frozen. Preparation of the crude lysate using the NWU in-house cell lysis method was done according to previously described methods (Mutingwende et al., 2015). For the NWU cell lysis method, lysis was carried out using the NWU lyser device: 250 µl of sample was mixed with 250 µl of a proprietary lysis buffer. The LMT was placed on the pre-set (95°C and 3600 rpm) lyser device for 7 min. Bacterial cell lysis was concurrently achieved through chemical, thermal and mechanical means (Mutingwende et al., 2015).

DNA concentration was assessed using the Qubit 4 Fluorometer (ThermoFisher Scientific, USA) along with the Qubit BR assay kit (ThermoFisher Scientific, USA), while quality was determined by nanodrop spectrophotometry on a Nanodrop One (ThermoFisher Scientific, USA). The integrity of extracted genomic DNA was evaluated by visualization of a 1.5% (w/v) agarose gel using GelRed® dye (Biotium, USA), after electrophoresis in the presence of a 1kb ladder as size reference standard.

## Representative gut sample

The *caecum* content from a concurrently running mouse study was investigated as a complex sample. As part of the initial study the *caecums* were dissected, snap frozen and stored at -80°C prior to further analysis. To generate sufficient sample volume, three *caecums* from C3HeB/JeF mice were physically cut open and the content was added to phosphate buffered saline (PBS). The mixture was vortexed at high speed for 2 minutes to generate a homogenized sample; 250 µl of this mix was then used for DNA extraction and lysate preparation in triplicate. DNA was extracted and/or prepared, subjected to quality control and samples were labelled as described under Sect. 2.1.1

## Amplicon library and flow cell preparation

Two sequencing runs were carried out; one for the sputum sample and one for the gut sample. For the sputum sample a 16S rRNA sequencing library was constructed according to the 16S Barcoding Kit (SQK-RAB204) Protocol (Oxford Nanopore Technologies, Oxford, UK) for sequencing on the ONT MinION platform. Library construction for the gut sample was performed according to the 16S metagenomics sequencing library preparation protocol (Illumina, San Diego, CA, USA) for sequencing on the Illumina MiSeq platform.

## Sputum sample

Sequencing of the sputum sample was carried out at the North-West University using the ONT 16S Barcoding Kit (SQK-RAB204) according to the ONT Protocol with the only difference being the use of an

inhibitor tolerant high-fidelity polymerase. Polymerase chain reaction barcoding amplification was conducted on a C1000™ Thermal Cycler (Bio-Rad, US). A 50 µl reaction volume consisting of: 1 µl (10 µM) of 16S barcode primer (Oxford Nanopore Technologies, Oxford, UK), 25 µl of Invitrogen Platinum SuperFi DNA Polymerase master mix (Thermo Fisher Scientific, USA), 10 µl of GC enhancer (Thermo Fisher Scientific, USA), 13 µl nuclease-free water (Sigma-Aldrich, USA) and 1 µl of template DNA (10 ng/µl) was prepared for each sample. In the case of the lysate produced by the NWU in-house cell lysis method, 1 µl of sample was added to the reaction mixture. PCR cycling conditions were set at 95°C for 1 min followed by 25 cycles of denaturation at 95°C for 20 s, annealing at 55°C for 30 s, extension at 65°C for 2 min and a final extension step of 65°C for 5 min before holding at 4°C. PCR products were cleaned using AMPure XP beads (Beckman Coulter, USA) and eluted in 10 µl of a buffer containing 10 mM Tris-HCl pH 8.0 and 50 mM NaCl. Following PCR, 1 µl of eluted sample was quantified using a Qubit fluorometer in order to pool the DNA barcoded libraries at an equal ratio. All barcoded libraries were pooled in the desired ratios to a total of 50–100 fmoles in 10 µl of 10 mM Tris-HCl pH 8.0 and/ 50 mM NaCl. Platform quality control (QC) was carried out using MinKNOW™ on 2 new R9.4.1 chemistry MinION™ flow cells before the flow cell was primed. In total 75 µl of sequencing mix consisting of the DNA library, sequencing buffer and library loading beads was prepared according to the ONT Protocol and added in a drop-wise fashion via the SpotOn sample port. The standard 48 h sequencing script was chosen with 1D live base calling.

