



# Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences

Brett Wagner Mackenzie, David W. Waite and Michael W. Taylor\*

Centre for Microbial Innovation & Centre for Brain Research, School of Biological Sciences, University of Auckland, Auckland, New Zealand

## Edited by:

David Berry, University of Vienna, Austria

## Reviewed by:

Ines Martinez, University of Nebraska - Lincoln, USA  
Christina A. Kellogg, United States Geological Survey, USA

## \*Correspondence:

Michael W. Taylor, School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand  
e-mail: mwt.taylor@auckland.ac.nz

The human gut contains dense and diverse microbial communities which have profound influences on human health. Gaining meaningful insights into these communities requires provision of high quality microbial nucleic acids from human fecal samples, as well as an understanding of the sources of variation and their impacts on the experimental model. We present here a systematic analysis of commonly used microbial DNA extraction methods, and identify significant sources of variation. Five extraction methods (Human Microbiome Project protocol, MoBio PowerSoil DNA Isolation Kit, QIAamp DNA Stool Mini Kit, ZR Fecal DNA MiniPrep, phenol:chloroform-based DNA isolation) were evaluated based on the following criteria: DNA yield, quality and integrity, and microbial community structure based on Illumina amplicon sequencing of the V4 region of bacterial and archaeal 16S rRNA genes. Our results indicate that the largest portion of variation within the model was attributed to differences between subjects (biological variation), with a smaller proportion of variation associated with DNA extraction method (technical variation) and intra-subject variation. A comprehensive understanding of the potential impact of technical variation on the human gut microbiota will help limit preventable bias, enabling more accurate diversity estimates.

**Keywords:** human gut microbiota, nucleic acids extraction, beta diversity, PERMANOVA

## INTRODUCTION

The human gut harbors the most substantial microbial communities within our bodies, with these communities exhibiting considerable inter- and intra-personal variability (Eckburg et al., 2005; Ley et al., 2006; Qin et al., 2010). Several diseases and disorders have been linked to dysbiosis (imbalance) in these gut communities, and recent studies have sought to identify changes to microbial community structure and function during health and disease (Bäckhed et al., 2004; Cantarel et al., 2012; Claesson et al., 2012; Qin et al., 2012; Hsiao et al., 2013; Gevers et al., 2014). Gut microbial community composition varies less within an individual than among different individuals, suggesting a strong environmental component (Turnbaugh et al., 2006; Qin et al., 2010; Caporaso et al., 2011; The Human Microbiome Project Consortium, 2012; Schloissnig et al., 2013).

Nucleic acids extraction from stool samples is the first step in describing microbial diversity using culture-independent methods. Differences in lysis efficiency, heterogeneous distribution of microbes across a sample, and adherence of microbes to the stool matrix can result in preferential lysis of certain cell types and misrepresentation of microbial community diversity (Li et al., 2003; Ariefdjohan et al., 2010; Maukonen et al., 2012). DNA extraction methods utilize different lysis procedures, such as mechanical (bead beating), chemical, enzymatic, and heat, and various method comparison studies have reported contradictory results regarding the most effective lysing procedure (Carrigg et al., 2007; Dridi et al., 2009; Maukonen et al., 2012; Yuan et al., 2012; Claassen et al., 2013). Methodological, or technical, variation

such as that due to different DNA extraction techniques, sequencing platforms, and/or sequence processing pipelines can influence descriptions of ecological diversity and observed biological variation, including inter-subject variation.

Although many studies have compared DNA extraction protocols for the gut microbiota (McOrist et al., 2002; Li et al., 2003; Yu and Morrison, 2004; Nechvatal et al., 2008; Lauber et al., 2010; Bahl et al., 2012; Maukonen et al., 2012; Yuan et al., 2012; Claassen et al., 2013; Henderson et al., 2013; Peng et al., 2013; Kennedy et al., 2014), a consensus as to the most efficient extraction method, which is most representative of gut microbial community diversity from stool samples, has not yet been reached. It is crucial to continually assess nucleic acids extraction methods, identifying those that are the most efficient, accurate and reproducible with the overall aim to standardize methods across gut microbiology studies, limiting preventable bias due to technical variation. Although a recent meta-analysis of stool samples showed strong clustering by experimental protocol, namely choice of PCR primer (Lozupone et al., 2013), the individual impacts from the sources of variation, including inter- and intra-subject variation and laboratory technique, have not been previously quantified.

This study examined multiple stool samples from three age- and sex-matched individuals using several DNA extraction methods to identify and quantify the degree to which choice of DNA extraction method (technical variation) impacted upon inter-subject variation (biological variation). In addition, data from an existing study of the human fecal microbiota (Yatsunenکو et al.,

2012) were included in order to further assess our predictions about inter-subject variability.

## MATERIALS AND METHODS

### SUBJECT SELECTION AND STOOL SAMPLE COLLECTION

Subjects for this study were chosen from a database of self-registered volunteers. Ethics approval for this project was granted by the Northern B Health and Disability Ethics Committee (Ref. 12/NTB/59). Three female subjects in their mid-20s (Subjects 1, 2, and 3), with no dietary restrictions, gastrointestinal disturbances or recent antibiotic usage (>6 months) were chosen for this study. All subjects were located within Auckland, New Zealand and consented to supply three entire stool samples at 48 h intervals into sterile polypropylene containers. Stool samples were immediately stored on ice until processing of the samples in the lab. Pre-processing of samples into smaller sub-samples occurred within 12 h of collection and these were stored at -80°C until microbial DNA was extracted. Samples for the Human Microbiome Project (HMP) extraction method were processed first, before homogenization, according to the HMP Manual of Procedures 2010 Version 11.0 (McInnes and Cutting, 2010). After sub-samples were processed according to the HMP protocol, the remaining stool sample was homogenized via stirring with a sterile metal spatula and sample sizes were standardized to 200 mg, enabling better comparisons between the DNA extraction methods.

### DNA EXTRACTION

DNA was extracted in triplicate from each stool sample using five commonly used microbial DNA extraction methods, namely the Human Microbiome Project extraction method (HMP), MoBio PowerSoil® DNA Isolation Kit (M), QIAamp® DNA Stool Mini Kit (Q), ZR Fecal DNA MiniPrep™ (Z), and one non-kit phenol:chloroform-based DNA isolation protocol (P) (Zoetendal et al., 2006) (Table 1). All bead-beating steps were performed in the Qiagen TissueLyser II at a frequency of 30 Hz for 2 min and centrifugation steps were carried out in an Eppendorf 5415D centrifuge.

#### The Human Microbiome Project Consortium Extraction Method

The HMP method uses the MoBio PowerSoil® DNA Isolation Kit, but includes a pre-processing protocol, in which 1 mL of supernatant from a centrifuged mixture of 2 mL stool sample

and 5 mL MoBio lysis buffer, was added to MoBio bead tubes for two 10 min heating steps at 65°C then 95°C prior to freezing at -80°C (McInnes and Cutting, 2010). The only subsequent deviation from the standard MoBio PowerSoil® DNA Isolation Kit protocol was a longer centrifugation step after the addition of Inhibitor Removal Technology® Solution C3 to precipitate a DNA pellet from any non-DNA organic and inorganic material, including humic acids, cell debris, and proteins. All other extraction steps were followed according to the manufacturer's instructions.

#### MoBio PowerSoil® DNA Isolation Kit

Samples were thawed on ice and added to the PowerBead Tubes provided in the kit using a sterile metal spatula. The MoBio PowerSoil® DNA Isolation Kit involves mechanical lysis of cells during a bead-beating step. Humic substances were then removed using patented Inhibitor Removal Technology. A salt solution was added to help the DNA bind to the silica spin column filter before the DNA was washed with an ethanol-based solution to remove residual salt, humic acids or any other contaminants. Last, a sterile elution buffer (10 mM Tris) released the DNA from the spin column filter, yielding DNA that is ready for downstream applications.

