Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Curr Protoc Hum Genet. Author manuscript; available in PMC 2015 July 14.

Published in final edited form as:

Curr Protoc Hum Genet. ; 82: 18.8.1–18.8.29. doi:10.1002/0471142905.hg1808s82.

Getting Started with Microbiome Analysis: Sample Acquisition to Bioinformatics

Ranjit Kumar[1], **Peter Eipers**[2], **Rebecca B. Little**[3], **Michael Crowley**[4], **David K. Crossman**[4], **Elliot J. Lefkowitz**[5], and **Casey D. Morrow**[2]

[1]Center for Clinical and Translational Sciences, University of Alabama at Birmingham, Birmingham Alabama 35294

[2]Department of Cell, Developmental and Integrative Biology, University of Alabama at Birmingham, Birmingham Alabama 35294

[3]Department of Nutrition Sciencesm, University of Alabama at Birmingham, Birmingham Alabama 35233

[4]Department of Genetics and Heflin Center for Genomic Sciencel, University of Alabama at Birmingham, Birmingham Alabama 35294

[5]Department of Microbiology, University of Alabama at Birmingham, Birmingham Alabama 35294

A Microbiome Analysis combines sample collection and processing, NextGen sequencing and bioinformatics analysis to provide unprecedented details of the composition of microbiota at different sites on the human body. While at first, it might appear that the components required for “in house” microbiome analysis may be beyond the capabilities of some research laboratories, in reality, the requisite components for establishing a microbiome analysis pipeline may already exist at many institutions, eliminating the need for investigators to contract out for microbiome analysis.

In the following sections we describe a Microbiome Analysis analytical pipeline from acquisition of samples, multiplex DNA sequencing using the Illumina MiSeq sequencing platform and bioinformatics data analysis to accommodate a wide spectrum of clinical and basic science investigators. Samples can be acquired from saliva (Basic Protocol 1), buccal swabs (Alternate Protocol 1), vaginal or skin samples (Basic Protocol 2), or fecal specimens

Address correspondence to: Ranjit Kumar, Center for Clinical and Translational Sciences, University of Alabama at Birmingham, Birmingham Alabama 35294, rkumar@uab.edu; Casey D. Morrow, Department of Cell, Developmental and Integrative Biology, Director, UAB Microbiome Resource, 1918 University Blvd. – MCLM 680, Birmingham, AL 35294, TEL: (205) 934-5705, caseym@uab.edu.

Internet Resources

The 16S rRNA gene has been sequenced for a wide range of cultured species and environmental isolates (Stackebrandt and Goebel, 1994; Woese and Gutell, 1989). The sequences are publicly available from several rRNA sequence databases including GreenGenes, the Ribosomal Database Project, and Silva (Cole et al., 2009; DeSantis et al., 2006; McDonald et al., 2012) that can be automatically searched for matches to sequences derived from samples of interest.

QIIME – http://www.qiime.org/

GreenGenes, – http://greengenes.lbl.gov/cgi-bin/nph-index.cgi

Ribosomal Database Project – http://rdp.cme.msu.edu/

Silva – http://www.arb-silva.de/

UniFrac – http://bmf.colorado.edu/unifrac/

MOTHUR – http://www.mothur.org/

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Basic Protocol 3 and Alternate Protocol 2). Support Protocols 1 and 2 provide information about transport of samples. Basic Protocols 4–6 describe DNA preparation, generation of the amplicon library and sample prep for NextGen sequencing. Support Protocol 3 describes how to quantify the DNA using PicoGreen. Finally, computational protocols using the QWRAP QIIME Wrapper are provided in Basic Protocols 7 and 8.

ACQUISITION OF SAMPLES FOR MICROBIOME ANALYSIS

The first step in establishing a microbiome analysis pipeline is to establish the necessary infrastructure for collection of samples. In this case, samples are collected from patients and "normal" controls. Some of the samples, such as those for the oral and vaginal microbiomes, can be collected during patient visits to their physician. A complex challenge is to collect fecal samples, which are used to study the gut microbiome. The main issue here is that fecal samples cannot be obtained "on demand" as you can for the other samples. In most cases, fecal samples are obtained at the patient's home and transported to the laboratory. Proper handling of these samples then is essential to avoid microbial growth that could result in biased microbiomes. That is, improper handling of the samples that result in extended times at room temperature or hotter would be expected to promote the replication of aerobic microbes, while anaerobic microbes in the stool would not grow (Nechvatal et al., 2008). We have found that two methods work well for collection and transport of fecal samples. For collection of small amounts of fecal material, enough for a few microbiome determinations, we favor the "wipe collection method", where the patient places a moist wipe used after defecation into a plastic bag before transport to the laboratory. If the patient is to bring the sample into laboratory the next day, freezing the wipe at –20 in a conventional freezer effectively kills the microbes (to prevent microbe growth which could bias the microbiome results) but still allows processing after transport to the laboratory. In cases where viable microbes are needed for study (such as those used for transplant into gnotobiotic mice), we have developed a method for collection and transport of the sample to preserve microbe viability. After the stool sample is placed into a specimen cup, about 50ml of a "modified" Cary-Blair medium is added to the fecal sample (this is the "clear solution in the orange capped tube" in the above example of instruction to the participant). Finally, it is also important for investigators to obtain patient information that could impact the microbiome such as diet, antibiotic use and body weight and we have provided examples for collection of this information.

Basic Protocol 1: Oral Sample Acquisition: Saliva

One of ways to access the oral microbiome is to obtain a saliva sample. The most straightforward way to obtain saliva is to have the patient into spit into a container, usually a small tube. It is always good form to first thank the patient for providing a sample, such as saliva. One important reason for this is that from our experience, the donor has to be patient because on average it takes 2 to 5 minutes to produce a saliva sample for most people. If the donor has an accompanying medical condition or is on medications, which cause dry mouth, the production of saliva could be more difficult and take longer.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Materials

Plastic collection tube: There are many collection tubes available. We have found a standard laboratory 50ml conical tube to be the least expensive.

The following are general directions that we give to the donors:

1. Please wait about 30 minutes after drinking, eating, smoking, or chewing gum to take the sample.
2. Open plastic collection tube.
3. Spit into the collection tube until the amount of liquid saliva (not counting the foam) is at the 5 ml line.
4. Screw the orange lid on plastic tube. Please make sure the lid is closed securely. The saliva samples should be stored at −20 or −80°C until processing.

Alternate Protocol 1: Oral Sample Acquisition: Buccal

A second, different oral microbiome can be obtained by swabbing the inside of the cheek. In this case, the microbes obtained by swabbing the inner cheeks contain not only microbes found in the saliva but also microbes that adhere to the epithelial cells of the cheek. This sampling method is called a buccal swab.

Materials—Any type of cotton tip sample will work, although we have found a suitable collection swab can be obtained from Fisher Scientific (catalog # 1490710).

1. To obtain buccal samples for analysis of the oral microbiome, use a soft cotton tip to rub the inside cheek to obtain free and attached microbes.

   The oral cavity also has several other sites that are amenable to microbiome analysis, including the tongue, tooth surface and gingival tissue. For the most part, the tongue and tooth surface can be obtained with the same swab for collection as above. For gingival tissue, a toothpick can be used, although it would be best to consult a dentist for the most appropriate method.

2. Store the swabs at −20 or −80°C until processing for DNA isolation.

Basic Protocol 2: Acquisition of Vaginal and Skin Samples

For vaginal and skin microbiome analysis, the samples are usually obtained at a clinic visit by a physician. The usual mode of collection is to use a swab (similar to that used for buccal collection). Copan Diagnostics (Murrieta, California) (http://www.copanusa.com/) sells flocked nylon swabs for collection. The swabs can be stored at −20 or −80°C until processing for DNA isolation.

Basic Protocol 3: Acquisition of Fecal Samples Using the Pre-Moistened Wipe Method

For just a small number of microbiome analyses, this method provides sufficient sample size. If the patient is to bring the sample into laboratory the next day, freezing the wipe at −20 in a conventional freezer effectively kills the microbes (to prevent microbe growth

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

which could bias the microbiome results) but still allows processing after transport to the laboratory.

Materials

Small or Medium Therapak box -if shipping; if participant is bringing the sample back to the clinic another container with a biohazard bag can be used. Therapak boxes can be obtained through Fisher Scientific. These boxes are assembled as kits with many of the additional supplies needed to ship the sample such as a biohazard bag, stickers for labeling the box “Exempt Human Specimen” or with the appropriate UN category, and absorbent material to wrap around collection vial.

Scott Naturals Moist Wipe placed in a plastic zip top bag

FedEx Clinical Pak (a bag overwrap for Clinical Sample shipment) – if shipping

Shipping label holder – if shipping

Collection instructions

Collection instructions for participant:

1. If possible, it is best to complete the stool sample in the morning and collect the first bowel movement of the day.
2. After bowel movement, wipe with the provided pre-moistened wipe. Please do not get urine on the wipe.
3. Fold the wipe in half and place in the biohazard lab bag.
4. Seal the biohazard lab bag and write date and time on the bag.
5. Place in the provided box.
6. Wash hands.
7. Close the box and check that the biohazard sticker is on the outside.
8. If shipping the sample, place box in FedEx Clinical Pak bag and seal. (See “Stool Method” and “Transporting samples to the laboratory” more information.)

Alternate Protocol 2: Acquisition of Fecal Samples Using the Stool Method

The following protocol should be followed by participants in cases where viable microbes are required for analysis. See Support Protocols 1 and 2 for further information about transporting the samples and handling them in the laboratory upon arrival.

Materials

Large brown paper bag (to pack all the kit items in for the participant)

Small or Medium Therapak box

Vial of 50 ml modified Cary Blair medium (see Reagents and Solutions).

2 or 2 ½ gallon plastic zipper bag or garbage bag to use as disposal bag for collection pan

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

pair of gloves

commode collection pan (i.e. "plastic hat" or "toilet hat")

2 plastic spoons (easier to scoop up a sample)

120 ml urine collection vial

collection instructions

Optional: 3 oz PolarPack freezer gel pack

Directions for Participant:

Always thank the participant for providing a stool sample. Tell the participant the stool sample collection and shipment should be the same day.