## Representative gut sample

The V3 and V4 regions of the 16S rRNA gene from the isolated microbiome were amplified on the C1000™ Thermal Cycler (Bio-Rad, US). A 25 µl reaction volume consisting of: 1 µl (5 µM) of forward primer, 1 µl (5 µM) of reverse primer, 12.5 µl of Invitrogen Platinum SuperFi DNA Polymerase master mix (Thermo Fisher Scientific, USA), 5 µl of GC enhancer (Thermo Fisher Scientific, USA), 4.5 µl nuclease-free water (Sigma-Aldrich, USA) and 1 µl of template DNA (20 ng/µl) was prepared for each sample. In the case of the lysate produced by the NWU in-house cell lysis method, 1 µl of sample was added to the reaction mixture. The following PCR conditions were used during amplification: initial denaturation at 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 55°C for 10 s and 72°C for 30 s and a final elongation step at 72°C for 300 s. After amplification, PCR products were purified using the Agencourt AMPure XP PCR Purification kit (Beckman Coulter, USA); and transported to the Agricultural Research Council Biotechnology Platform (ARC-LNR), Pretoria, South Africa, for sequencing on the Illumina MiSeq platform according to the standard protocol.

## Bioinformatics analysis

Fast 5 files generated from the sputum sample sequencing runs on the MinION™ were base called and demultiplexed, followed by removal of adapter and primer sequences using ONT's Guppy™ sequencing software (version 3.2.4). The resultant fastq files were then filtered to remove reads with a Phred score below 7; and lengths below 1200 and above 1500bp, using NanoFilt

(<https://github.com/wdecoster/nanofilt>) (De Coster et al., 2018). Taxonomy was assigned to sequences using the `sintax` command from Usearch (Edgar, 2018), with a `sintax` cut-off of 0.8. Sequences were classified using the RDP 16S training set v16 (RTS) database comprising 13,212 sequences belonging to 2,126 genera. Following classification, the output `sintax` files were processed with in-house R scripts to produce OTU tables and excel summary files. The OTU table along with a mapping file were then fed into the MicrobiomeAnalyst online software suite for further evaluation, which included alpha and beta diversity determination, rarefaction curve generation and clustering analyses (Dhariwa *et al.*, 2017; Chong et al., 2020).

Fastq files produced by the Illumina MiSeq platform were processed and analysed using the `mothur` (version 1.42.3) shell program (Schloss et al., 2009). The sequences were aligned with SILVA-based bacterial reference alignment (version July 2019). Chimeric reads were filtered out using the abundant sequences (`dereplicate = t`) as a reference with the `VSEARCH` command in the shell program. Host DNA sequences along with other undesirable sequences were removed with the `remove.lineage` command where the DNA of chloroplasts, mitochondria, archaea and eukaryotes were removed. Sequences that shared a minimum of 97% pairwise nucleotide identity were clustered into operational taxonomic units (OTUs). OTUs were classified up to species level using the 16S rRNA RDP reference (version 16) and the `classify.otu` command in the `mothur` shell program. The abundance file obtained using the `make.shared` command and the taxonomy file obtained using the `classify.otu` command was exported for further analyses (Schloss et al., 2009).

## Statistical analysis

To evaluate the influence of each extraction method on the DNA quantity/yield and quality (A260/A280 and A260/A230); ANOVA (one-way analysis of variance) was employed with Tukey's post-hoc test for multiple pairwise comparisons [15]. Statistical testing was carried out using Statistica (v13.1) (Statsoft, Inc, USA) and visualised in GraphPad Prism (v8) (GraphPad, Inc., USA). Statistical analyses and visualisation of microbiome data were done with the assistance of MicrobiomeAnalyst on default settings unless specified otherwise (Dhariwa *et al.*, 2017; Chong et al., 2020). Data was normalised using total sum scaling (TSS). The alpha diversity of samples was measured by observed, Chao1 and Shannon diversity indexes (Corcoll et al., 2017). The observed and Chao1 indices act as measures of species richness. The observed species index measures the number of distinguishable taxa in every sample; whereas the Chao1 index is a qualitatively measure, which beside species richness also takes into account the ratio of singletons, and hence gives more weight to rare species. The Shannon diversity index on the other hand, is a measure of both richness and evenness of the microbes of the given sample; thus addressing the question of whether there is evenness and possible domination in the main genera/species found in the sample (Xia & Sun, 2017).

Beta-diversity calculations, which describes the diversity in a microbial community between different samples, were visualized using principal coordinate analysis plots (PCoA). Beta-diversity PCoA plots were based on Bray–Curtis distances, and compared using the nonparametric analysis of similarities

(ANOSIM) test. Heat maps of the most abundant genera classified to the genus level were generated using complete hierarchical clustering by Euclidian distance (Dhariwa *et al.*, 2017; Xia & Sun, 2017; Chong *et al.*, 2020). These tests were performed at the feature level with only the higher abundance taxa (> 0.1% of total).

## Ethics

This study was approved by the North-West University Research Ethics Committee under ethics numbers: NWU-00127-18-A1 and NWU-00584-19-A5.

## Results

### Yield and quality of prepared nucleic acids

The concentration and quality of the prepared nucleic acids were compared for the various methods used for both sample types as summarised in Table 1. Method H without the addition of a purification step yields a crude lysate and thus no quality assessments could be obtained.