#### QIAamp® DNA Stool Mini Kit

After samples were thawed on ice, Buffer ASL was added to help lyse bacterial cells and samples were placed in the Tissue Lyser II for 2 min at 30 Hz. The homogenate was then incubated in a 70°C water bath for 5 min. Stool particles were pelleted, then an InhibitEX Tablet was added to the supernatant to adsorb inhibitors. Proteinase K and Buffer AL were added to the supernatant to digest proteins. The DNA was bound to a spin column filter and impurities were washed from the sample using 96–100% ethanol and proprietary Buffer AW2. The extracted microbial DNA was eluted with 200 µL Buffer AE.

#### ZR Fecal DNA MiniPrep™ Kit

ZR Fecal DNA MiniPrep™ is a bead-beating and spin-column filter extraction kit. Similar to other bead-beating kits, a lysis solution was added to the sample in the ZR BashingBead™ Lysis Tube to help lyse bacterial cells. Fecal DNA Binding Buffer and centrifugation in the spin column tube then bound the extracted DNA to the spin filter. The bound DNA was washed, and an extra elution step resulted in twice-filtered DNA.

**Table 1 | Comparison of the five microbial DNA extraction methods used in this study.**

Extraction method	Kit abbreviation	Recommended fecal starting amount (mg)	Lysis Type	Elution volume (µL)	DNA Isolation
Human Microbiome Project Extraction Method	HMP	1 mL supernatant	Heat, Mechanical	100	Spin column
MoBio PowerSoil® DNA Isolation Kit	M	250	Mechanical	100	Spin column
Qiagen QIAamp® DNA Stool Mini Kit	Q	180–220	Heat, Chemical, Enzymatic	200	Spin column
Zymo ZR Fecal DNA MiniPrep™	Z	150	Mechanical	100	Spin column
Phenol: chloroform-based DNA isolation	P	200	Mechanical	100	Phase separation

A total of 135 samples were analyzed from 5 extraction methods, comprising 3 sub-samples from each of 3 entire stool samples from 3 subjects.

### Phenol:chloroform-based DNA isolation

This protocol for DNA extraction by phase separation was adapted from Zoetendal et al. (2006). Samples were thawed on ice and added to 2 mL tubes containing 0.3 g of 0.1 mm diameter silica zirconia beads. Buffer-saturated phenol was added to the tube containing the stool sample and beads, and homogenized in the Tissue Lyser II for 2 min at 30 Hz. The homogenized sample was briefly cooled on ice before adding chloroform and isoamyl alcohol (24:1). A 3 M sodium acetate salt solution was used to bind the extracted DNA from the upper aqueous phase, while 95% ethanol washed away impurities. The resulting DNA pellet was dried on the bench top, then rehydrated in 100  $\mu$ L TE Buffer.

### EVALUATION OF DNA YIELD, QUALITY AND INTEGRITY

Final yield and quality of extracted DNA were determined spectrophotometrically using the NanoDrop<sup>®</sup> ND-1000 (NanoDrop Technologies Inc., Wilmington, USA). DNA yield was also determined fluorometrically using the Broad Range (BR) kit on the Qubit<sup>®</sup> Fluorometer 1.0 (Invitrogen Co., Carlsbad, USA). Pure genomic DNA is indicated by an A260/A280 nm ratio between 1.8 and 2.0. Integrity of genomic DNA was determined by visualizing 3  $\mu$ L of extracted DNA on a 1% agarose gel (w/v) containing SYBR Safe DNA Gel Stain (Invitrogen Co., Carlsbad, USA) run in 0.5X TBE buffer at 100 V for 45 min. All values were normalized to 200 mg starting material and 100  $\mu$ L elution volume, to allow for accurate comparisons between methods.

### STATISTICAL ANALYSES OF DNA YIELD AND QUALITY DATA

Statistical analyses, including coefficient of variation (CV) tests for reproducibility, Shapiro-Wilk test for normality, Kruskal-Wallis group test, and Mann-Whitney U pairwise tests with Benjamini-Hochberg (“BH”) *p*-value adjustment for multiple pairwise tests (Benjamini and Hochberg, 1995), were conducted on quantitative and qualitative data. Mean values for yield and quality were determined. All quantitative and qualitative statistical analyses were conducted in “R” version 2.15.0 (R Development Core Team, 2012).

### PCR AMPLIFICATION AND ILLUMINA MISEQ PREPARATION

Extracted DNA was diluted to 5 ng/ $\mu$ L in PCR-grade water for PCR amplification targeting the hypervariable V4 region of the bacterial and archaeal 16S rRNA genes (Caporaso et al., 2010c; Klindworth et al., 2013). PCR was performed using an Applied Biosystems<sup>®</sup> GeneAmp<sup>®</sup> PCR System 9700. Each PCR reaction contained: 2.5  $\mu$ L 10X High Fidelity PCR Buffer, 1  $\mu$ L 50 mM magnesium sulfate, 0.5  $\mu$ L 2.5 mM dNTPs, 0.1  $\mu$ L Platinum<sup>®</sup> Taq DNA Polymerase High Fidelity, 18.4  $\mu$ L PCR-grade autoclaved water, and 0.5  $\mu$ L 10  $\mu$ M 515 F primer (5′—AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT ATG GTA ATT GTG TGC CAG CMG CCG CGG TAA—3′ Caporaso et al., 2010c) composed of an Illumina adapter, a forward primer pad, a two-base linker sequence (“GT”) and the 16S rRNA gene-targeting primer (underlined). To each PCR reaction, 1  $\mu$ L of each DNA template was added together with 1  $\mu$ L of 5  $\mu$ M unique 806R-barcode reverse fusion primers (5′—CAA GCA GAA GAC GGC ATA CGA GAT NNNNNNNNNNNN GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TGG ACT ACH VGG GTW

TCT AAT—3′ Klindworth et al., 2013), composed of the Illumina adapter, a unique 12-base error-correcting Golay barcode, the reverse primer pad, a two-base linker sequence (“CC”) and the 806R primer (underlined). A positive control of *Escherichia coli* genomic DNA and a negative control of 1  $\mu$ L randomly selected reverse primer with PCR-grade autoclaved water was performed for each PCR. Thermocycling conditions were: initial denaturation at 94°C for 3 min followed by 30 cycles of denaturation at 94°C for 45 s, annealing at 55°C for 45 s, and extension at 72°C for 90 s. A final extension step was performed at 72°C for 10 min.

Amplifications were completed in triplicate and 20  $\mu$ L aliquots of each sample from the three PCR amplifications were pooled. After checking amplified PCR products on a 1% agarose gel (w/v) containing SYBR Safe DNA Gel Stain, pooled products were purified using the MoBio UltraClean<sup>®</sup>-htp 96 Well PCR Clean-Up Kit (MoBio Laboratories, Carlsbad, CA, USA). The manufacturer’s protocol was followed and the only deviation was to increase the centrifugation time to 4 min to achieve the appropriate force in the centrifuge. All purified samples were measured fluorometrically using the High Sensitivity (HS) kit on the Qubit<sup>®</sup> Fluorometer 1.0 (Invitrogen Co., Carlsbad, USA). Impurities and contaminants were assessed for a random selection of samples using an Agilent DNA 1000 chip (Agilent Technologies, Inc., Waldbronn, Germany).