1. Place plastic hat over toilet seat to collect stool (no urine).
2. Using provided spoon, fill up the cup about half full (close to the 60 ml line). Gloves are provided if you want to use them.
3. Add the entire clear solution in the provided tube to the stool sample in the cup (the solution smells bad – it is OK).
4. Screw the cap onto the specimen cup with stool/solution. Please wrap provided tape around the top.
5. Wrap paper towel or included absorbent pad around collection and place in biohazard bag.
6. Put the stool sample back into the box with tape and put box with sample into FedEx ClinicalPak (that we provide). Note: FedEx will ship sample only if it is in box and box is in Pak.
7. You can discard the plastic hat, orange-capped tube and gloves in plastic bag in trash.
8. We have included a FedEx billing to send the sample to the Investigator. Please call FedEx for pick up or take to FedEx place- there will be no cost for shipping to you.

Support Protocol 1: Transporting Fecal Samples to the Laboratory

Once a FedEx account is set up, labels can be generated online. To protect participant identity, the receiving lab can be listed as the return address also. The labels can then be printed out or either emailed to the study coordinator or straight to the participant. We use Next Day shipping to ensure samples are received and processed before overgrowth can occur. FedEx and all other postal carriers require compliance with the International Air Transport Association (IATA) guidelines for patient specimens and possibly infectious substances. The packages are required to be labeled as either "Exempt Human Specimen" or with the appropriate UN code if an infectious disease is likely or confirmed. All samples require the following: 1) a leak-proof primary receptacle; 2) a leak-proof secondary packaging (i.e. biohazard bag with receptacle wrapped in absorbent material); 3) an outer package of adequate strength to transport the sample with at least one surface having

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

minimum dimensions of 100 mm × 100 mm; 4) and FedEx requires their ClinicalPak outer wrap.

Support Protocol 2: Processing of Fecal Samples After Shipment

One of the unintended benefits of the shipping process is that through the natural agitation during shipping, the sample is fully mixed in the modified Cary-Blair medium. The fecal samples are diluted in Cary-Blair medium to 0.1mg/ml for a total volume of 20 mL with 10% by volume glycerol. Aliquots of 5 mL are dispensed into cryovial tubes and stored at −80°C. These samples represent the "fecal library" of archived microbes that can be used for transplant studies if desired. We also store 4 200μL aliquots (without added glycerol) that are dispensed into 4 small screw cap vials. These are used for later DNA extraction/ microbiome analysis and also stored at −80°C.

MULTIPLEXED 16S AMPLICON SEQUENCING ON THE MiSeq® SYSTEM

In this section, we describe the process whereby the samples collected in the first section are processed for DNA using commercially available kits. PCR is used with primers specific for the V4 region of the 16S rRNA gene for amplification. (Caporaso et al., 2011; Kozich et al., 2013) (Figure 1). A previous study from Caparoso et al (2011) described a microbiome analysis using degenerate PCR primers specific for the V4 region (Caporaso et al., 2011) (Figure 2). The 5′ primer consists of the adaptor sequence, which basically allows binding to the DNA chip, a "pad" and "link" sequence which is used for binding of the sequencing primer and to help facilitate PCR and the sequence (highlighted in grey and black) that is complementary to the V4 region of the 16S rRNA gene. This primer sequence contains degenerate bases that allow binding to the more conserved regions surrounding V4 (Caporaso et al., 2011). The 3′ primer contains many of the same elements as the 5′ (adaptor, link and pad), but also contains a 6 base pair region (NNNNNN) for the "bar coded" index sequences. These sequences are used to distinguish separate PCR reactions (and thus separate samples) after the sequencing reaction (Figure 3). We have used 94 barcoded 3′PCR primers that allow sequencing of 94 samples per run of an Illumina MiSeq DNA chip. We find that this number of samples per run allows the quantification of the PCR product using a convenient 96 well microtiter plate.

Basic Protocol 4: DNA Isolation

Although there are many ways to isolate DNA from samples, we have found that the use of "kits" is an effective way to prepare DNA from a large number of samples. We have used the "Fecal DNA isolation" kit that can be purchased from Zymo Research (catalog # D6010). The kit provides all of the necessary reagents and columns required for the isolation of total DNA from fecal samples. We have also had success using the same kit for isolation of DNA from oral, vaginal and skin samples.

Materials

Microbiome sample (see Basic Protocols 1–3 and Alternate Protocols 1–2)

Fecal DNA Isolation Kit (Zymo Research, #D6010)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Tris-EDTA buffer

1. Isolate total DNA from the stored individual samples following the instructions for the Zymo Research Fecal DNA Isolation Kit. Add 750μl of the "Lysis Solution" to the samples (usually 100 μl of saliva or add the Lysis solution directly to the swab containing the buccal, vaginal or skin sample for 5 minutes, then remove the swab). Follow the subsequent processing steps to isolate DNA.
2. Use the isolated DNA immediately for PCR or store in standard Tris-EDTA buffer at 4°C.
3. Quantitate the isolated PCR DNA prior to PCR using a micro spectrometer (www.nanodrop.com/) (Kozich et al., 2013).

Basic Protocol 5: Generate Amplicon Library by PCR

Once the sample DNA is prepared, PCR is used with unique bar coded primers to amplify the V4 region of the 16S rRNA gene to create an "amplicon library" from individual samples. The PCR products from the individual samples are electrophoresed on an agarose gel and visualized by UV illumination. The PCR product is excised from the gel and purified using a commercial gel extraction kit. The purified product is then ready for DNA sequencing.

Materials—Degenerate PCR primers specific for the V4 region (Eurofins – mwg/operon (http://www.operon.com). The primers were ordered at 50 nmol scale with desalting purification. Storage stock diluted with 10 mM Tris pH 8.0 to 100 μM, then diluted 10× in water to 10μM for use in PCR reactions).

New England Biolabs LongAmp Taq PCR kit (http://www.neb.com, cat # E5200S). The kit contains all of the necessary reagents for the PCR reactions.

Qiagen QIAquick Gel Extraction Kit (cat # 28704).

1. Set up individual reactions as follows:

   10 μL of 5× Reaction Buffer

   1.5μLof dNTPs

   2μL of the 5 'Primer diluted as described above

   2 μL of the 3′ Primer diluted as described above

   1.5 μL of the "LongAmp" enzyme supplied in the kit

   30 μL of the Template DNA prepared using the "Fecal DNA Isolation kit. The concentration of DNA should be at 2–5 ng/ul

   3μL of H_2O

   The total reaction volume is 50 μL

   Cycling conditions:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Initial denature

94° C 1 minute

32 cycles of:

94°C 30 seconds

50°C 1 minute

65°C 1 minute

Final extension:

65°C 3 minutes

Final hold: 4° C

2. Load the entire PCR reaction on a 1.0% agarose/Tris-borate-EDTA gel and electrophorese.

3. The PCR product is approximately 380 base pair predicted product size. Visualize by UV illumination and photograph, then excise and purify from the agarose using Qiagen QIAquick Gel Extraction Kit according to manufacturer's instructions. (http://www.qiagen.com, cat # 28704).

Basic Protocol 6: Sample Preparation for Loading onto the Illumina MiSeq

Sequence the PCR products using NextGen sequencing Illumina MiSeq platform. The MiSeq is a single flowcell, single lane instrument that can generate approximately 9Gb of sequence data from a paired end 250bp run (Caporaso et al., 2012; Schloss et al., 2009). We have found that sequencing using the paired 250 base pair end kits from Illumina works well for the V4 region in the microbiome analysis. A comprehensive discussion of the machine and work flow can be found at the following link:

http://res.illumina.com/documents/products/datasheets/datasheet_miseq.pdf

Materials

Tris pH8.5:Tween 20 (10mM:0.1%)

5–10% PhiX (Illumina)

2N NaOH

ice cold hybridization buffer (HT1 from Illumina)

1. Following sample quantitation with Pico Green (see Support Protocol 3), dilute the PCR products to a final concentration of 4nM. Then process the samples for sequencing on the Illumina MiSeq as follows:

2. Dilute individual samples to 4nM in Tris pH8.5:Tween 20 (10mM:0.1%). (ng of DNA/(avg. fragment size × 660Da) × 1.0 $\times 10^6$

3. Combine equal volume of 4nM libraries into one 1.5mL microfuge tube (e.g 5μL of each sample).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

4. Spike in 5–10% PhiX (Illumina) or other balanced genome.

   The addition of the balanced genome (e.g. PhiX or another genomic type library) is necessary to achieve proper focusing and template generation due to the low diversity of base composition in the first 4 nucleotides within the PCR amplicon of the 16S rDNA V4 region.

5. In a separate tube, dilute fresh 2N NaOH to 0.2N NaOH in water.
6. Combine 5μL of combined libraries with 5μL of 0.2N NaOH in a separate microfuge tube.
7. Incubate at 25°C for 5min.
8. Add 990 μL of ice cold hybridization buffer (HT1 from Illumina). This results in a 20pM solution of library. Mix well by vortex.
9. Dilute libraries to 12.5pM by mixing 375μL of 20pM library with 225μL hybridization buffer.
10. Add all 600μL of 12.5pM library/hyb mix to position 17 of the MiSeq cassette.
11. Using a 500μL HandiStepper tip with HandiStepper remove 500μL of Read 1 sequencing primer mix from position 12 of the MiSeq cassette and transfer to a 1.5mL microfuge tube.
12. Add 3.4μL of 100mM custom Read 1 sequencing primer and mix well. Final concentration will be 0.5mM (Figure 4).
13. With HandiStepper return Read 1 solution to position 12. Mix well with HandiStepper.
14. Repeat with Index primer in position 13 and Read 2 sequencing primer in position 14 (**see** Figure 4 **for index and read 2 primer**).

Support Protocol 3: Quantification of PCR DNA using Pico Green

The purified amplicon PCR products are quantitated using Pico Green dsDNA reagent. The following is for the Tecan Infinite M200 fluorescent plate reader but can be applied to any fluorometric plate scanner.