Table 1  
DNA Quality

Sample	Nanodrop Concentration (ng/ $\mu$ L)	A260/A280	A260/A230
<b>Sputum sample</b>			
H	-	-	-
HQ	15,03 $\pm$ 0,13 <sup>a</sup>	1,95 $\pm$ 0,03 <sup>a</sup>	1,96 $\pm$ 0,38 <sup>b</sup>
Q	186,53 $\pm$ 5,89 <sup>b</sup>	1,87 $\pm$ 0 <sup>b</sup>	2,15 $\pm$ 016 <sup>b</sup>
S	20,2 $\pm$ 0,31 <sup>a</sup>	2,12 $\pm$ 0,03 <sup>c</sup>	0,13 $\pm$ 0,12 <sup>a</sup>
<b>Gut sample</b>			
H	-	-	-
HQ	10.43 $\pm$ 2.74 <sup>a</sup>	1.77 $\pm$ 0.01 <sup>b</sup>	1.33 $\pm$ 0.27 <sup>a,b</sup>
Q	610.5 $\pm$ 111.49 <sup>b</sup>	2.06 $\pm$ 0.04 <sup>a</sup>	1.86 $\pm$ 0.41 <sup>b</sup>
S	79.87 $\pm$ 23.72 <sup>a</sup>	1.96 $\pm$ 0.05 <sup>a</sup>	0.67 $\pm$ 0.34 <sup>a</sup>

Values sharing a common letter belong to the same group according to one way ANOVA with Tukey's post-hoc test for multiple pairwise comparisons. These values do not differ significantly according to the test ( $p > 0.05$ ). Values not sharing a common lower case letter indicate a statistically significant difference in comparison to a value belonging to other groups ( $p < 0.05$ ).

## Sputum sample

Initial nanodrop results (Table 1) indicate the nucleic acid concentration obtained with method Q was significantly higher ( $p < 0.05$ ) when compared to HQ, and S. All methods with the exception of method S (2.1) had a A260/A280 absorbance ratio between 1.7-2.0, suggesting the preparation of pure nucleic acids. Both methods HQ and Q had A260/A230 ratios within the acceptable range, while method S had a significantly lower A260/A230 ratio. The integrity of the prepared DNA was assessed by agarose gel electrophoresis. Method S yielded high molecular weight intact DNA with minimal shearing, while method Q produced large amounts of high molecular weight DNA, but shearing was also evident. Method HQ on the other hand produced sheared low molecular weight DNA.

## Representative gut sample

Upon examination of DNA obtained from the gut sample, a similar pattern in the average nucleic acid concentrations obtained can be seen (Table 1). Method Q again had a significantly higher nucleic acid concentration when compared to other methods.

Again all had acceptable A260/A280 absorbance ratios between 1.7-2.0, while only method Q had A260/A230 ratios within the acceptable range. DNA integrity was assessed by agarose gel electrophoresis and revealed similar results to those obtained with the selected sputum sample.

NGS analyses of bacterial diversity and composition of samples from various DNA preparation methods

NGS analyses of bacterial diversity and composition from various DNA preparation methods were carried out on sequencing reads produced for both sample types.

## Sputum sample

A single sequencing run on a MinION™ flow cell of R9.4 chemistry, produced 5 169 844 reads from 12 DNA barcodes. Platform quality control (QC) analysis preceding the sequencing run revealed a total of 1356 available pores, split into 4 groups. The mean read length was 1352 base pairs (bp), with an average Phred score of 10.01. The alpha diversity based on the number of observed species, Chao1 and Shannon diversity indices, was compared according to the DNA preparation method (Fig. 1: A).

As seen in Fig. 1A, there was no significant difference between methods HQ, Q and H, based on the Shannon, Chao1 and observed indices, while method S ranked significantly higher than any of the other methods for all three alpha diversity measures. Clustering analyses were performed by means of a hierarchical clustering heat map (Fig. 2), which revealed three clusters: Cluster one, consisting of method H and HQ, and one cluster each for methods Q and S respectively. In samples extracted with method S, Gram-negative bacterial genera were more abundant, while samples extracted with method Q resulted in an increased abundance for Gram-positive bacterial genera.

Principal coordinate analysis of beta-diversity yielded similar results, showing significant clustering according to the method. Methods H and HQ clustered together, while methods Q and S formed unique clusters (Supplementary Fig. 1(A)). Analyses of similarity confirmed significant variability among the preparation methods ([ANOSIM] R: 0.74691; p-value < 0.001).

## Gut sample

The resulting run on the Illumina Miseq platform produced 4 582 278 reads for 12 samples, with an average quality of 32.58 and an average length of 453 bp. Again the alpha diversity based on the number of observed species, Chao1 and Shannon diversity indices, was compared according to the DNA preparation method (Fig. 1: B). For the representative gut sample, there was no significant difference between methods H, HQ, Q and S based on the Shannon, Chao1 and observed indices. Clustering analyses performed by means of a hierarchical clustering heat map, revealed random clustering among samples, also suggesting no statistically significant difference based on the various preparation methods (Fig. 3).