Purified PCR products were diluted in 10 mM Tris to achieve equimolar concentrations across all samples and 2  $\mu$ L of each sample was combined into a sterile microcentrifuge tube. Illumina MiSeq 2  $\times$  250 bp paired-end sequencing was performed by the Centre for Genomics, Proteomics and Metabolomics through NZ Genomics Ltd at the University of Auckland. Sequence data were uploaded to the NCBI Sequence Read Archive, project number SRP051334.

### SEQUENCE DATA PROCESSING

Forward and reverse paired-end sequence reads were merged according to the fastq-join parameter (Aronesty, 2011) within the join\_paired\_ends.py command in QIIME (Caporaso et al., 2010b). The unique\_seqs command in the mothur software identified unique sequences (Schloss et al., 2009). Abundance data were appended to unique sequences and operational taxonomic units (OTUs) were constructed using the UPARSE pipeline, based on the program usearch (Edgar, 2013). Briefly, individual OTUs were constructed by binning sequences into clusters of greater than 97% sequence similarity (-cluster\_otus, -minsize 1), discarding putative chimeric OTUs in the process. A representative sequence of each OTU was further tested for chimeric artifacts using the SILVA reference database provided as part of the UPARSE pipeline (-uchime\_ref). Abundance data were then reincorporated into the dataset by mapping the initial sequences against the representative OTUs (-usearch\_global, -id 0.97). The resulting table was converted to biom format for analysis in QIIME using the inbuilt “convert” function of the biom software package.

Rarefaction was performed using the command alpha\_rarefaction.py as part of the calculate\_core\_analyses.py command within QIIME 1.8 (Caporaso et al., 2010b). A total of 100 permutations were performed at 10 equal intervals

between the minimum (1) and maximum (11,739) sequence depths. Alpha diversity was estimated for each DNA extraction method using the “observed species” metric within QIIME. Shannon and Simpson diversity indices measuring richness and evenness, were used to estimate diversity within each of the DNA extraction methods. Taxonomy was assigned to each sequence using uclust consensus taxonomy assigner, based on 97% sequence similarity with the Greengenes reference database version 13.5 (DeSantis et al., 2006; Caporaso et al., 2010a). A phylogenetic tree was built using FastTree 2.1.3 (Price et al., 2009) and weighted UniFrac, unweighted UniFrac, and Bray-Curtis dissimilarity matrices were generated to perform beta diversity measures on the data set (Lozupone and Knight, 2005). All three beta diversity measures returned overall similar results, and we chose to use the unweighted UniFrac metric to compare beta diversity, because of its previous successful application distinguishing microbial communities within the human microbiome (Lozupone and Knight, 2005; Wu et al., 2011; Lozupone et al., 2013). Unweighted UniFrac PCoA biplots were visualized in the EMPeror Visualization Program (Vázquez-Baeza et al., 2013).

### STATISTICAL ANALYSES OF SEQUENCING DATA

To assess which bacteria are driving the differences between Subjects and DNA extraction methods, paired, two-tailed *t*-tests, with Bonferroni adjustment for multiple pairwise comparisons, were conducted between all Subjects and DNA extraction methods for taxon-assigned OTUs.

Permutational multivariate analysis of variance (PERMANOVA) statistical analyses and pairwise tests were conducted in PERMANOVA+ in PRIMER v6 software (Anderson et al., 2008). The PERMANOVA test allows robust, unbiased analysis of multivariate data based on complex experimental designs and models (Anderson et al., 2008). PERMANOVA analyses return a *p*-value for significance and also the  $R^2$  value, which is indicative of the amount of variation attributed to a specific treatment within a model. PERMANOVA analysis was conducted on the unweighted UniFrac matrix, and values were obtained using type III sums of squares with 9999 permutations of residuals under a reduced model.

### COMBINED SEQUENCE DATA ANALYSIS

We combined our raw sequencing data with the raw sequencing reads from an international study conducted by Yatsunenko et al. (2012) (hereafter referred to as the Yatsunenko data) to examine predictions regarding inter-subject variation within our cohort. The same V4 region of the 16S rRNA gene was amplified in the Yatsunenko data as in our study, and our raw sequencing reads were reprocessed with the raw reads from the Yatsunenko data in accordance with the protocol used in their study. Due to the documented impact of subject age on the human gut microbiota (Lozupone et al., 2012; Yatsunenko et al., 2012), samples obtained from infants or children <3 years of age (as described in the Yatsunenko metadata) were removed from the data set. The combined data set sequences were quality filtered in QIIME, and reads <90 bp were eliminated. Representative sequences from the Yatsunenko data were aligned against the SILVA seed database and a lane mask constructed to represent the base positions covered

by the first 90 nucleotides of these sequences. Our own sequences were then aligned and filtered using this lane mask, resulting in a standardization of sequence lengths. All sequences were then dereplicated in mothur, and usearch was used for closed-reference OTU picking with the Greengenes database (May 2013 release). All sample depths were normalized to 10,303 sequences per sample within the combined data sets. Taxonomy assignments and phylogenetic trees were inferred from the reference OTUs.

Two dimensional NMDS plots were generated in R (version 3.1.2) using the package vegan (version 2.2–1) (Oksanen et al., 2015) and the influence of bacterial families on the ordination was tested using the envfit function. Vectors with a statistically significant contribution to the ordination were identified following Benjamini-Hochberg (“BH”) false discovery rate correction ( $p = 0.05$ ) and overlaid onto the ordination space. The size of taxa nodes is based upon the average abundance of the taxon in the communities. Analysis of similarity (ANOSIM) was used to identify changes in microbial community structures between countries (Clarke, 1993). ANOSIM tests between countries were performed in mothur, using 10,000 permutations to determine significance and “BH” correction applied to all *p*-values.

## RESULTS

### INFLUENCE OF DNA EXTRACTION METHOD ON YIELD AND QUALITY OF EXTRACTED DNA

Mann-Whitney U pairwise tests, with “BH” *p*-value adjustment, suggest that the only pairwise comparison that did not produce significant differences for DNA yield and quality was that comparing the M and P methods ( $p = 0.12$ ). The P method exhibited the highest median yield (59.15  $\mu\text{g/mL}$ ,  $p < 0.002$ ), but was not significantly different when compared to the median yield from the MoBio PowerSoil® DNA Isolation Kit ( $p = 0.12$ , median yield  $M = 45.4 \mu\text{g/mL}$ ). The lowest median yield ( $Z = 2.62 \mu\text{g/mL}$ ,  $p < 0.001$ ), as well as the lowest A260/A280 ratios ( $Z = 1.06$ ,  $p < 0.001$ ), resulted from use of method Z (Figure 1). The median value of A260/A280 ratios from M were the closest to 1.8, indicating pure, high quality DNA (HMP = 1.88,  $M = 1.79$ ,  $Q = 2.03$ ,  $Z = 1.06$ ,  $P = 1.75$ ).

Analysis of agarose gel images revealed very faint bands from Z extractions (Figure S1). These results are in agreement with the consistently low yield measurements obtained for this kit. Method P yielded a high amount of sheared DNA, visualized by the strong smear toward the bottom of the gel. Methods HMP, M, and Q showed consistently strong bands at the top of the gel, indicating high molecular weight DNA. The Human Microbiome Project gel results indicated the least amount of shearing of all the methods.

Only two kits, M and Q, proved reproducible for DNA yield with coefficient of variation (CV) values < 1 ( $M = 0.42$ ,  $Q = 0.53$ ). However, the median A260/A280 ratios recovered by kit Q were greater than 2.0, which is outside the range of pure DNA as indicated by the Nano Drop® ND-1000 spectrophotometer manual. All methods were reproducible for DNA quality (CV values: HMP = 0.09,  $M = 0.03$ ,  $Q = 0.11$ ,  $Z = 0.12$ ,  $P = 0.07$ ).