Materials

9 2mL screw top tubes with caps

Stock λ DNA (100mg/mL)

5mL of 1× Tris-EDTA

(96 samples/plate)

6 96-well dilution plates

150mL of 1× Tris-EDTA (TE)

Pico Green solution (Invitrogen Catalog # P11495)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

35mL pipette trough

Two 96 well tube racks

Vortexer

P10, P20 and P200 pipettes and tips

1. To create a standard curve using λ DNA, make 1:20 dilutions yielding 5μg/mL in 1 mL total volume (950μL of 1× TE and 50μL of λ DNA). Make Stock DNA standards according to Table 1.
2. To create dilution plates, arrange tubes in 96-tube rack and create a plate layout, noting the location of each sample (e.g., well A01 = IDxxxx).
3. Depending on the probable concentration of the DNA, either make a 1:100 or 1:1000 dilution. Remember, the final concentration has to be with the range of the standard curve (0.02–2 ng/μL).

   All dilutions to be done in separate duplicate dilutions.
4. To prepare the 1:100 dilution, add 1 μL of sample to a dry deep-well 96-well plate. Depending on speed of technician, pipette DNA into 2 (slow)-4 (fast) rows before adding TE.
5. With a multi-channel pipette, add 99 μL of 1× TE to each well containing DNA.
6. In increments of between 2–4 rows, securely cover completed wells with a clear plate sealer cut into halves or fourths.
7. When **all** wells have 100 μL @ 1:100, vortex plate 1 min at 1000 rpm on high-speed shaker, centrifuge plate briefly at 280 rpm to re-collect all samples in wells.
8. In appropriately sized plastic tube, covered in foil to effectively block the light, make a 1:200 dilution of Pico Green, with enough volume to add 50μl per sample, including standard!

   Keep Pico Green solution in the dark in 50 μL aliquots in a −20°C freezer. Follow Invitrogen instructions for further handling.

   Example calculations:

   Pico Green Solution for 2 full quantitation plates (duplicates of same sample plate) + standards – this should be:

   [(96 samples × 2) + 8 standards + 1 blank control)] = 201 samples

   ROUND UP SAMPLES – 205 SAMPLES, IN THIS EXAMPLE.

   205 samples × 50 μL per sample = 10,250 μL or 10.25 mL total volume needed.

   To calculate the quantity of Pico Green

   V1 × C1 = V2 × C2; V1 is the volume of Pico Green you will need.

   V1 × (200× Pico Green) = 10.25 mL × (1×)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

V1 = 10250/200 = 51.25 L = volume of Pico Green stock

10250 µl – 51.25 µl = 10198.75 µL Tris-EDTA buffer

- Cover 15 mL conical tube with foil
- Add 10.198 mL of 1× TE buffer
- Add 51.25 µL of thawed Pico Green
- Invert to mix

9. Add the 1× Pico Green solution to a 50 mL reservoir and, using a multi-channel pipette, add 50µL of Pico Green Solution to each well of the black quantitation plates, including blanks. Pump-mix each well 3 times to ensure homogeneity.

Add Pico Green to one plate at a time. After adding Pico Green to one plate, cover the reservoir containing the remaining Pico Green solution by setting on it an un-detached foil plate sealer. Proceed to reading the plate containing Pico Green, and return to repeat the process with the next plate. If running a standard on a plate separate from the samples, process the standard plate first and the others in succession. Cover plate with foil plate seal.

10. Read the plate on a Tecan Infinite M200 fluorescent plate reader (although the plate can be applied to any fluorometric plate scanner). The settings for the Tecan Infinite M200 reader are:

Excitation: 480 nm

Emission: 520 nm

Shake Linear 3 secs.

Optimal Gain Setting for Standard Plate

Manual Gain set to the gain value determined on the Standard Plate

Additional information can be found at helpdesk-us@trcan.com.

MICROBIOME DATA ANALYSIS

The microbiome analysis package **QIIME** (*Quantitative Insights Into Microbial Ecology*) provides the vast majority of useful, important bioinformatics tools needed to support a complete and thorough microbiome analysis (Caporaso et al., 2010b; Lozupone et al., 2006; Navas-Molina et al., 2013). However, the package is quite large with multiple components, and the naïve user has to spend a significant amount of time to review the manuals and scripts in order to design a successful analytical workflow. This can be quite overwhelming for the novice user who just wants to carry out a basic or preliminary analysis. So, in an effort to support preliminary microbiome data analysis and data sharing in a high throughput environment supporting multiple investigators, we have developed a package called **QWRAP** (**Q**IIME **wrap**per). QWRAP acts as a software wrapper, guiding the user through the use of publically available tools such as FASTQC and FASTX for quality checking and quality-filtering followed by QIIME tools to perform microbiome analysis. QWRAP scripts are written using the bash, perl, python and R languages. QWRAP works on de-multiplexed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Illumina fastq files and can be used with either single end reads or overlapping paired end reads. Following analysis, QWRAP generates a static HTML report that can be viewed in any web browser (e.g. Internet Explorer, Firefox, Google Chrome). Since the web pages are provided as static HTML files, the entire results folder can be shared among other users so that they can view the data on their local workstations. A flowchart for using QWRAP is presented in Figure 5. The front page of the HTML report is displayed in Figure 6. Below we describe the utilization of QWRAP to perform a basic microbiome analysis, and discuss the different third-party tools utilized and reports generated during the analysis. The files generated by QWRAP are also designed to be directly compatible with other analytical programs and tools, such as some of the more advanced tools provided by QIIME. Therefore most of the protocols described below can be easily extended to provide more detailed and specialized analyses.

Necessary resources

As a prerequisite, users should have a basic idea of how to install software on a Linux operating system. The following tools must be installed for proper functioning of the tool QWRAP. All these tools can be either installed on a Linux machine or within a QIIME virtual Box (see http://qiime.org/install/virtual_box.html).

- **a.** QIIME: There are several ways to install QIIME (http://qiime.org/install/index.html). Virtual box based installation are the easiest to install and support the analysis of small data sets and can be installed on a Windows PC. (Kuczynski et al., 2011).
- **b.** USEARCH (http://www.drive5.com/usearch/). After downloading the 32bit Linux binary, rename the binary file to usearch61
- **c.** FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
- **d.** FASTX (http://hannonlab.cshl.edu/fastx_toolkit/)
- **e.** R (http://www.r-project.org/)

Basic Protocol 7: Installing QWRAP and Other Dependencies

QWRAP is available for download from GitHub at https://github.com/QWRAP/QWRAP. It can be installed using the following command on a Linux terminal:

```
git clone git://github.com/QWRAP/QWRAP.git
```

This command will download all the QWRAP scripts to the current directory in a folder named QWRAP. Now you need to add the location of this folder to the system PATH variable so that you can run the program from any location on computer.

For example, if the absolute path for the QWRAP folder is “ `/home/username/QWRAP`” then you can add the following line to file “.bashrc” which should be present in your home directory (`/home/username/.bashrc`):

```
export PATH=${PATH}:/home/username/QWRAP/
```

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(You must logout and re-login once to make it effective). The following web link provides detailed instruction on how to add file and folder locations to the PATH variable: http://www.cyberciti.biz/faq/unix-linux-adding-path/.

You should test all the installed software before starting any analysis. The details on how to test are provided in the critical parameters heading in the commentary section of this manuscript.

Basic Protocol 8: Microbiome Analysis using QWRAP

QWRAP is designed for de-multiplexed datasets. So if your reads are multiplexed, they need to be de-multiplex using Illumina software before using QWRAP.

1. Create a new directory: It's a good idea to create a new directory " `ANALYSIS`" for the storage of files generated during the analysis. This directory will also be used for the preparation of the HTML report:

    ```
    mkdir ANALYSIS
    cd ANALYSIS
    ```

2. Get the raw data. Move all of the raw data files (fastq format) into a single folder, for example " `example_rawdata`". You can provide either compressed (.fastq.gz) or uncompressed fastq files (.fastq). Usually, each fastq file represents one sample. These files should be given a name that reflects a proper sample name. The QWRAP program operates on all the files present in a given folder with an extension .fastq or .fastq.gz. The filename should not include an underscore (_) in it.

    For example, the file " `example_rawdata.tar.gz`" (present in the QWRAP directory) contains multiple fastq files (8 samples, 4 sample groups) stored in a single archive. Copy the file into the " `ANALYSIS`" directory, and uncompress it to extract the (individually compressed) fastq files for the 8 samples:

    ```
    tar xvf example_rawdata.tar.gz
    ```

    You can check the contents of the folder as

    ```
    ls example_rawdata/A1.fastq.gz A2.fastq.gz B1.fastq.gz
    B2.fastq.gz C1.fastq.gz C2.fastq.gz D1.fastq.gz D2.fastq.gz
    ```

    This dataset will be used as an example throughout the analysis protocol.

3. Initial quality check of the raw data. In this step, we run an initial data quality check on all files present in the raw data folder. If the raw data is located in the folder "example_rawdata", then, from the "ANALYSIS" directory, run the program as follows:

    ```
    quality_check_rawdata.sh example_rawdata
    ```

    This command calls the program FASTQC and runs it across all of the files present in the " `example_rawdata`" directory. This script creates the folder fastqc_rawdata containing FASTQ reports for all samples. Inside the folder, an

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

HTML file “FASTQC_overview.html” is created which provides an overview of the quality statistics for all samples, and also provides a more detailed report for individual samples. This report can be used to choose the best parameters for quality filtering. (See next step.)

4. Quality filtering of raw data. The quality filtering of raw data is performed in two steps.

   a. All the reads are trimmed to a user defined sequence length.

   b. Only those reads with a QScore >20 over at least 80% of the bases are retained.

The program “ `quality_filter_single.sh`” requires the location of the folder containing the raw data and the user-defined parameter that specifies the length of the final reads (trimming all reads to the specified length as necessary). To modify the QScore parameter for quality filtering step 2, edit the parameters QC_PERCENT and QC_SCORE in the program script “ `quality_filter_single.sh`” (located in the QWRAP folder). To trim the reads to a length of 200, execute the program as follows:

```
quality_filter_single.sh example_rawdata 200
```

If the user does not want to do trimming of reads, they can just provide the initial length of reads as the length parameter. The program does the quality filtering and outputs the filtered fastq files in a directory called “ `filtered_fastq`”. The program also creates fasta files for all samples in the current directory “ `ANALYSIS`” which are used for subsequent analyses.