This was again confirmed by principal coordinate analysis of beta-diversity (Supplementary Fig. 1(B)), which revealed no statistically significant difference between samples based on preparation method. Analyses of similarity confirmed no significant variability among the preparation methods ([ANOSIM] R: -0.11728; p-value < 0.759).

## Discussion

The A260/A280 ratio is viewed as the primary measure of DNA purity, and DNA with a ratio between 1.70 and 2.00 is generally accepted as pure (Lucena-Aguilar et al., 2016; Teng et al., 2018). All methods evaluated produced relatively pure DNA when examining the A260/A280 absorbance ratio; although the A260/A230 ratio was generally lower for all methods evaluated, especially for method S. The expected value for the A260/A230 absorbance ratio is between 2.0 and 2.22. The low A260/A230 may be due to such factors as carbohydrate carryover, proteins, lipids, EDTA, phenol or salts such as guanidine HCL, a common substance found in various commercial DNA extraction kits (Lucena-Aguilar et al., 2016). None the less, the A260/A230 ratio is viewed as a secondary measure of DNA quality, due to the instability of this value when certain elution buffers are used to elute the DNA (Lucena-Aguilar et al., 2016). PCR amplicons were successfully obtained from all samples; thus the results indicate that all methods yielded nucleic acid of sufficient quality and quantity for sequencing of the 16S rRNA gene (Pollock et al., 2018). Fragmentation of DNA was assessed by agarose gel electrophoresis. The DNA extracted with methods S and Q yielded high molecular weight DNA. Method H was originally designed to form part of a TB diagnostic system and as such the lysis procedure is incredibly robust primarily producing sheared single stranded DNA. The method was originally incorporated with real time qPCR, involving the amplification of short fragments of DNA; and the potential impact of shearing on larger PCR products has not been previously investigated (Kennedy *et al.*, 2009).

NGS analyses of bacterial diversity and composition were compared between different DNA preparation methods and with two sequencing approaches. Evaluations of a representative gut sample with the incorporation of short read sequencing on the Illumina Miseq platform indicated no statistically significant difference between either DNA preparation methods at the end of analyses of the microbiome. Analysis of the sputum sample with the incorporation of long read sequencing, on the other hand showed a statistically significant difference when compared to one of the commercial methods (S). When comparing methods H and HQ it is evident that purification of the DNA obtained with method H did not have any significant impact on the end results and that the differences observed were the result of the level of integrity of the DNA and the sequencing method applied. The DNA obtained with methods H, HQ and Q was consistently sheared, which may not be a problem when sequencing short fragments; but for single molecule or long-read sequencing technologies, particularly those from Pacific Biosciences (PacBio) and ONT, this can be a major issue (Lim et al., 2018). The MinION 16S workflow incorporates full length 16S amplicons of around 1500bp, and it would thus follow that a workflow incorporating larger amplicons will fail to capture the true diversity of a sample if there is a low amount of intact DNA, as illustrated by the current results.

A limitation of the present study was the use of only one representative sample for each evaluation. It was therefore not possible to determine whether the variation between the preparation methods was less than the inter-subject variation. According to previous studies the inter-subject variation generally tends to be greater than the technical variation. Hence, it is expected that analysis of additional samples would be in accordance with those of prior accounts (Lim et al., 2018; Kennedy et al., 2014). This study focussed on an evaluation of a new method, producing a crude lysate and its efficiency in comparison to already existing validated commercial DNA extraction methods (Wagner Mackenzie et al., 2015; Lim et al., 2018). Various studies tend to use mock microbial samples of known composition to evaluate new methods, but these mock samples also have their limitations: They are often presented in a liquid or broth and as such lack the complexity of real biological samples such as sputum, in which the viscosity of the sample is a major stumbling block. Furthermore, mock microbial communities tend to have a lower degree of diversity when compared to actual clinical samples, which harbour a much more diverse bacterial population. Given the growing significance of the microbiome, new products and tools are consistently being developed to study the human microbiome and its various relations to health and disease (Mohajeri et al., 2018). Despite the mentioned limitations, the current study provides useful information, describing the potential of a LMT to carry out downstream microbiome analyses with improved speed and safety.

## Conclusion

To conclude, the NWU lysis system is a promising prospect for clinical microbiome studies due to its lysis efficiency. In addition to lysis efficiency the system also provides improved safety due to the fact that lysis of clinical specimens take place within a closed container. In its current form though the lysis method is only compatible with short read sequencing because of the sheared nature of the DNA produced and further optimisation will be required before the lysis step can be incorporated into a

portable sequencing technology such as the ONT MinION. In addition, this study also highlights the importance of evaluating the integrity of DNA, as shearing may greatly impact the final composition of the microbiome depending on the sequencing technology employed.