### INITIAL SEQUENCE DATA ANALYSIS RESULTS

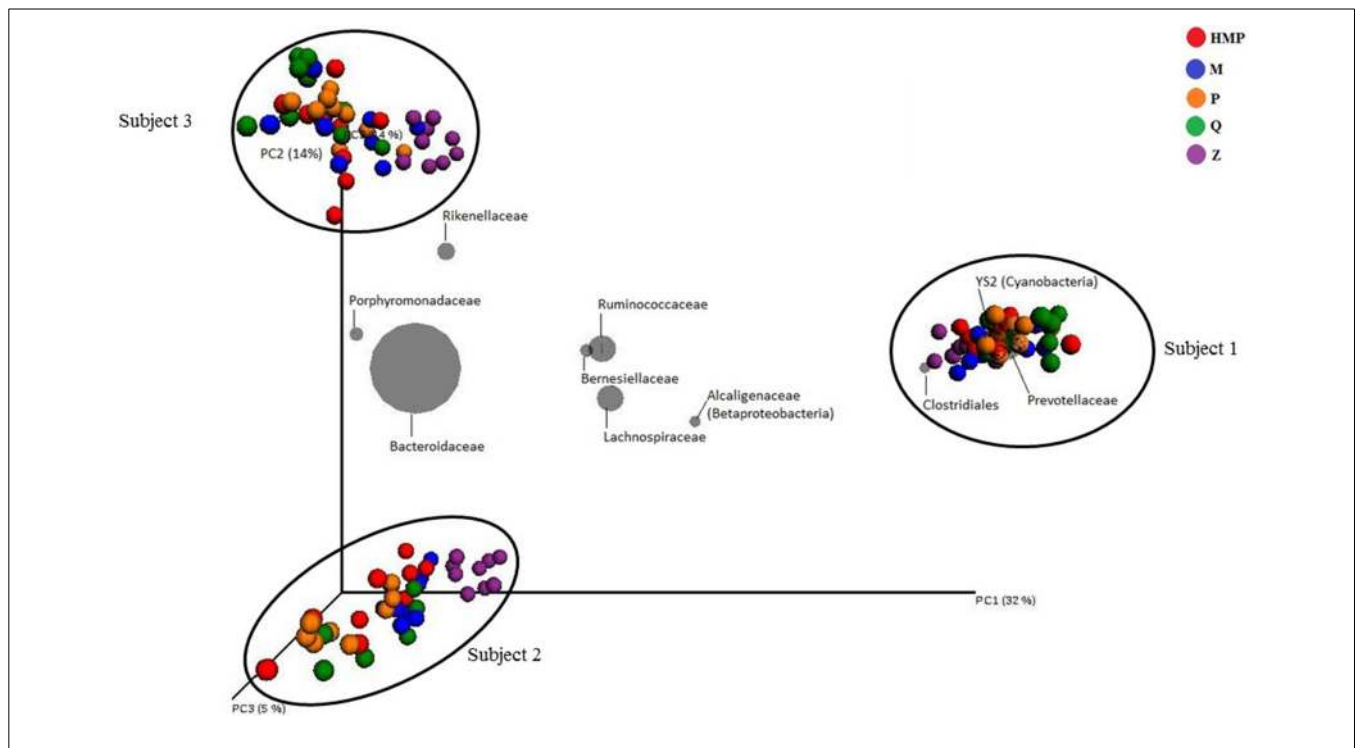
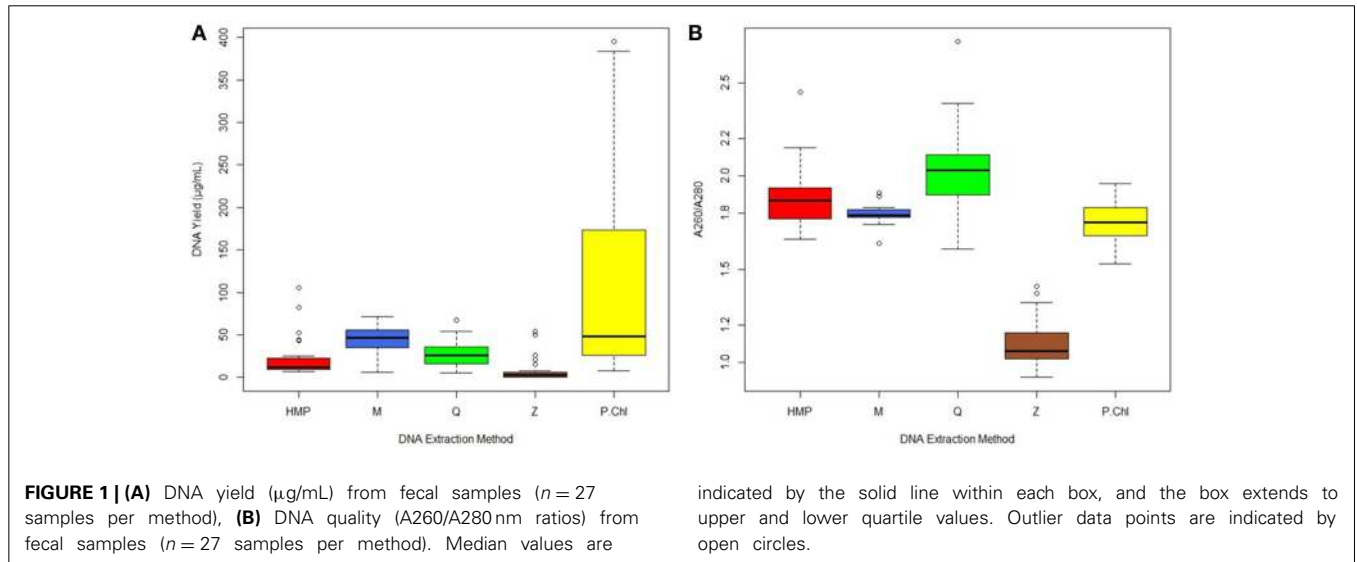
A total of 135 samples were sequenced ( $n = 45$  per subject,  $n = 27$  per extraction method), with quality filtered (q score 30)

reads resulting in 9.74 million sequencing reads with an average length of 253 bp. After removing chimeras (4.7% of the unique sequences), *de novo* OTU picking returned 5403 97%-OTUs.

**INTER-SUBJECT AND DNA EXTRACTION METHOD DIFFERENCES WERE IDENTIFIED AS MAJOR SOURCES OF VARIATION BY PERMANOVA**

PERMANOVA analysis of microbial community profiles rarefied to 11,739 sequences per sample for unweighted UniFrac revealed that the largest portion of variation could be attributed

to inter-subject differences (Figure 2, Table S1). PERMANOVA analysis of weighted UniFrac and Bray-Curtis dissimilarity matrices exhibited similar overall results to those from unweighted UniFrac (results not shown). Inter-subject differences, then extraction method, and any combination thereof, all contributed significantly to observed total variation. The number of samples from individual subjects (intra-subject variation), and method combined with sample number, were not significant contributors to overall variation, suggesting that one sample



**FIGURE 2 | Relative unweighted UniFrac phylogenetic distances between subjects imaged using a biplot.** Superimposed on the PCoA plot are gray spheres indicating the most abundant bacterial families. The sizes of

the spheres represent the mean relative abundance of the respective taxon and the location of the spheres within the plot indicate subject-specific associations. Samples within subjects are colored by extraction method used.

per subject is sufficient for cross-sectional studies. Inter-subject differences explained 34% of the variation within the model ( $R^2 = 0.34$ ,  $p = 0.0001$ ), with extraction method alone explaining 9% of variation ( $R^2 = 0.09$ ,  $p = 0.0006$ ). Subject combined with extraction method explained an additional 9% of variation ( $R^2 = 0.09$ ,  $p = 0.0001$ ), and subject combined with sample number explained 7% ( $R^2 = 0.07$ ,  $p = 0.0001$ ). Variation between samples within each subject (intra-subject variation) did not contribute a significant proportion of variation to the model ( $p > 0.05$ ).