5. Final quality check of filtered data. The program “ `quality_check_filterdata.sh`” uses the directory `filtered_fastq` to generate the quality report for the filtered dataset:

```
quality_check_filterdata.sh filtered_fastq
```

This script creates a folder `fastqc_filterdata` with FASTQC reports for all files of the filtered dataset. Inside the folder, the HTML file “ `FASTQC_overview.html`” is created which provides a combined overview of the quality statistics for all samples and also provides a more detailed report for individual samples.

6. Generating QIIME compatible input files and running QIIME scripts. All the fasta files present in the current directory (`ANALYSIS`) will be included in the analysis. Based on the quality report of the filtered dataset, if the user wants to remove certain samples from the analysis (perhaps because of low sampling depth), just delete the corresponding fasta file from the “ `ANALYSIS`” directory.

The program “ `microbiome-workflow1.sh`” runs inside the `ANALYSIS` directory and creates several files that are required for QIIME-based analysis:

```
microbiome-workflow1.sh
```

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

The program generates the following files:

a) seqs.fna: All fasta files present in a current folder are merged into this single fasta file named " `seqs.fna`". The sample name extracted from the name of the original fasta file becomes the header for each read to distinguish different samples. A snapshot of the file is provided below. Since read headers are not very informative, they are renamed as 1,2… to save disk space.

```
>A1_1
AACGTAGGTCACAAGCGTTGTCCGGAATTACTGGGTGTAAAGGG
AGCGCA
>A1_2
TACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGG
AGCGTA
…
>B1_1
TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGG
AGCGTA
>B1_2
AACGTAGGGTGCAAGCGTTGTCCGGAATTACTGGGTGTAAAGGG
CGCGCA
…
```

b) mapping.txt: A default mapping file is created with a single column containing sample names (following a header) as shown below.

```
#SampleID
#Mapping file for the QIIME analysis
A1
A2
B1
B2
C1
C2
D1
D2
```

Author Manuscript

The mapping file can/should be edited to separate the samples into experimental groups. For example, since these 8 samples belong to four groups (A, B, C, D) a new column called "Group" can be added as shown below. Each column is separated by a tab.

```
#SampleID	Group
#Mapping file for the QIIME analysis
A1	A
A2	A
B1	B
B2	B
C1	C
C2	C
D1	D
D2	D
```

c) sample_order.txt: This file is used to determines the ordering of samples in the OTU table. This is especially important as the same order is used for generating taxonomy bar charts. You can change the order of samples here to reorder them in OTU table.

```
A1
A2
B1
B2
C1
C2
D1
D2
```

d) script.sh: This contains all of the commands required to generate the OTU table and taxonomy bar chart. Most of the commands are from the QIIME package and are arranged in the order required to perform subsequent analyses. More details about these commands can be found at http://qiime.org/scripts/. Parameters used for each command can be edited within this file. An example of the "script.sh" file can be found in the QWRAP GitHub archive. In brief, the script performs the following functions:

a. Chimera filtering using the program "USEARCH"

b. Denovo OTU Clustering at 97% sequence similarity using the program "UCLUST".

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

c. Picking representative of OTUs based on abundance

d. Assigning taxonomy to OTUs using RDP classifier (threshold 0.8) and the Greengenes database (DeSantis et al., 2006).

e. Sorting the OTU table based on the file "sample_order.txt"

f. Summarizing OTUs into taxonomic groups

g. Creating a normalized OTU table by converting raw numbers from the original OTU table into percent abundance, and merging taxa information in a single file.

h. Creation of a filtered list of the top 10, 25 and 100 OTUs and taxa.

i. A new filtered OTU table (where OTU abundance > 0.0005%) is created with new taxonomy charts and summary statistics. This filtered file is used for all subsequent analysis.

j. .

Execute the file "script.sh" as

```
sh script.sh
```

This will execute all the commands present in the file and generate the OTU table and taxonomic charts.

For the analysis of alpha and beta diversity, another program " `microbiome-workflow2.sh`" is provided. This program generates a file called "script_adv.sh" which contains QIIME commands to perform several jobs including sequence alignment, phylogenetic tree generation, and the calculation of alpha diversity and beta diversity plots. An example of "script_adv.sh" file can be found in the GitHub archive. The script requires the user-defined parameter "sampling depth" (read depth). Since different samples may have different read depths, read depth should be normalized across all samples. When a read depth is provided, a random sampling event is used to rarify the OTU table. If the sampling depth is 22986 (in our example dataset), you can run the script as

```
microbiome-workflow2.sh 22986
```

If no sampling depth is provided, the program will automatically calculate the minimum sampling depth from the file " `otu_table.stats`"

This script performs the following functions:

a. Generation of a multiple sequence alignment of OTUs using the program "PYNAST" and creating a phylogenetic tree using the program "FASTTREE".

b. Rarify the OTU table based on minimum sampling depth.

c. Calculate alpha diversity (using chao1, observed_species, PD_whole_tree, shannon, and simpson diversity indicies) and generate plots

d. Calculate beta diversity (using bray_curtis, unweighted_unifrac, and weighted_unifrac distances) and generate plots

e. Generate a UPGMA tree of all samples

Execute the file "script_adv.sh" as:

```
sh script_adv.sh
```

7. Generating HTML report.

Although most of the results are generated in the `ANALYSIS` folder and can be explored manually, we have developed a script which will summarize most of the results into a user friendly HTML report. The HTML report is a static set of files so the whole `ANALYSIS` folder can be moved or shared with another person or the results can also be shared using a webserver.

Run the following program to generate the HTML report.

```
report_microbiome.sh
```

This creates the html file " `microbiome_report.html`" in the `ANALYSIS` directory and can be opened using any web browser.

The most updated version of the protocol (QWRAP-Readme.docx) is available at https://github.com/QWRAP/QWRAP. On the github page click the button "RAW" to view this file. This readme file also has instructions to perform paired end microbiome data analysis.

Reagents and Solutions

Cary-Blair Medium—Cary-Blair medium has been developed and used for transport of microbe samples. It consists of predominantly buffered salts with a non-existent nutrient content to retard the growth of the microbes in the fecal sample. We have modified the formula to add reducing agents to promote the survival of the anaerobic microbes during transport. The modified Cary-Blair media is as follows:

To 1 liter of autoclaved water, add with stirring:

Na thioglycolate 3.0 g

Na(H)phosphate 1.1 g

NaCl 5.0 g

Cysteine 5.0 g

Resazurin (oxygen indicator) 0.01 g

Stir until completely in solution. Dispense into 50mL conical tubes. Cap tubes securely and seal with Parafilm. Store at room temperature in the dark.

The solution should be clear. If the solution turns blue, it indicates oxygen and the solution should be discarded.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Commentary

Background Information—Up to the late 20th century, the study of microbes focused predominantly on those organisms that could be grown in the laboratory. To study an organism, it was necessary to grow an organism, which was accomplished by plating samples on specialized media selective for the growth of that organism and identifying features such as the morphological characteristics of colonies. However, it was clear that this approach had limitations to understand microbial diversity since there were a vast number of microbial species that could not be grown in the laboratory. The development of DNA-based culture-independent methods to analyze the DNA extracted directly from a sample rather than from individually cultured microbes, have resulted in an explosion of studies on the complexity of microbial communities (Arumugam et al., 2011; Peterson et al., 2009). These methods support the detection of unculturable organisms in addition to the culturable organism present in a sample.

Analysis of the data obtained from microbiome sequencing includes identification of all bacterial taxa present in the samples. Taxonomic detection is based on the sequencing of the variable regions of the 16S ribosomal RNA genes present in all prokaryotic organisms (Figure 1). Several software packages that can be used for microbiome data analysis are QIIME (Caporaso et al., 2010b; Lozupone et al., 2006; Navas-Molina et al., 2013), MOTHUR (Schloss et al., 2009), RDP Tools (Olsen et al., 1992), ARB (Ludwig et al., 2004), and VAMPS (Sogin et al., 2009). Depending on the sequencing protocol, the reads (DNA sequences) can be either single end reads or paired end reads. For microbiome analysis, single end reads are handled directly by most tools, however a number of analytical programs may not be able to directly use paired-end reads. Users should check the software manual before proceeding with paired end reads. If there is a predicted overlap among the forward and reverse end reads, then tools such as Usearch (Edgar, 2004), or FLASH (Magoc and Salzberg, 2011) can be used to merge paired reads into longer single reads, which then can be handled by most programs. Using longer reads helps to provide more accurate taxa information at higher taxonomic levels (genus, species).

Quality score information present in the FASTQ files is used for quality filtering. Quality filtering vastly improves the diversity estimates for samples processed through Illumina sequencing (Bokulich et al., 2013). Based on the quality of the reads, one can decide to filter the raw data to remove low quality reads and/or low quality bases. One can trim individual (or all) reads from the 3′ end where the quality drops below a certain Q score (e.g. a Q score of 30), or entirely remove the read if its overall quality falls below a certain threshold (i.e., the presence of a certain number of low quality bases containing the character “N”). Another quality control step is removal of chimeric sequences. Chimeras are sequences that are created in the PCR amplification process from two or more templates instead of a single parent template. Chimeric sequences can be removed with programs such as BLAST (Altschul et al., 1990), ChimeraSlayer (Haas et al., 2011), or UCHIME (Edgar, 2004).

Sequences that have passed the quality control filtering step are first clustered at a predetermined sequence similarity level, usually 97% to support their taxonomic classification. QIIME provides multiple algorithms and methods to perform OTU clustering

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(e.g. cd-hit (Li and Godzik, 2006), UCLUST (Edgar, 2010), and USEARCH (Edgar, 2010); additional methods can be found in (Navas-Molina et al., 2013). The resulting clusters of nearly identical sequences are referred to as **Operational Taxonomic Units (OTUs)**. An OTU table is generated that includes the count of the number of sequence reads that represent each OTU, along with the taxonomic classification of that OTU. This OTU table is used for subsequent analyses. The relative abundance of each taxon present in a set of samples can be determined from the OTU table and used to generate figures representing taxa abundance. These summary charts can be generated for each specific taxonomic level in the hierarchy (e.g., phylum, family, genus, and/or species) depending on the depth of the taxonomic assignment of OTU's. This abundance information can be presented in a variety of formats including bar charts and heat maps.