## Declarations

## Funding

This study was performed, and the manuscript compiled with the support of funding provided by the South African Department of Science and Innovation in conjunction with the International *Mouse* Phenotyping Consortium (*IMPC*) and HANKS TB Diagnostics (Pty) Ltd. The authors also acknowledge financial support from the National Research Foundation of South Africa (DST-NRF Innovation Master's Scholarship Grant No.117255). Opinions expressed, findings and concluding statements are those of the authors and are not necessarily attributed to the NRF.

## Nucleotide sequence data

Genbank All data is publicly available from Genbank and the NCBI SRA under Bio projects PRJNA675726 (SRR13014341-SRR13014338) and PRJNA675484 (SRR1301004 -SRR13010037).

## References

1. Behrouzi A, Nafari AH, Siadat SD (2019) The significance of microbiome in personalized medicine. *Clin Transl Med* 13;8(1):16. doi: 10.1186/s40169-019-0232-y.
2. Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E (2018) Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin Microbiol Infect* 24(4):335-341. doi: 10.1016/j.cmi.2017.10.013.
3. Lim MY, Song EJ, Kim SH, Lee J, Nam YD (2018) Comparison of DNA extraction methods for human gut microbial community profiling. *Syst Appl Microbiol* 41(2):151-157. doi: 10.1016/j.syapm.2017.11.008.
4. Chong J, Liu P, Zhou G, Xia J (2020) Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nat Protoc* 15(3):799-821. doi: 10.1038/s41596-019-0264-1.
5. Corcoll N, Österlund T, Sinclair L, Eiler A, Kristiansson E, Backhaus T, Eriksson KM (2017) Comparison of four DNA extraction methods for comprehensive assessment of 16S rRNA bacterial diversity in marine biofilms using high-throughput sequencing. *FEMS Microbiol Lett* 1;364(14). doi: 10.1093/femsle/fnx139.
6. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34(15):2666–2669.

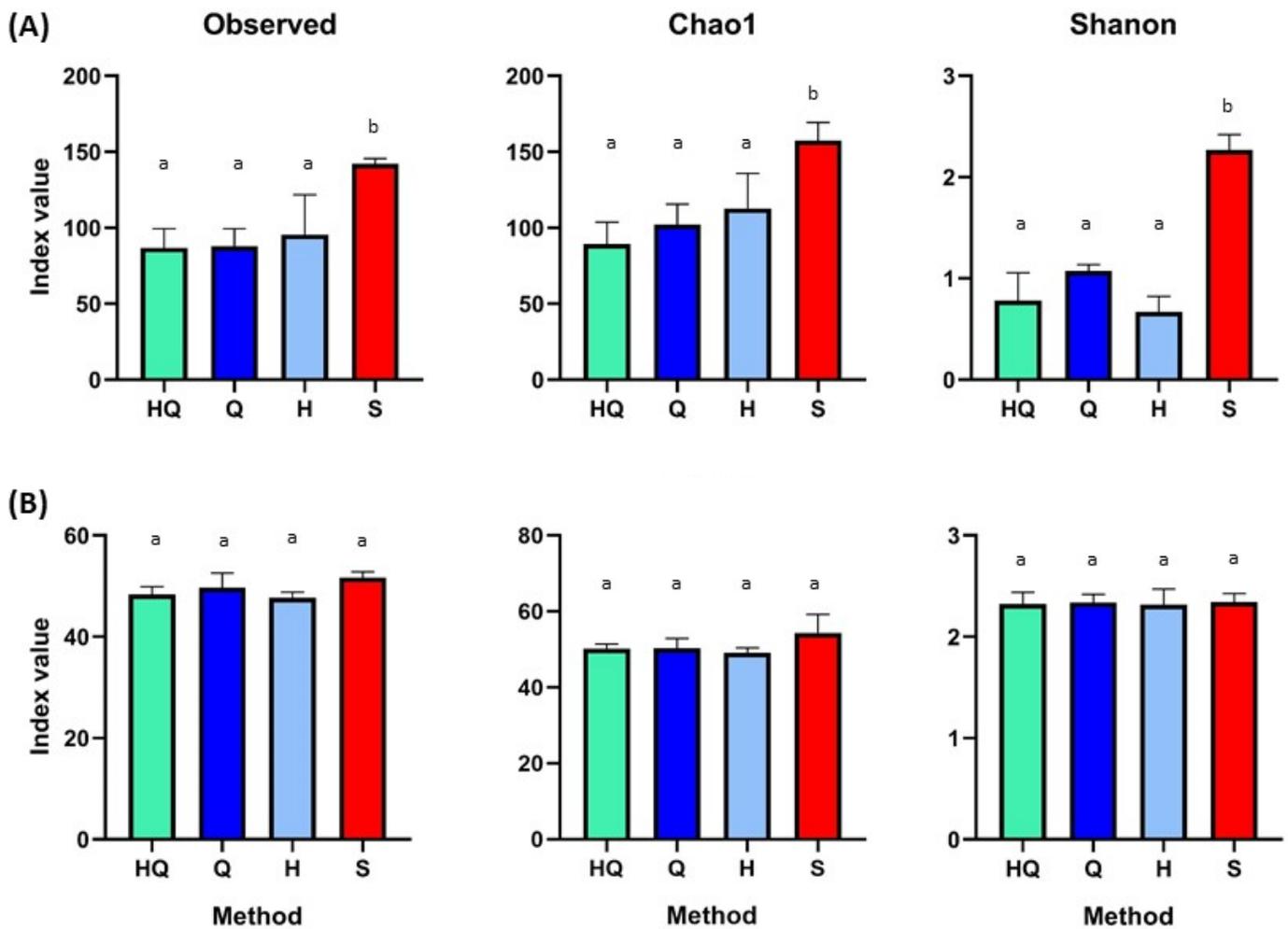
<https://doi.org/10.1093/bioinformatics/bty149>.

7. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J (2017) MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res* 3(45):180-188. doi: 10.1093/nar/gkx295.
8. Di Resta C, Galbiati S, Carrera P, Ferrari M (2018) Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities. *EJIFCC* 29(1):4-14.
9. Edgar RC (2018) Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *PeerJ* 6. <https://doi.org/10.7717/peerj.4652>
10. Fiedorová K, Radvanský M, Němcová E, Grombířiková H, Bosák J, Černochová M, Lexa M, Šmajš D, Freiberger T (2019) The Impact of DNA Extraction Methods on Stool Bacterial and Fungal Microbiota Community Recovery. *Front Microbiol* 17;10:821. doi: 10.3389/fmicb.2019.00821.
11. Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, Louis P, Thomson JM; UK IBD Genetics Consortium, Satsangi J, Flint HJ, Parkhill J, Lees CW, Hold GL (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* 24;9(2):e88982. doi: 10.1371/journal.pone.0088982.
12. Lucena-Aguilar G, Sánchez-López AM, Barberán-Aceituno C, Carrillo-Ávila JA, López-Guerrero JA, Aguilar-Quesada R (2016) DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. *Biopreserv Biobank* 14(4):264-70. doi: 10.1089/bio.2015.0064.
13. Mohajeri MH, Brummer RJM, Rastall RA, Weersma RK, Harmsen HJM, Faas M, Eggersdorfer M (2018) The role of the microbiome for human health: from basic science to clinical applications. *Eur J Nutr* 57:1-14. doi: 10.1007/s00394-018-1703-4.
14. Mutingwende I, Vermeulen U, Steyn F, Viljoen H, Grobler A (2015) Development and evaluation of a rapid multiplex-PCR based system for *Mycobacterium tuberculosis* diagnosis using sputum samples. *J Microbiol Methods* 116:37-43. doi: 10.1016/j.mimet.2015.06.007.
15. Pollock J, Glendinning L, Wisedchanwet T, Watson M (2018) The Madness of Microbiome: Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies. *Appl Environ Microbiol* 84(7). <https://doi.org/10.1128/AEM.02627-17>.
16. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75(23):7537-41. doi: 10.1128/AEM.01541-09.
17. Teng F, Darveekaran Nair SS, Zhu P, Li S, Huang X, Li J, Xu J, Yang F (2018) Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci Rep* 8(1):16321. <https://doi.org/10.1038/s41598-018-34294-x>.
18. Videvall E, Strandh M, Engelbrecht A, Cloete S, Cornwallis CK (2017) Direct PCR Offers a Fast and Reliable Alternative to Conventional DNA Isolation Methods for Gut Microbiomes. *mSystems*. 2(6).