Taxa biplots can be used to visualize and explore the taxonomic factors driving clustering patterns within the PCoA. In a biplot, bacterial taxa are plotted according to weighted average of taxa within all samples, where the weights of the bacterial taxa represent sequence abundance, and the location of the taxa spheres indicate which bacteria are driving clustering patterns. Taxonomic resolution at the family-level best captured the groups of bacteria driving the clustering patterns observed between subjects. The taxa biplots indicate that clustering by Subject is driven by differences in abundance of *Bacteroidaceae*, *Ruminococcaceae*, *Lachnospiraceae*, and *Prevotellaceae* (Figure 2). After primary clustering by Subject, the samples within each subject exhibit secondary clustering due to DNA extraction method. These observations are supported by the PERMANOVA results.

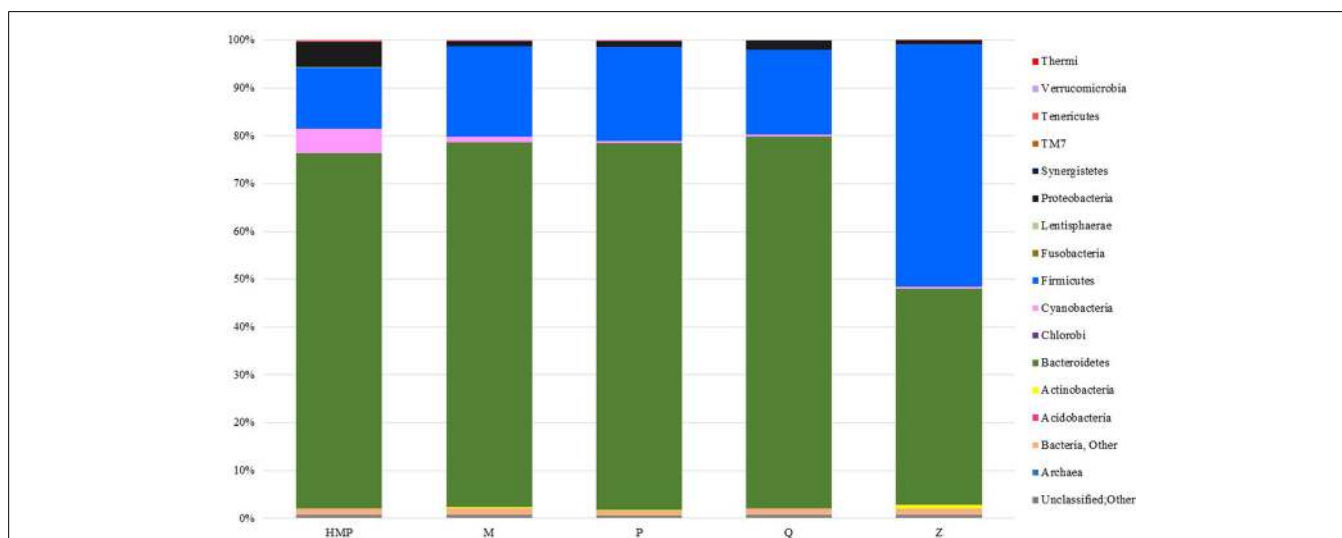
#### EFFECTS OF DNA EXTRACTION METHOD ON MICROBIAL COMMUNITY PROFILES

DNA extraction method accounted for 9% of the variation within this experimental model. PERMANOVA pairwise tests based on unweighted UniFrac metrics revealed significant differences between HMP and Z methods ( $p = 0.044$ ), P and Z methods ( $p = 0.021$ ), and suggested that the largest difference in DNA extraction method was between Z and Q methods ( $p = 0.018$ ). No significant difference between Z and M methods was detected ( $p = 0.113$ ). No other significant differences were detected

between methods for unweighted UniFrac PERMANOVA pairwise analyses.

Members of the bacterial phyla *Bacteroidetes* and *Firmicutes* accounted for most of the taxon-assigned OTUs observed for all five extraction methods, while sequences representing *Actinobacteria*, *Cyanobacteria*, *Lentisphaerae*, *Proteobacteria*, *Tenericutes*, and *Verrucomicrobia* were also recovered using all of the extraction methods (Figure 3, Table S2). Unclassified OTUs (sequences unresolved at domain level) represented 0.58–0.75% and unclassified bacteria represented 1.0–1.3% of the total number of sequences for each of the DNA extraction methods (Table S2). The percentage of sequences represented by each phylum varied between DNA extraction methods. Method Q yielded the largest proportion of *Bacteroidetes* (77.7%), whereas Z had the lowest (45.2%), but the highest proportion of *Firmicutes* (50.8%). The HMP method yielded the lowest proportion of *Firmicutes* (12.8%); however, this method had the highest proportions of *Cyanobacteria* (5.1%) and *Proteobacteria* (5.3%) (Figure 3). Only two methods, HMP and Q, detected *Fusobacteria* (<0.001%), which is interesting since these are the only extraction methods that include a heat lysis component. Phyla that were only detected using protocol Z include *Acidobacteria*, *Chlorobi* and *Thermi*. HMP was the only method to detect members of the candidate phylum TM7.

Multiple pairwise comparisons between DNA extraction methods for observed OTUs richness revealed significant differences between method Z and two other methods, P ( $p = 0.02$ ) and M ( $p = 0.04$ ). Method Z yielded the highest number of observed OTUs, followed by the HMP method, Q, M, and P methods (Table 2). The mean Shannon diversity across all DNA extraction methods was  $5.33 \pm 0.51$  and the mean Simpson diversity was  $0.89 \pm 0.06$ . Method Z had the highest Shannon and Simpson diversity measurements, followed by the HMP method, P, M, and Q methods (Table 2). Pairwise comparisons between the extraction methods for Shannon and



**FIGURE 3 | Taxa plot summarizing the relative abundance of taxon-assigned OTUs identified for bacterial and archaeal phyla in the stool samples taken from each extraction method.** Each method represents sequencing information from 27 samples.

Simpson indices revealed that method Z was the only kit that was significantly different from all other methods ( $p \leq 0.002$ ).

A total of 110 OTUs significantly impacted upon the differences observed between DNA extraction methods ( $p < 0.05$ ), of which the top 52 OTUs were members of the phylum *Firmicutes* (Table S3). Members of the *Firmicutes* family *Lachnospiraceae* were the main drivers of variation, separating Z from the other extraction methods. *Parabacteroides distasonis* sequences were found in significantly higher abundance in both P and Q methods. *Bilophila* sp. was the only member of the phylum *Proteobacteria* that significantly impacted upon variation between methods. Within the phylum *Actinobacteria*, *Bifidobacterium adolescentis* was reported at significantly higher abundances in M and Z methods. *Bacteroides* sp. was also reported at significantly different levels across the DNA extraction methods.

PERMANOVA pairwise comparisons between the microbial community structures obtained using the HMP and MoBio PowerSoil® DNA Isolation Kit protocols did not suggest they were significantly different. However, analysis of the community structures from the phylum- and genus-level taxa plots revealed minor differences in the recovery of Gram-positive and Gram-negative bacteria. The HMP method recovered significantly higher proportions of Gram-negative *Bacteroidetes*, *Cyanobacteria* and *Proteobacteria*. The MoBio PowerSoil® DNA Isolation Kit recovered higher proportions of *Actinobacteria* and *Firmicutes*, which are both Gram-positive taxa. It is unclear why the HMP method, which uses the MoBio PowerSoil® DNA Isolation Kit, reported a higher abundance of Gram-negative bacteria, and consequently, lower abundances of Gram-positive bacteria.