A phylogenetic tree of all identified OTU sequences can be used to assist in the visualization and clustering of microorganisms from samples. Several multiple sequence alignment programs are available such as PyNAST (Caporaso et al., 2010a), Infernal (Nawrocki et al., 2009), MUSCLE (Edgar, 2004) and clustalw (Larkin et al., 2007) to align the OTU sequences prior to phylogenetic inference. A phylogenetic tree can be generated from the aligned sequences using programs like FastTree (Price et al., 2009), clustalw (Larkin et al., 2007), clearcut (Evans et al., 2006), raxml (Stamatakis et al., 2005) or MUSCLE (Edgar, 2004).

A variety of statistical analyses can be performed to assess the differences in OTU distribution and abundance between samples and groups. Alpha diversity is a measure of diversity *within* a sample or environment. Different methods available for alpha diversity calculations can take into account factors such as richness (number of OTUs/species present in a sample) and evenness (relative abundance of different OTUs/species and their even distribution), (Jost, 2007; Lozupone et al., 2007). A few commonly used metrics include:

- **Observed species:** measures unique OTUs in the sample.
- **Chao1**: estimates the species richness.
- **Shannon's index:** measures both richness and evenness.
- **Simpson's index:** measures both richness and evenness, but less affected by the presence of rare species when compared to Shannon's index.
- **PD (Phylogenetic Distance):** includes phylogenetic distance into the diversity calculation.

In a typical microbiome experiment, different samples may have been sequenced at different sampling depth (number of reads per sample). A common question is whether or not the samples have been sequenced at sufficient depth to properly calculate the diversity. This question can be answered using rarefaction analysis. The term "rarefaction" refers to a random collection of sequences from a given sample, with a specified depth (number of sequences). The rarefaction curve plots the number of individuals sampled (at a particular depth) versus the alpha diversity. For each biological sample, a random sample is generated (without replacement) with 10%, 20%, 30%, up to 100% of all sequences, and alpha diversity is calculated for each random sample. When the curve starts to level off, the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

assumption is that you've attained the maximum alpha diversity (depending on the specific metrics used) for that sample. If the curve does not level off, then the sample(s) has not been sequenced to a sufficient depth. Due to sequencing depth variation across all samples, it is advised that the diversity analysis and comparisons be performed and reported at an equivalent depth of sequencing across all samples.

Beta diversity provides a measure of the distance or dissimilarity between each sample pair. When more than two samples are used, the beta diversity is calculated for every pair of samples to generate a distance/dissimilarity matrix. Various metrics have been developed to calculate beta diversity and each has its own advantages and disadvantages. A few of the commonly used beta diversity metrics are:

- **Bray-Curtis:** Non-phylogeny based method that takes abundance into account;
- **Un-weighted UniFrac**: Uses the presence and absence of OTUs between samples along with their phylogenetic distances to determine beta diversity;
- **Weighted UniFrac:** Uses the abundance information for each OTU along with their phylogenetic distances to determine beta diversity.

Principal Coordinates Analysis (PCoA) can be used for visualization of the data present in the beta diversity distance matrix in the form of 2-Dimensional or 3-Dimentional plots known as PCoA plots, which can reveal various clustering patterns present within the samples. PCoA transforms the distance matrix into a new set of orthogonal axes where the first axis (usually called PC1) can be used to explain the maximum amount of variation present in the dataset, followed by the second axis (PC2), and so on. In a PCoA 2D plot, the first two principal components (PC1 and PC2) are used whereas in a 3D plot the first three components (PC1, PC2 and PC3) are used for plotting. The output from the PCoA analysis can be used to generate trees using clustering methods such as hierarchical clustering.

One of the goals of the human microbiome project, funded as an initiative of the NIH Roadmap for Biomedical Research, was to determine whether there were associations with changes in the microbiome and health and disease (Peterson et al, 2009). The most studied and complex environment is the human gut that is colonized by microbes that outnumber the number of cells in the host by 10 fold (Arumugam et al., 2011; Eckburg et al., 2005; Faith et al., 2013; Huttenhower et al., 2012; Peterson et al., 2009). In most mammals, the gut microbiome is dominated by three bacterial phyla: *Firmicutes, Bacteriodetes, and Actinobacteria*. Although the composition of bacterial phylum in humans is similar, there is variation in species composition within the human population due to differences in diet and health status of the human host (Arumugam et al., 2011; Lozupone et al., 2012; Peterson et al., 2009; Turnbaugh et al., 2009a; Wu et al., 2011). The gut microbiome has been shown to have a major impact on host metabolism and homeostasis (Cho and Blaser, 2012; Ley, 2010). In particular, studies have linked an imbalance of the gut microbiome (dysbiosis) with energy utilization from diet, development of obesity and possibility diabetes (Musso et al., 2011; Tilg and Kaser, 2011; Turnbaugh et al., 2006). Central to the involvement of the gut microbiome with these disorders is the realization the interaction between the gut microbes and the host immune system is critical in the development of the immune response and the modulation of chronic inflammation (Littman and Pamer, 2011). Given the role of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

inflammation in the development of cancer and chronic disease (Coussens and Werb, 2002), it is not surprising that links are being discovered that point to a significant role in gut dysbiosis with the development of these disorders. Thus, we anticipate that many alterations of the gut microbiome (as well as other site microbiomes) will be seen as investigators begin to apply these analysis techniques to a variety of new diseases as well as in depth studies on normal homeostasis.

One of the exciting aspects of microbiome research that we have observed from our experience has been the promotion of translational research linking basic research. The general pipeline we have established allows investigators to collect the samples either newly acquired or archived, to be used for microbiome analysis. This allows involvement of clinical researchers that have access to patients and samples they may have collected over time, to also participate in the analysis.

Critical Parameters for Success of Analysis/Trouble Shooting/Costs

Sample Acquisition—Sample acquisition is the first and probably most important step in the microbiome analysis. Collecting, processing and archiving of samples is not the most exciting aspect of the analysis, but careful and thorough collection methods are crucial for a high impact study. One of the challenges in human microbiome work is to convince people to participate. This is especially true of studies that require fecal (stool) samples. Providing a private environment to discuss and answer questions with the participant about sample collection is vital to overcoming this unease. Likewise, when samples are collected in the clinical setting, privacy should be provided. For instance, a saliva sample can essentially be collected anywhere, although few people are comfortable spitting into a tube in front of others in a waiting room.

Undoubtedly, fecal samples induce the most hesitation in participants. Below are factors the investigators should be aware of and modify as needed to ameliorate hesitation to participate.

- Providing a kit with clear instructions, items for hygienic collection and disposal, and a contact person for issues with the kit or shipping the sample
- Recognize and encourage strong family/caregiver support, especially in diseases requiring assistance with activities of daily living
- Assure the participant FedEx is accustomed to all types of clinical shipments, including fecal samples, and the delivery person does not know the shipment is stool instead of blood or any other type of clinical sample

Collection of sample metadata—A microbiome analysis alone is insufficient to fully appreciate the interactions between the microbes and the host in normal homeostasis and disease. It is important then, to obtain additional patient information that can put the results from the microbiome analysis in the proper perspective. In fact, some of this information might be used to exclude patients from a study because of a known impact on the microbiome that could complicate the subsequent understanding of the results. Of course, Institutional Review Board approval should be sought not only for collection of the samples,

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

but also for the collection of additional information (i.e. diagnosis reported by participant or from medical chart review). Some of this information that might be collected includes:

- Age, sex, weight, body mass index (BMI)
- History of chronic diseases and recent acute illnesses (i.e. cold, stomach flu)
- Diet
    - Best method – food records of every food and beverage consumed for 3–7 days
    - Next best – Food Frequency Questionnaire (FFQ) for the population of interest
    - Minimum – a questionnaire such as below should be used to capture dietary habits or recent changes.
- Medications with a particular focus on ones which cause diarrhea (antibiotics, chemotherapy) – This is particularly important for the gut microbiome since it is known that antibiotics in particular can alter the microbiome.
- The use of diet supplements, including probiotics, fiber/prebiotic supplements, and herbs

A sample information form is provided in Figure 7. While it might not be important to obtain all information on the form (it can be amended according to the study protocol), the general "rule of thumb" is that it is easier to ask for the information before the analysis rather than try to re-connect with the patient after the study where there might be gaps in memory.

Finally, the methods we have described for acquisition of fecal samples will allow the investigator to archive samples that can be used in future studies with gnotobiotic/germ free animal models to study the impact of complex microbial populations on host homeostasis and disease. These germ free animal facilities are predominantly for rodents, but microbiome transplantation has also been accomplished for other model systems such as zebra fish, drosophila and *C. elegans* (Goodman et al., 2011; Tlaskalova-Hogenova et al., 2011). The power of linking microbiome transplantation with animal models has been highlighted in recent publications using transplanted "human" microbiota into gnotobiotic animals to study the impact of diet on obesity (Faith et al., 2013; Turnbaugh et al., 2009b).

Multiplexed 16S amplicon sequencing on the MiSeq system—To establish an "in house" microbiome analysis, it is critical to have access to the appropriate NextGen sequencing machine. As the interest in NextGen sequencing increased, new machines that are more cost effective such as the Illumina MiSeq, which can be used for high speed, multiplexed 16S amplicon sequencing. In addition, barcoded PCR amplification using degenerate primers originally taken from Caparoso et al. (2011) have been described by Kozich et al. (2013) for use on the MiSeq (Caporaso et al., 2011; Kozich et al., 2013). Finally, a recent study has evaluated the *cost per analysis* and found that a MiSeq platform has greater potential for use with 16S rRNA gene sequence studies because it generates

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

longer sequence reads while balancing the overall costs making it reasonable for investigators (Kozich et al., 2013).