<https://doi.org/10.1128/mSystems.00132-17>

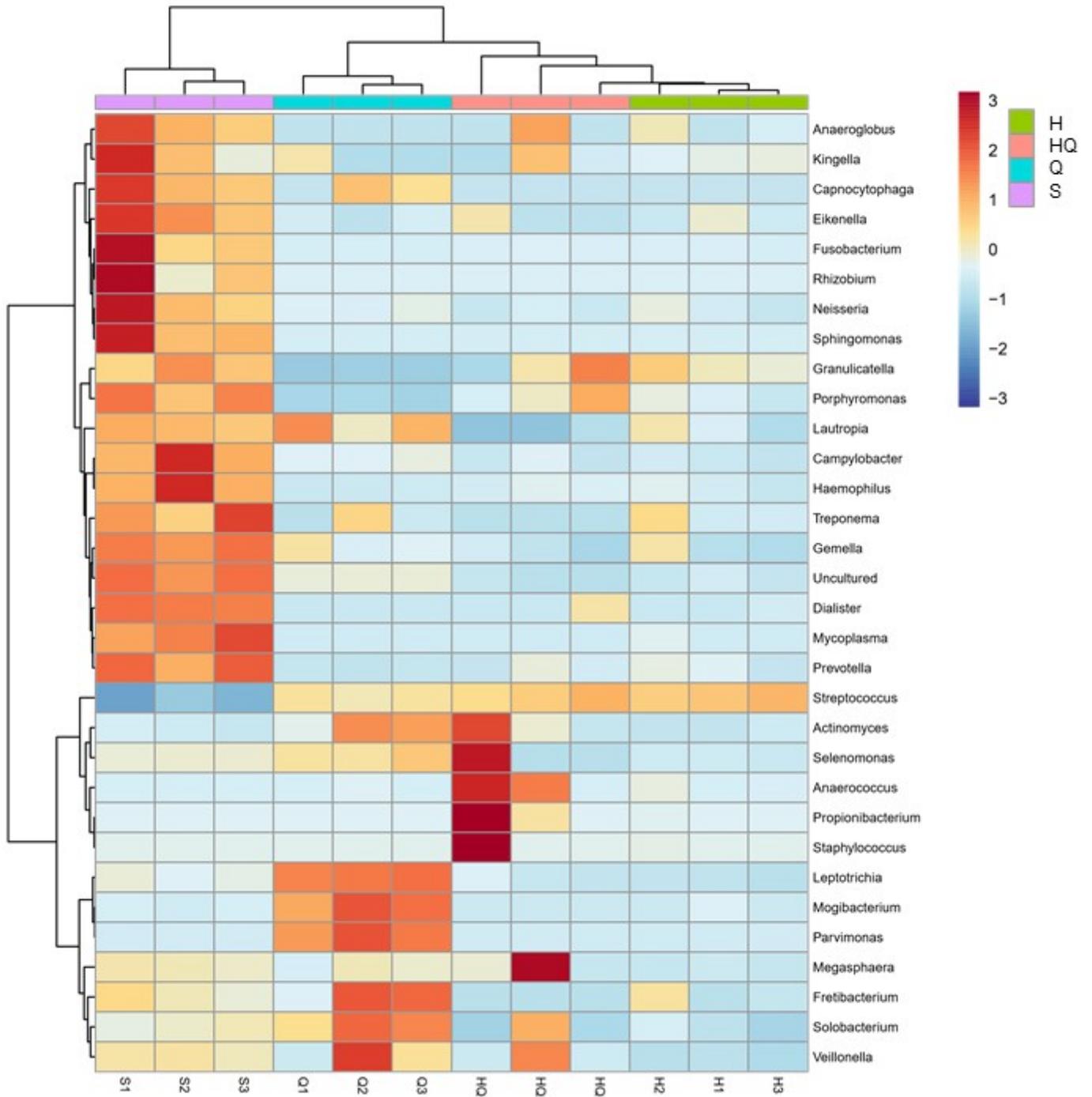
19. Vrijenhoek T, Kraaijeveld K, Elferink M, de Ligt J, Kranendonk E, Santen G, Nijman IJ, Butler D, Claes G, Costessi A, Dorlijn W, van Eyndhoven W, Halley DJ, van den Hout MC, van Hove S, Johansson LF, Jongbloed JD, Kamps R, Kockx CE, de Koning B, Kriek M, Lekanne Dit Deprez R, Lunstroo H, Mannens M, Mook OR, Nelen M, Ploem C, Rijnen M, Saris JJ, Sinke R, Sistermans E, van Slegtenhorst M, Sleutels F, van der Stoep N, van Tienhoven M, Vermaat M, Vogel M, Waisfisz Q, Marjan Weiss J, van den Wijngaard A, van Workum W, Ijntema H, van der Zwaag B, van IJcken WF, den Dunnen J, Veltman JA, Hennekam R, Cuppen E (2015) Next-generation sequencing-based genome diagnostics across clinical genetics centers: implementation choices and their effects. *Eur J Hum Genet* 23(9):1142-50. doi: 10.1038/ejhg.2014.279
20. Wagner Mackenzie B, Waite DW, Taylor MW (2015) Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Front Microbiol* 18(6):130. doi: 10.3389/fmicb.2015.00130.
21. Xia Y, Sun J (2017) Hypothesis Testing and Statistical Analysis of Microbiome. *Genes Dis* 4(3):138-148. doi: 10.1016/j.gendis.2017.06.001.