### INTER-SUBJECT VARIATION

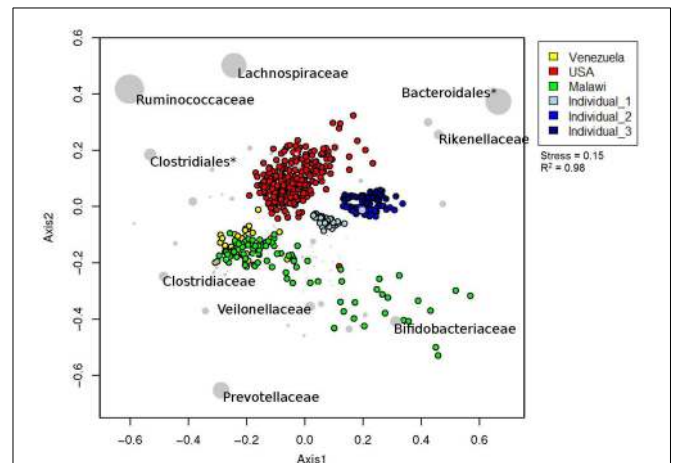
Inter-subject variation contributed the largest proportion of variation (Figure 2, Table S1). Sequencing data from 3 samples, including 45 subsamples, from each subject were combined and are included in analyses of inter-subject variation. Based on relative abundance of taxon-assigned OTUs, members of the bacterial phyla *Bacteroidetes* (Subject 1: 63.4, Subject 2: 73.3, and Subject 3: 75.3%) and *Firmicutes* (26.1, 21.7, and 21.7%) comprised the majority of sequences from each subject's gut microbiota.

A total of 915 OTUs were significantly different between the subjects ( $p < 0.05$ ) in terms of sequence read abundances. The largest proportion of inter-subject variation could be attributed to the increased prevalence of *Prevotella* in Subject 1

( $p < 10^{-25}$ ). Significantly higher proportions of *Ruminococcaceae* ( $p < 10^{-25}$ ) and *Cyanobacteria* ( $p < 10^{-23}$ ) in Subject 1 were also important drivers of inter-subject variation. Higher proportions of *Bacteroides* sp. observed in Subject 2, and higher proportions of *Sutterella* sp. observed in Subjects 1 and 2, but not Subject 3, were significantly associated with inter-subject variation ( $p < 10^{-25}$  and  $< 10^{-24}$ , respectively). A significantly higher proportion of *Bacteroides uniformis* was observed in Subject 3 ( $p < 10^{-23}$ ).

### INTER-SUBJECT VARIATION IN NEW ZEALAND, AMERICAN, MALAWI AND AMERINDIAN POPULATIONS

By including our data with those from a larger, international study (Yatsunenکو et al., 2012), we were able to test our predictions and initial results regarding inter-subject differences. We removed any subjects <3 years old from the Yatsunenکو data set, so that the effect of age on gut microbiota did not affect clustering. The samples clustered primarily by geography/culture (American vs. agrarian vs. New Zealand), and these differences were large enough to outweigh the technical differences observed between the three New Zealand subjects (Figure 4). Notwithstanding



**FIGURE 4 | Relative unweighted UniFrac phylogenetic distances between New Zealand, American, Malawian and Amerindian subjects imaged using two dimensional NMDS plot.** Vectors with a statistically significant contribution to the ordination are overlaid onto the ordination space. The sizes of taxa nodes are based on the average abundances of the taxon in fecal microbial communities from each group of subjects. Asterisk (\*) denotes a cluster that could not be resolved beyond taxonomic order.

**Table 2 | Mean estimates of alpha diversity metrics (mean  $\pm$  S.D.,  $p$ -values) calculated for each method from OTU tables rarefied to 11,739 sequences per sample.**

Method	Chao1	Observed OTUs	Shannon diversity	Simpson diversity
HMP	842 $\pm$ 64.9 ( $p = 1.0$ )	533 $\pm$ 44.8 ( $p = 0.32$ )	5.28 $\pm$ 0.41 ( $p = 0.01$ )	0.90 $\pm$ 0.04 ( $p = 0.02$ )
M	818 $\pm$ 90.1 ( $p = 0.33$ )	528 $\pm$ 46.0 ( $p = 0.04$ )	5.18 $\pm$ 0.64 ( $p = 0.01$ )	0.87 $\pm$ 0.09 ( $p = 0.02$ )
P	825 $\pm$ 61.4 ( $p = 0.24$ )	524 $\pm$ 38.4 ( $p = 0.02$ )	5.19 $\pm$ 0.52 ( $p = 0.01$ )	0.89 $\pm$ 0.07 ( $p = 0.01$ )
Q	825 $\pm$ 87.3 ( $p = 0.52$ )	531 $\pm$ 48.8 ( $p = 0.16$ )	5.13 $\pm$ 0.69 ( $p = 0.01$ )	0.87 $\pm$ 0.09 ( $p = 0.01$ )
Z	865 $\pm$ 61.0	562 $\pm$ 42.6	5.94 $\pm$ 0.30	0.94 $\pm$ 0.03

Each method represents data from 27 samples.  $P$ -values reported in the table are only associated with pairwise comparisons with Z method. No other pairwise comparisons between methods resulted in significant  $p$ -values ( $p < 0.05$ ).

**Table 3 | ANOSIM statistics with “BH” adjusted *p*-values on the comparisons between gut microbial community structures of subjects from New Zealand, Malawi, America and Venezuela.**

Country	<i>R</i> -value	<i>p</i> -value, “BH” adjusted
New Zealand vs. Malawi vs. America vs. Venezuela	0.797	0.0001
New Zealand vs. Malawi	0.827	0.0001
New Zealand vs. America	0.799	0.0001
New Zealand vs. Venezuela	0.996	0.0001
Malawi vs. America	0.839	0.0001
Malawi vs. Venezuela	0.041	0.1542
America vs. Venezuela	0.815	0.0001

the very low numbers of subjects from New Zealand, the 2D NMDS suggested that a transition from western (American and New Zealand) to agrarian (Malawian and Venezuelan) cultures could be associated with an enrichment in members from the bacterial families *Prevotellaceae*, *Clostridiaceae*, *Veillonellaceae*, and *Bifidobacteriaceae*. Transitioning from American and agrarian cultures toward New Zealand subjects was associated with an increase in sequences classified to the order *Bacteroidales* and the family *Rikenellaceae* (Figure 4). American subjects were associated with an enrichment in sequences classified to the order *Clostridiales*, and *Ruminococcaceae* and *Lachnospiraceae* families.

Results from ANOSIM tests revealed an overall significant difference in gut microbial community structures between New Zealand, Malawian, American, and Venezuelan countries (Table 3). ANOSIM pairwise tests suggested Malawian and Venezuelan gut community profiles were the most similar, while New Zealand and Venezuelan subjects were the most different. Although the bacterial profiles from New Zealand subjects were significantly different when compared to all other countries, they were most similar and grouped closer to those from America (Figure 4, Table 3).

## DISCUSSION

Understanding the origins of variation within an experimental model, and limiting known sources of bias, may help resolve inconsistencies within the literature regarding gut microbial dysbiosis and disease. A recent meta-analysis examined beta diversity among microbial communities generated from human stool samples across several human microbiome studies (Lozupone et al., 2013). Samples clustered strongly by age and geography/culture, which was to be expected (Yatsunenkov et al., 2012; Lozupone et al., 2013; Suzuki and Worobey, 2014). However, in data sets comparing fecal samples from subjects with similar ages and geographical/cultural background, secondary clustering by experimental protocol was observed. Our study aimed to identify and quantify the relative contribution of preventable technical bias associated with DNA extraction method.