Although the MiSeq is a very user-friendly machine in our hands, one of the important considerations for proper sequencing is cluster generation. "Clusters" refers to the number of sites on the sequencing flow cell that contain a unique sequence molecule or read. The numbers of clusters are important because the optimum cluster density will maximize the number of reads obtained from the set of multiplexed samples loaded onto the cell, providing the optimum sample read depth needed to support bioinformatics analysis. The MiSeq version 2 and version 3 cluster kits are design to provide approximately 1.2 million clusters/mm2. We generally obtain between 70,000 to 120,000 reads per sample for a 96-sample analysis using conditions described in Section 2.

There are several factors that can contribute to reduction of cluster densities. First, during PCR and subsequent gel purification steps, it is important that the amplification results in a single band of approximately 250 base pairs. Although the PCR primers are in excess it may be beneficial to reduce the amount of primers so they become limiting, thus ensuring all the primers are used. Free primers can become entangled with the PCR product when running an agarose gel. These free primers will attach to the flow cell and generate non-productive clusters, which will reduce cluster density of the target amplicons.

Second, proper cluster generation requires accurate quantitation of the DNA (i.e. Pico Green assay).

Third, improper denaturation will reduce the number of molecules available for cluster generation. It's important to use between 0.15N–0.2N NaOH in the denaturation process. One can also heat the samples to 95°C following NaOH denaturation to ensure the DNA is single stranded prior to addition to the flow cell. Conversely, using NaOH at too high a concentration will destroy the DNA and render it unable to cluster.

Fourth, it is very important that the hybridization buffer be kept on ice prior to addition to the denatured libraries. Similarly, the diluted, single stranded libraries should also be kept on ice and used within a 2-hour period to inhibit the reformation of double strand amplicons.

Fifth, if one has applied the solutions above and under-clustered flow cells are still obtained, then it is possible to increase the pM amount of DNA that is added. For example, it is suggested that a final concentration of mixed libraries plus PhiX are at 12.5pM. It may be necessary, in some instances, to increase the amount of libraries to 15pM or 20pM. This must be done experimentally as it is impossible to know, a priori, if 12.5pM DNA will result in low cluster numbers.

Ensuring successful installation of QWRAP and other tools required for analysis—Instructions concerning the installation and testing of each of the following required tools can be found on their respective web sites. To ensure that all of the required tools are working, the following commands can be executed from within the "ANALYSIS" folder. If a particular tool is not accessible an error message similar to "command not found" will be obtained.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- FASTQC (provides program version number)

```
fastqc -v
```

- FASTX (provides help documentation)

```
fastx_trimmer -h
```

- QIIME (print the configuration of QIIME package)

```
print_qiime_config.py
```

- Usearch (provides program version number)

```
usearch61
```

- QWRAP

```
check_qwrap.sh
```

If you receive any error then it's possible that the location of the program is not present in your PATH variable. You can check the list of all folders accessible in your PATH variable using following command.

```
echo $PATH | tr ":" "\n"
```

If the program is not in your path, you need to follow the instructions for modifying your PATH variable provided above in the section on installing QWRAP and other dependencies.

Anticipated Results

To illustrate a microbiome analysis, we present the analysis of four human fecal samples designated as A, B, C and D. The total DNA from each of these samples was isolated and, each sample was processed in duplicate (e.g. A1 and A2) in separate PCR reactions; each PCR product was then sequenced using the Illumina MiSeq.

Taxonomy charts—An example of the taxa distribution at the phylum level for the 8 human fecal samples (A1, A2, B1, B2….) is presented in Figure 8. In addition to the bar graphs, heat-maps and correlation plots are also commonly used to graphically represent the microbial abundance data present in the OTU tables. An example of a heat-map is presented in Figure 9.

Alpha diversity and rarefaction charts—In Figure 10, the Shannon diversity index was used to compare within sample diversity; higher Shannon numbers indicate greater diversity within the sample. As can be seen, the D1/D2 samples exhibit lower Shannon diversity than the A1/A2 samples, consistent with what we observed from the bar graphs.

Beta diversity and Principal coordinates analysis (PCoA)—The beta diversity analysis was performed on four pair of fecal samples using un-weighted UniFrac distances and PCoA plots (Figure 11) were generated. The 3-Dimensional PCoA plot shows the distinct clustering of samples from the four different groups. However, the duplicate samples (e.g. A1 and A2) generally clustered close together; in fact the A1 and A2 cluster so

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

close that the two red spheres were not separately resolved on this picture. In contrast, the C1 and C2 samples cluster farther from each other (than the others) but still clearly cluster separately from samples A, B or D.

Time Considerations

The time considerations for sample acquisition can vary with respect to the sample. For example, oral and vaginal samples can be collected during the same clinic visit, while this may not be possible for fecal samples. Once the samples have been obtained, the process of DNA isolation can usually be accomplished in a few hours for 12 to 24 samples. The PCR reactions, gel electrophoresis and isolation take four to six hours for the same 12–24 samples. Multitasking can reduce the overall time needed to acquire the 96 samples for a complete run on the MiSeq. The actual DNA sequencing takes about 48 hours to complete (for paired end 250 base pairs), with an additional few hours for FASTQ conversion. The bioinformatics analysis can take several hours to several days depending on the dataset and availability of computer resources.

In our experience, once the complete pipe line has been established, the entire process from sample collection to complete analysis will take between 10 days to two weeks with one person for sample preparation/PCR (including isolation), one person for DNA quantitation/ MiSeq and one person for bioinformatics. (Note that the analysis time includes time spent waiting for the various computational tasks to be completed by the computer performing the analysis.)

Acknowledgments

The following are acknowledged for their support of the Microbiome Resource at the University of Alabama at Birmingham: School of Medicine, Comprehensive Cancer Center (P30AR050948), Center for AIDS Research (5P30AI027767), Center for Clinical Translational Science (UL1TR000165) and Heflin Center. RBL acknowledges support from the Training Program in Brain Tumor Biology (2T32NS048039). We thank Adrienne Ellis for preparation of the manuscript.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990; 215:403–410. [PubMed: 2231712]

Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Dore J, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Merieux A, Melo Minardi R, M'Rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P. Enterotypes of the human gut microbiome. Nature. 2011; 473:174–180. [PubMed: 21508958]

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods. 2013; 10:57–59. [PubMed: 23202435]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal. 2012; 6:1621–1624. [PubMed: 22402401]

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(Suppl 1):4516–4522. [PubMed: 20534432]

Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nature Reviews. Genetics. 2012; 13:260–270.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research. 2009; 37:D141–145. [PubMed: 19004872]

Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002; 420:860–867. [PubMed: 12490959]

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology. 2006; 72:5069–5072. [PubMed: 16820507]

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Diversity of the human intestinal microbial flora. Science (New York, NY). 2005; 308:1635–1638.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research. 2004; 32:1792–1797. [PubMed: 15034147]

Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics (Oxford, England). 2010; 26:2460–2461.

Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. The long-term stability of the human gut microbiota. Science (New York, NY). 2013; 341:1237439.

Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G, Gordon JI. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:6252–6257. [PubMed: 21436049]

Haas BJ, Garges S, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methe B, DeSantis TZ, Consotium HM, Petrosino JF, Knight R, Birren BW. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 2011; 21:494–504. [PubMed: 21212162]

Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, et al. Structure, function and diversity of the healthy human microbiome. Nature. 2012; 486:207–214. [PubMed: 22699609]

Jost L. Partitioning diversity into independent alpha and beta components. Ecology. 2007; 88:2427–2439. [PubMed: 18027744]

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and Environmental Microbiology. 2013; 79:5112–5120. [PubMed: 23793624]

Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG, Knight R. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Current Protocols in Bioinformatics/editoral board, Andreas D. Baxevanis … [et al.]. 2011:17. Chapter 10: Unit 10.

Ley RE. Obesity and the human microbiome. Current Opinion in Gastroenterology. 2010; 26:5–11. [PubMed: 19901833]

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics (Oxford, England). 2006; 22:1658–1659.

Littman DR, Pamer EG. Role of the commensal microbiota in normal and pathogenic host immune responses. Cell Host & Microbe. 2011; 10:311–323. [PubMed: 22018232]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Applied and Environmental Microbiology. 2007; 73:1576–1585. [PubMed: 17220268]

Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature. 2012; 489:220–230. [PubMed: 22972295]

Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics (Oxford, England). 2011; 27:2957–2963.

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME journal. 2012; 6:610–618. [PubMed: 22134646]

Musso G, Gambino R, Cassader M. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. Annual Review of Medicine. 2011; 62:361–380.

Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, McMurdie PJ, Vazquez-Baeza Y, Xu Z, Ursell LK, Lauber C, Zhou H, Song SJ, Huntley J, Ackermann GL, Berg-Lyons D, Holmes S, Caporaso JG, Knight R. Advancing Our Understanding of the Human Microbiome Using QIIME. Methods in Enzymology. 2013; 531:371–444. [PubMed: 24060131]

Nechvatal JM, Ram JL, Basson MD, Namprachan P, Niec SR, Badsha KZ, Matherly LH, Majumdar AP, Kato I. Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. Journal of Microbiological Methods. 2008; 72:124–132. [PubMed: 18162191]

Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M. The NIH Human Microbiome Project. Genome Research. 2009; 19:2317–2323. [PubMed: 19819907]

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. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology. 2009; 75:7537–7541. [PubMed: 19801464]

Stackebrandt E, Goebel BM. Taxonomic Note: A Place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bact. 1994; 44:846–849.

Tilg H, Kaser A. Gut microbiome, obesity, and metabolic dysfunction. The Journal of Clinical Investigation. 2011; 121:2126–2132. [PubMed: 21633181]

Tlaskalova-Hogenova H, Stepankova R, Kozakova H, Hudcovic T, Vannucci L, Tuckova L, Rossmann P, Hrncir T, Kverka M, Zakostelska Z, Klimesova K, Pribylova J, Bartova J, Sanchez D, Fundova P, Borovska D, Srutkova D, Zidek Z, Schwarzer M, Drastich P, Funda DP. The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. Cellular & Molecular Immunology. 2011; 8:110–120. [PubMed: 21278760]

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. A core gut microbiome in obese and lean twins. Nature. 2009a; 457:480–484. [PubMed: 19043404]

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006; 444:1027–1031. [PubMed: 17183312]

Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Science Translational Medicine. 2009b; 1:6ra14.