## Figures



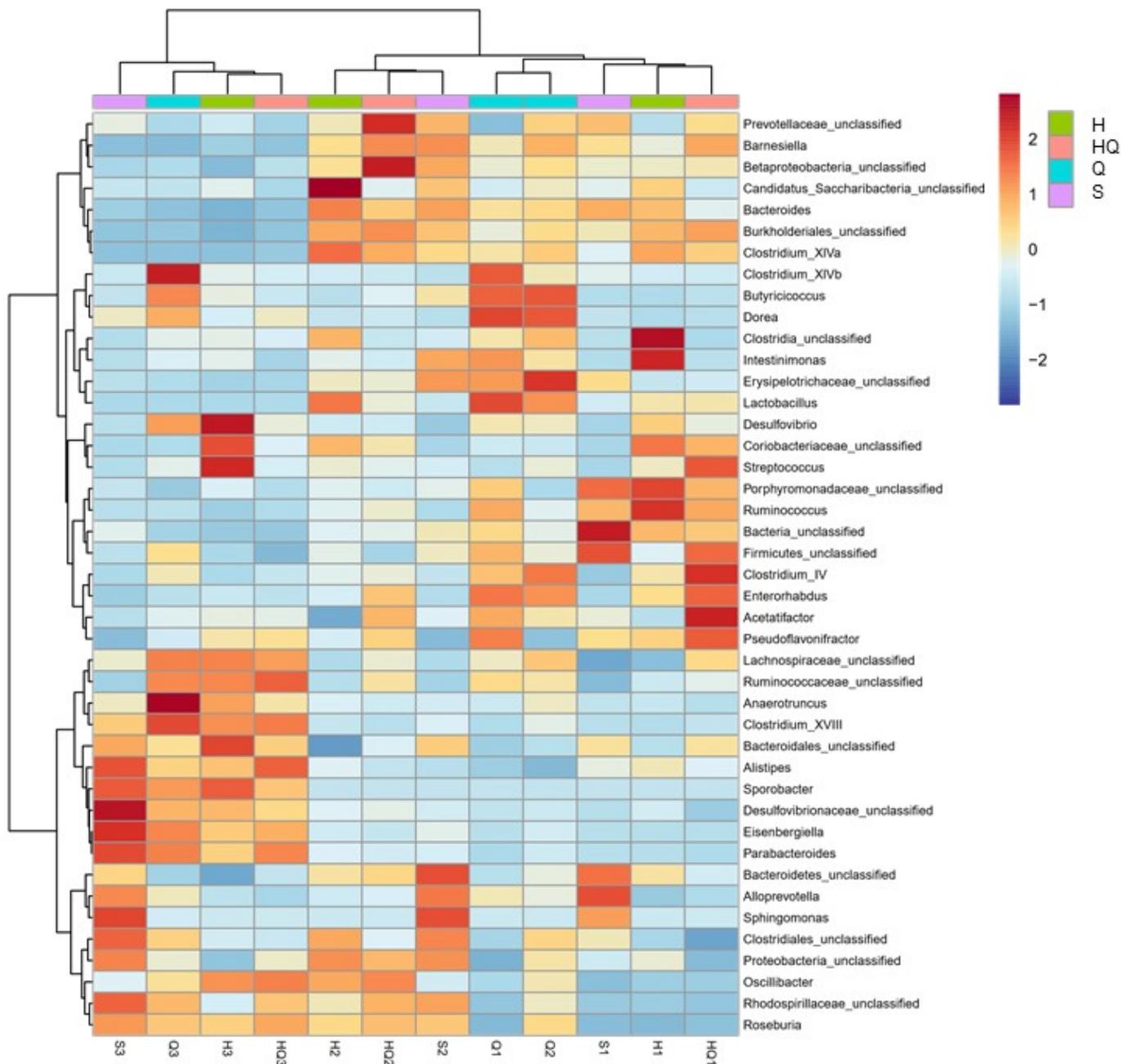
**Figure 1**

Alpha-diversity, measured by Shannon, Chao1 and observed diversity indices for the selected clinical sputum sample (A) and the representative gut sample (B). Error bars show the standard error of the mean (SEM). Methods sharing a common lower case letter belong to the same group according to one way ANOVA with Tukey's post-hoc test for multiple pairwise comparisons. These values do not differ significantly according to the test ( $p > 0.05$ ). Values not sharing a common lower case letter indicate a statistically significant difference in comparison to a value belonging to other groups ( $p < 0.05$ ).



**Figure 2**

Hierarchical clustering heat map based on the relative abundance of the most abundant genera classified to the genus level for the selected clinical sputum sample. Individual cells in the heat map are colour-coded according to the row Z-scores



**Figure 3**

Hierarchical clustering heat map based on the relative abundance of the most abundant genera classified to the genus level from the representative gut sample. Individual cells in the heat map are colour-coded according to the row Z-scores.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S1.png](#)
- [S2.jpeg](#)

- S3.jpeg