This study draws attention to the importance of DNA extraction method when interpreting microbial community diversity measurements. Previously, lysis technique has been cited as a contributing factor when extracting microbial nucleic acids (Carrigg

et al., 2007; Feinstein et al., 2009). One recent study (Claassen et al., 2013) identified few significant differences between DNA extraction methods when examining lysis technique; however, a different study (Yuan et al., 2012) reported the most effective cell lysis and DNA recovery from bead beating and/or mutanolysin (Yuan et al., 2012; Claassen et al., 2013). In another comparison of lysis techniques, the most effective lysing procedure was hot phenol and bead beating, suggesting that a combination of lysing procedures captures the most accurate community composition (Wu et al., 2010). Our study examined a variety of lysing techniques; however, Shannon and Simpson diversity indices did not reveal significant differences in microbial community diversity specifically pertaining to lysing methods.

The MoBio PowerSoil® DNA Isolation Kit had the second highest median yield of DNA per extraction, and recovered high quality DNA with minimal shearing. While the phenol:chloroform method gave the highest median yield, this result is heavily influenced by the wide range of recovered DNA (7.56–395 ng/μL). Although phenol:chloroform extractions are still widely used, they may take considerably more time than kit methods, and often require extensive clean-up steps before PCR amplification to remove left-over phenol and humic acids. Additionally, human error can affect the reproducibility of these extractions when analysing recovered nucleic acid yield and quality. Compared with the other extraction methods, a large amount of sheared DNA was visualized at the bottom of the phenol:chloroform gel electrophoresis images and DNA yield was not reproducible.

Microbial DNA extracted using the ZR Fecal DNA MiniPrep™ kit returned consistently low DNA yields and poor A260/A280 ratios (<1.8). Analysis of the gel electrophoresis image supports the low amounts of extracted DNA, as bands were barely visible. Results from a study by Claassen et al. indicated much higher amounts of extracted DNA from ZR Fecal DNA MiniPrep™ kit when compared to other DNA extraction methods tested in that study, and when compared to the results presented in our work. Additionally, Claassen et al., reported a substantially higher median quality of DNA extracted using the ZR Fecal DNA MiniPrep™ kit (A260/A280 = 1.678) when compared to the results presented here. Closer examination of the data from that study revealed a substantial range of A260/A280 results below 1.6, even though the results were considered reproducible (CV < 1.0) (Claassen et al., 2013). Significantly lower DNA yield and quality from samples extracted using this kit, and subsequent PCR bias may help explain differences in community structure between ZR Fecal DNA MiniPrep™ and the other methods. Richness and diversity estimates for ZR Fecal DNA MiniPrep™ were significantly higher when compared to the other extraction methods. Claassen et al. (2013) reported the highest Shannon index for the ZR Fecal DNA MiniPrep™ kit, but the value was not significantly different from those estimated for the QIAamp® DNA Stool Mini Kit or MoBio PowerSoil® DNA Isolation Kit. By contrast, the results from our methods comparison suggest the ZR Fecal DNA MiniPrep™ kit was the only method that proved significantly different from the other methods across richness and diversity measures.



Comparisons between the bacterial community profiles resulting from the different DNA extraction methods revealed substantially increased overall bacterial diversity and recovery of members from the *Firmicutes* phylum in the samples extracted using ZR Fecal DNA MiniPrep™. A previous methods study that examined the ZR Fecal DNA MiniPrep™ kit also reported inflated representation of members from the *Firmicutes* phylum, as well as poor DNA quality (Henderson et al., 2013).

The methods examined in this study contribute to the ongoing debate regarding the most accurate and reproducible DNA extraction method. Within the limitations of this study, the results indicate that the most effective microbial DNA extraction method is the MoBio PowerSoil® DNA Isolation Kit. The reproducible high yield and quality of extracted DNA, as well as minimal shearing, support this decision. The community profile was significantly different to that of the ZR Fecal DNA MiniPrep™ kit, however it was similar to the microbial community profiles derived from all other methods. In addition to analyses of biological samples, the inclusion of a mock community, composed of known organisms in known quantities, will help elucidate how technical variation impacts upon recovered microbial community profiles.

The results for biological (inter-subject) vs. technical (DNA extraction method) variation within the New Zealand cohort demonstrate that while DNA extraction does impact upon interpretation of beta diversity, technical variation does not overcome observed community differences between subjects. Furthermore, intra-subject biological variation revealed no significant impact on the observed total variation within the experimental model. These results are consistent with other studies that reported greater inter-subject variation than intra-subject variation for stool samples from adults (Costello et al., 2009; Caporaso et al., 2011; Ursell et al., 2012).

The known, strong effect of inter-subject biological variation, especially related to geography/culture, on the human gut microbiota was observed in the data set presented here. As previously described, the abundance of *Prevotellaceae* was associated with Malawi and Amerindian subjects (Yatsunenko et al., 2012), as well as Subject 1 from New Zealand. Even though significant differences were observed in pairwise ANOSIM tests between subjects from New Zealand, Malawi, America, and Venezuela, the New Zealand subjects were more similar to those from America than the agrarian cultures. A paucity of information exists regarding the New Zealand adult gut microbiota, and the three subjects presented in this study may not be an accurate representation of the average New Zealand gut community. Increasing the number of New Zealand samples may help moderate the extreme biological differences depicted between the three subjects sampled here. Additionally, increasing the number of New Zealand subjects will help describe their gut communities within the global context, especially in relation to other “western” cultures.

High labor costs and time constraints have led to the development of many commercially available DNA extraction kits. Such kits allow researchers to quickly and efficiently extract DNA, with minimal clean-up steps before amplification. While this study is limited by the exclusion of a mock community, or any way of knowing the actual microbial composition within

the stool samples, significant differences in community structure were observed between extraction methods, and these differences were the second-greatest contributing factor to variation observed within this study.

## ACKNOWLEDGMENTS

This work was supported by funding from a University of Auckland Faculty Research Development Fund grant to MT (grant 9841 3702219). DW was supported by a University of Auckland Doctoral Scholarship. A special thanks to Gerald Tannock for his helpful advice and Peter Tsai for his bioinformatics expertise. The work described here was carried out within the broad framework of the Minds for Minds autism research initiative.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2015.00130/abstract>

## REFERENCES

- Anderson, M. J., Gorley, R. N., and Clarke, K. R. (2008). *PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods*. Plymouth: PRIMER-E Ltd.
- Ariefdjohan, M. W., Savaiano, D. A., and Nakatsu, C. H. (2010). Comparison of DNA extraction kits for PCR-DGGE analysis of human intestinal microbial communities from fecal specimens. *Nutr. J.* 9:23. doi: 10.1186/1475-2891-9-23
- Aronesty, E. (2011). *ea-utils: “Command-Line Tools for Processing Biological Sequencing Data.”* Available online at: <http://code.google.com/p/ea-utils>.
- Bäckhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., et al. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15718–15723. doi: 10.1073/pnas.0407076101
- Bahl, M. I., Bergström, A., and Licht, T. R. (2012). Freezing fecal samples prior to DNA extraction affects the *Firmicutes* to *Bacteroidetes* ratio determined by downstream quantitative PCR analysis. *FEMS Microbiol. Lett.* 329, 193–197. doi: 10.1111/j.1574-6968.2012.02523.x
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300.
- Cantarel, B. L., Lombard, V., and Henrissat, B. (2012). Complex carbohydrate utilization by the healthy human microbiome. *PLoS ONE* 7:e28742. doi: 10.1371/journal.pone.0028742
- Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T., Anderson, G., and Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26, 266–267. doi: 10.1093/bioinformatics/btp636
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth0510-335
- Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., et al. (2011). Moving pictures of the human microbiome. *Genome Biol.* 12:R50. doi: 10.1186/gb-2011-12-5-r50
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2010c). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4516–4522. doi: 10.1073/pnas.1000080107
- Carrigg, C., Rice, O., Kavanagh, S., Collins, G., and O’Flaherty, V. (2007). DNA extraction method affects microbial community profiles from soils and sediment. *Appl. Microbiol. Biotechnol.* 77, 955–964. doi: 10.1007/s00253-007-1219-y
- Claassen, S., du Toit, E., Kaba, M., Moodley, C., Zar, H. J., and Nicol, M. P. (2013). A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. *J. Microbiol. Methods* 94, 103–110. doi: 10.1016/j.mimet.2013.05.008