Woese CR, Gutell RR. Evidence for several higher order structural elements in ribosomal RNA. Proceedings of the National Academy of Sciences of the United States of America. 1989; 86:3119–3122. [PubMed: 2654936]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. Linking long-term dietary patterns with gut microbial enterotypes. Science (New York, NY). 2011; 334:105–108.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

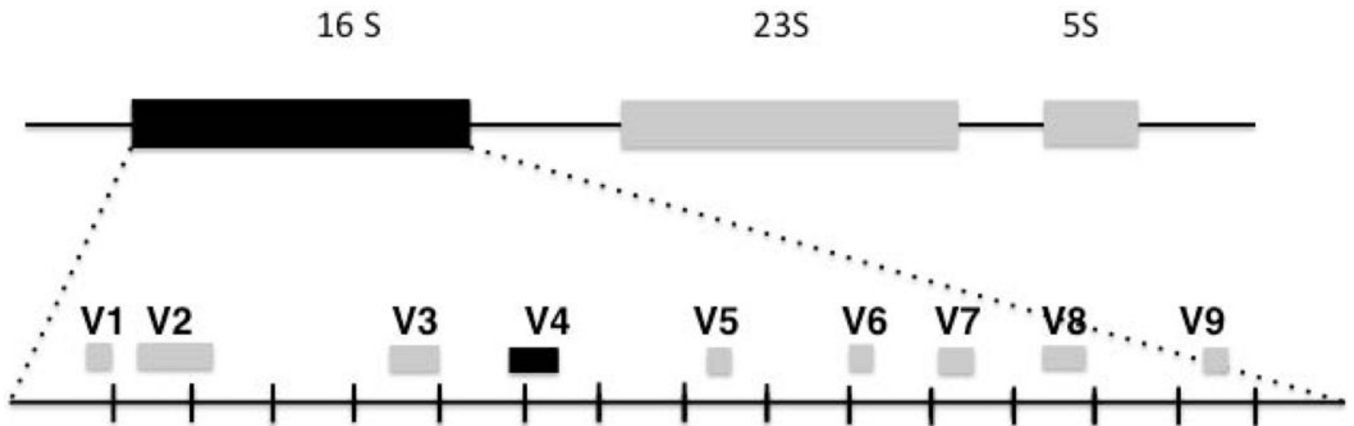
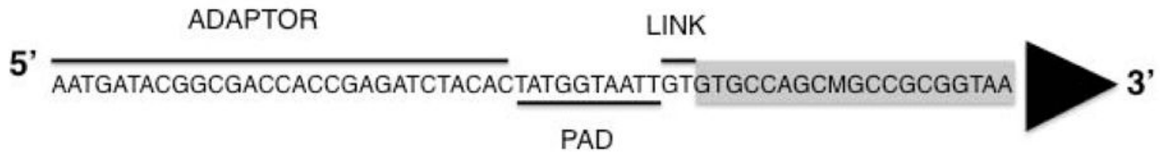


Figure 1.
The 16 S, 23S and 5S rRNA gene. The expanded 16S gene depicts the variable regions. The target V4 region for microbiome analysis is depicted in black.

CTTCCACTTAAATGAGACTTGTGCCAGCMGCCGCGGTAA......ATTAGAWACCCBDGTAGTCCATACAGGTGAGCACCTTGTA...
GAAGGTGAATTTACTCTGAACACGGTCGKCGGCGCCATT.....TAATCTWTGGGVHCATCAGGTATGTCCACTCGTGGAACAT...

5’ PCR PRIMER



3’ PCR PRIMER

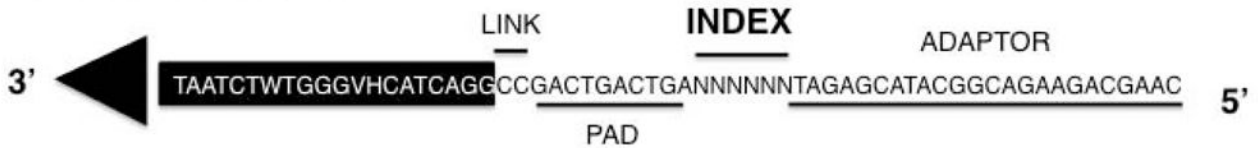


Figure 2.

V4 region and 5′ and 3′ PCR primers. The target nucleotides for the PCR primers are depicted in grey (5′region) and black (3′region). The 5′ and 3′ primers include adaptor Illumina sequences for NexGen sequences and “pad “sequences to facilitate PCR amplification. The 3′ primer also contains 6 base index primers (see Figure 3).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	A		B		C		D		E
1	ATCACG	23	CCACTC	44	ATTATA	65	TGTCGA	86	TGGAGT
3	TTAGGC	24	GCTACC	45	GAATGA	66	TGGCTC	87	TACCGC
4	TGACCA	25	ATCAGT	46	TCGGGA	67	GCGACA	88	TGTGCG
5	ACAGTG	26	GCTCAT	47	CTTCGA	68	TCGAGC	89	GATTAT
6	GCCAAT	27	AGGAAT	48	TGCCGA	69	TCGATA	90	GCCTAG
7	CAGATC	28	CTTTTG	49	TCTCGT	70	CATGCT	91	GATGTA
8	ACTTGA	29	TAGTTG	50	TACAGC	71	TCGTAC	92	ACTCCT
9	GATCAG	30	CCGGTG	51	GGTGAT	72	GTGGTG	93	GTCACG
10	TAGCTT	31	ATCGTG	52	CACGAC	73	TAGTAT	94	GCGAGC
11	GGCTAC	32	TGAGTG	53	TCCGCT	74	TGCGCT	95	GTAGAT
12	CTTGTA	33	CGCCTG	54	TATGCA	75	ATGGCT	96	ACTTGG
13	AGTCAA	34	GCCATG	55	TCGACT	76	GTTCTC		
14	AGTTCC	35	AAAATG	56	ATCTAC	77	CGTAAG		
15	ATGTCA	36	TGTTGG	57	ACGTTA	78	GCGTTC		
16	CCGTCC	37	ATTCCG	58	GATCTG	79	GTTGTT		
17	GTAGAG	38	AGCTAG	59	GAGCTG	80	GGACTT		
18	GTCCGC	39	GTATAG	60	TATCGC	81	CTCACA		
19	GTGAAA	40	TCTGAG	61	TGCTCG	82	CTGCTA		
20	GTGGCC	41	GTCGTC	62	ACGACT	83	TGACAT		
21	CGAAAC	42	CGATTA	63	ATATGC	84	TGTAAC		
22	CGTACG	43	GCTGTA	64	CTCGTA	85	AGCAGA		

Figure 3.

Listing of 96 6 base pair index sequences. For location of the 6 base pair sequences see Figure 2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

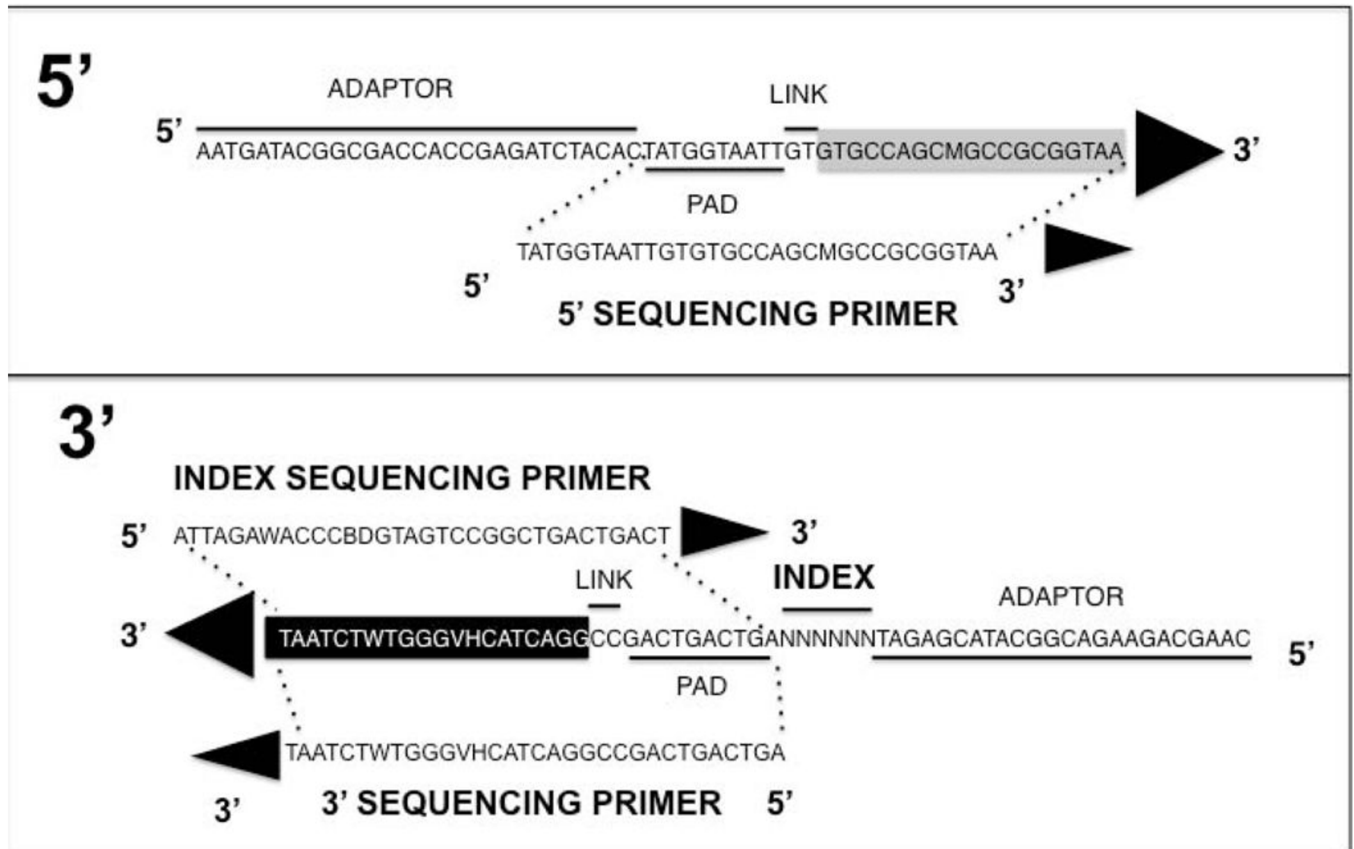


Figure 4.