- Claesson, M. J., Jeffery, I. B., Conde, S., Power, S. E., O'Connor, E. M., Cusack, S., et al. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488, 178–184. doi: 10.1038/nature11319
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x
- Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science* 326, 1694–1697. doi: 10.1126/science.1177486
- DeSantis, T., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E., and Keller, K. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05
- Dridi, B., Henry, M., El Khéchine, A., Raoult, D., and Drancourt, M. (2009). High prevalence of *Methanobrevibacter smithii* and *Methanospaera stadtmanae* detected in the human gut using an improved DNA detection protocol. *PLoS ONE* 4:e7063. doi: 10.1371/journal.pone.0007063
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., et al. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638. doi: 10.1126/science.1110591
- Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604
- Feinstein, L. M., Sul, W. J., and Blackwood, C. B. (2009). Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl. Environ. Microbiol.* 75, 5428–5433. doi: 10.1128/AEM.00120-09
- Gevers, D., Kugathasan, S., Denson, L., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392. doi: 10.1016/j.chom.2014.02.005
- Henderson, G., Cox, F., Kittelmann, S., Miri, V. H., Zethof, M., Noel, S. J., et al. (2013). Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS ONE* 8:e74787. doi: 10.1371/journal.pone.0074787
- Hsiao, E. Y., McBride, S. W., Hsien, S., Sharon, G., Hyde, E. R., McCue, T., et al. (2013). Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* 155, 1451–1463. doi: 10.1016/j.cell.2013.11.024
- Kennedy, N. A., Walker, A. W., Berry, S. H., Duncan, S. H., Farquarson, F. M., Louis, P., et al. (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:e88982. doi: 10.1371/journal.pone.0088982
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, e1. doi: 10.1093/nar/gks808
- Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R., and Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol. Lett.* 307, 80–86. doi: 10.1111/j.1574-6968.2010.01965.x
- Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848. doi: 10.1016/j.cell.2006.02.017
- Li, M., Gong, J., Cottrill, M., Yu, H., de Lange, C., Burton, J., et al. (2003). Evaluation of QIAamp® DNA stool mini kit for ecological studies of gut microbiota. *J. Microbiol. Methods* 54, 13–20. doi: 10.1016/S0167-7012(02)00260-9
- Lozupone, C. A., Stombaugh, J., Gonzalez, A., Ackermann, G., Wendel, D., Vázquez-Baeza, Y., et al. (2013). Meta-analyses of studies of the human microbiota. *Genome Res.* 23, 1704–1714. doi: 10.1101/gr.151803.112
- Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature* 489, 220–230. doi: 10.1038/nature11550
- Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228
- Maukonen, J., Simões, C., and Saarela, M. (2012). The currently used commercial DNA-extraction methods give different results of clostridial and actinobacterial populations derived from human fecal samples. *FEMS Microbiol. Ecol.* 79, 697–708. doi: 10.1111/j.1574-6941.2011.01257.x
- McInnes, P., and Cutting, M., (2010). *Manual of Procedures for Human Microbiome Project, Core Microbiome Sampling, Protocol A, HMP Protocol # 07-001*. Available online at: [http://www.hmpdacc.org/doc/sops\\_2/manual\\_of\\_procedures\\_v11.pdf](http://www.hmpdacc.org/doc/sops_2/manual_of_procedures_v11.pdf).
- McOrist, A. L., Jackson, M., and Bird, A. R. (2002). A comparison of five methods for extraction of bacterial DNA from human faecal samples. *J. Microbiol. Methods* 50, 131–139. doi: 10.1128/JCM.42.12.5913-5916.2004
- Nechvatal, J. M., Ram, J. L., Basson, M. D., Namprachan, P., Niec, S. R., Badsha, K. Z., et al. (2008). Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *J. Microbiol. Methods* 72, 124–132. doi: 10.1016/j.mimet.2007.11.007
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2015). *Vegan: Community ecology*. Available online at: <http://CRAN.R-project.org/package=vegan>.
- Peng, X., Yu, K.-Q., Deng, G.-H., Jiang, Y.-X., Wang, Y., Zhang, G.-X., et al. (2013). Comparison of direct boiling method with commercial kits for extracting fecal microbiome DNA by Illumina sequencing of 16S rRNA tags. *J. Microbiol. Methods* 95, 455–462. doi: 10.1016/j.mimet.2013.07.015
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077
- The Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60. doi: 10.1038/nature11450
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna. Available online at: <http://www.r-project.org/>.
- Schloissig, S., Arumugam, M., Sunagawa, S., Mitrava, M., Tap, J., Zhu, A., et al. (2013). Genomic variation landscape of the human gut microbiome. *Nature* 493, 45–50. doi: 10.1038/nature11711
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Suzuki, T. A., and Worobey, M. (2014). Geographical variation of human gut microbial composition. *Biol. Lett.* 10:20131037. doi: 10.1098/rsbl.2013.1037
- Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031. doi: 10.1038/nature05414
- Ursell, L. K., Clemente, J. C., Rideout, J. R., Gevers, D., Caporaso, J. G., and Knight, R. (2012). The interpersonal and intrasubject diversity of human-associated microbiota in key body sites. *J. Allergy Clin. Immunol.* 129, 1204–1208. doi: 10.1016/j.jaci.2012.03.010
- Vázquez-Baeza, Y., Pírrung, M., Gonzalez, A., and Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience* 2:16. doi: 10.1186/2047-217X-2-16
- Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108. doi: 10.1126/science.1208344
- Wu, G. D., Lewis, J. D., Hoffmann, C., Chen, Y.-Y., Knight, R., Bittinger, K., et al. (2010). Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol.* 10:206. doi: 10.1186/1471-2180-10-206
- Yatsunenkov, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., et al. (2012). Human gut microbiome viewed across age and geography. *Nature*, 486, 222–227. doi: 10.1038/nature11053
- Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36, 808–812. doi: 10.2144/3605A0808
- Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., and Forney, L. J. (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS ONE* 7:e33865. doi: 10.1371/journal.pone.0033865

Zoetendal, E. G., Heilig, H. G. H. J., Klaassens, E. S., Boonjink, C. C. G. M., Kleerebezem, M., Smidt, H., et al. (2006). Isolation of DNA from bacterial samples of the human gastrointestinal tract. *Nat. Protoc.* 1, 870–873. doi: 10.1038/nprot.2006.142

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 November 2014; accepted: 04 February 2015; published online: 18 February 2015.

*Citation:* Wagner Mackenzie B, Waite DW and Taylor MW (2015) Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Front. Microbiol.* 6:130. doi: 10.3389/fmicb.2015.00130  
This article was submitted to *Microbial Symbioses*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2015 Wagner Mackenzie, Waite and Taylor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.