Sequencing primers for 16S PCR product. The 5′ and 3′ sequencing primers used for MiSeq sequencing are depicted. The “index” sequencing primer is also depicted.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

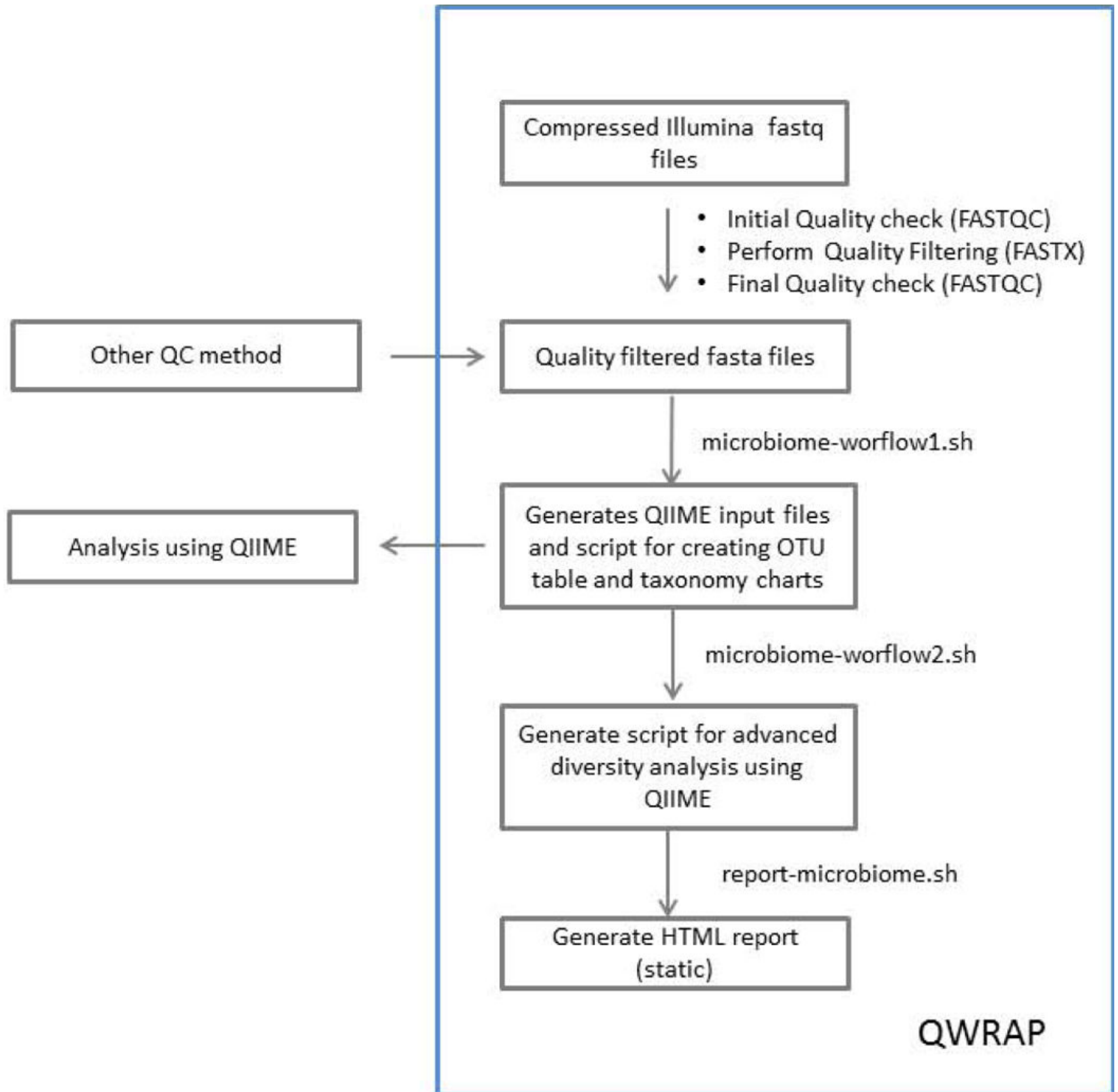


Figure 5.
The QWRAP workflow for data analysis.

Author Manuscript

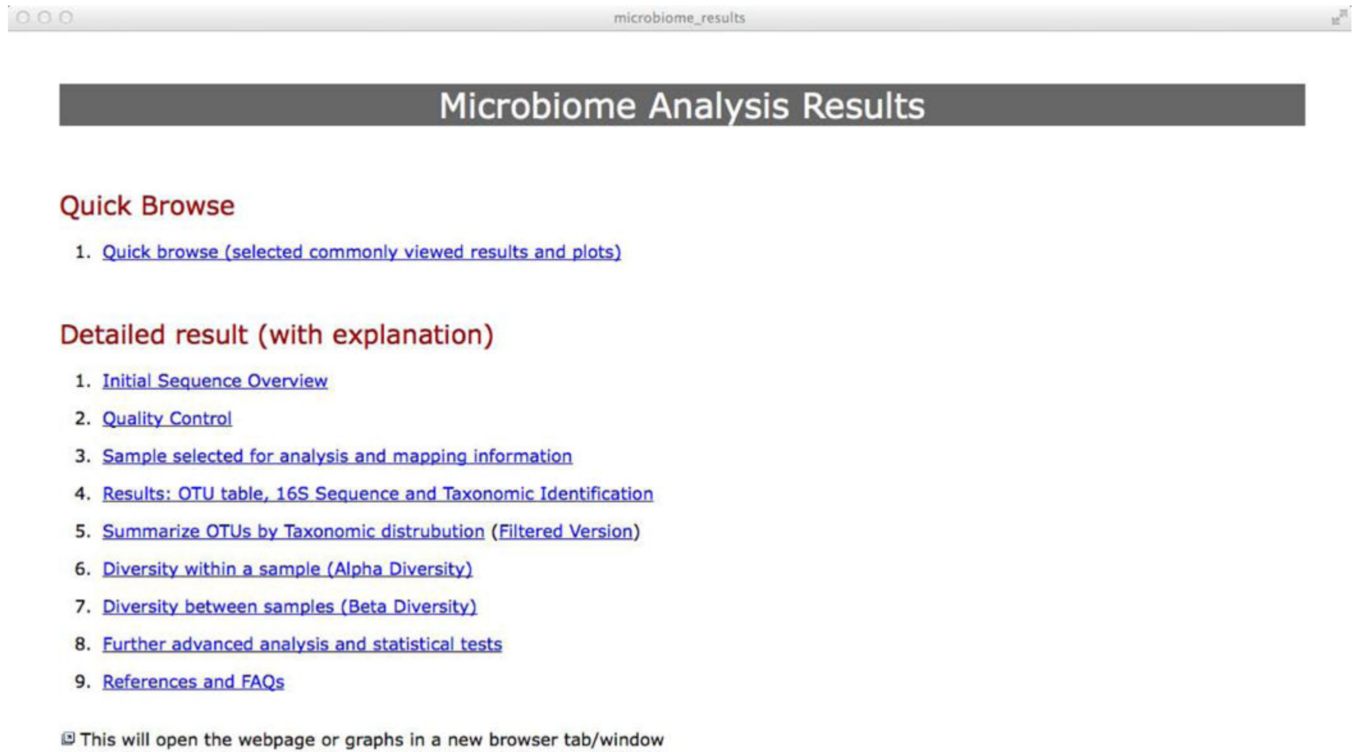


Figure 6.
Sample Microbiome Analysis Report from QWRAP.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sample Information Form

Please place this completed form in the outside pocket of the biohazard bag

on white vial: ________________________________

Date of stool collection: ___________________________

Time of stool collection: ____________________am / pm

Stool consistency: ☐ Formed ☐ Loose/Soft ☐ Watery ☐ Bloody

Questions

Your answers to the following questions will help us better understand any changes we see in your gut bacteria. Please circle your response.

In the last 72 hours (3 days)…		
Overall, has your normal diet or eating pattern changed	Yes	No
Have you had nausea or vomiting?	Yes	No
Have you had diarrhea?	Yes	No
Have you had constipation?	Yes	No

In the past week…		
Overall, has your normal diet or eating pattern changed'	Yes	No
Has your intake of yogurt or probiotic supplements (Align, Philip's Colon Health) changed?	Yes	No
Has your intake of fiber (found in fruits, vegetables, beans, and whole grain products) or fiber supplements (Metamucil, etc., including cereals such as Fiber One, All-Bran) changed?	Yes	No
Have you been on antibiotics?	Yes	No
Have you started any new vitamins?	Yes	No
Have you had chemotherapy?	Yes	No
Have you had radiation therapy?	Yes	No

Over the past 6 months, how many times in a typical week did you eat…			
Fried chicken, fish, seafood, or meat?	0-3	4-7	more than 7
Fried vegetables (including French fries)?	0-3	4-7	more than 7
High fat dairy products (whole milk and regular fat ice cream, cheeses, and sour cream)?	0-3	4-7	more than 7
Over the past 6 months, how many times on a typical day did you eat…			
A serving of food that is a good source of fiber (fruits, vegetables, beans, and whole grain products fiber supplements (Metamucil, etc., including cereal as Fiber One, All-Bran)?	0-3	4-7	more than 7

Do you have Diabetes?	Yes	No
If yes, Type I or Type II?	Type I	Type II
If Type II, is it well controlled? Are your numbers within the goals set by your doctor?	Yes	No
If diabetic, in the morning before eating, is you blood sugar usually 125 mg/dl and above or below 125 mg/dl?	Above	Below

Include contact information for questions and assistance with collecting or shipping the sample.

Figure 7.

Sample Information Form. A template form given to patients for donation of sample for microbiome analysis.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

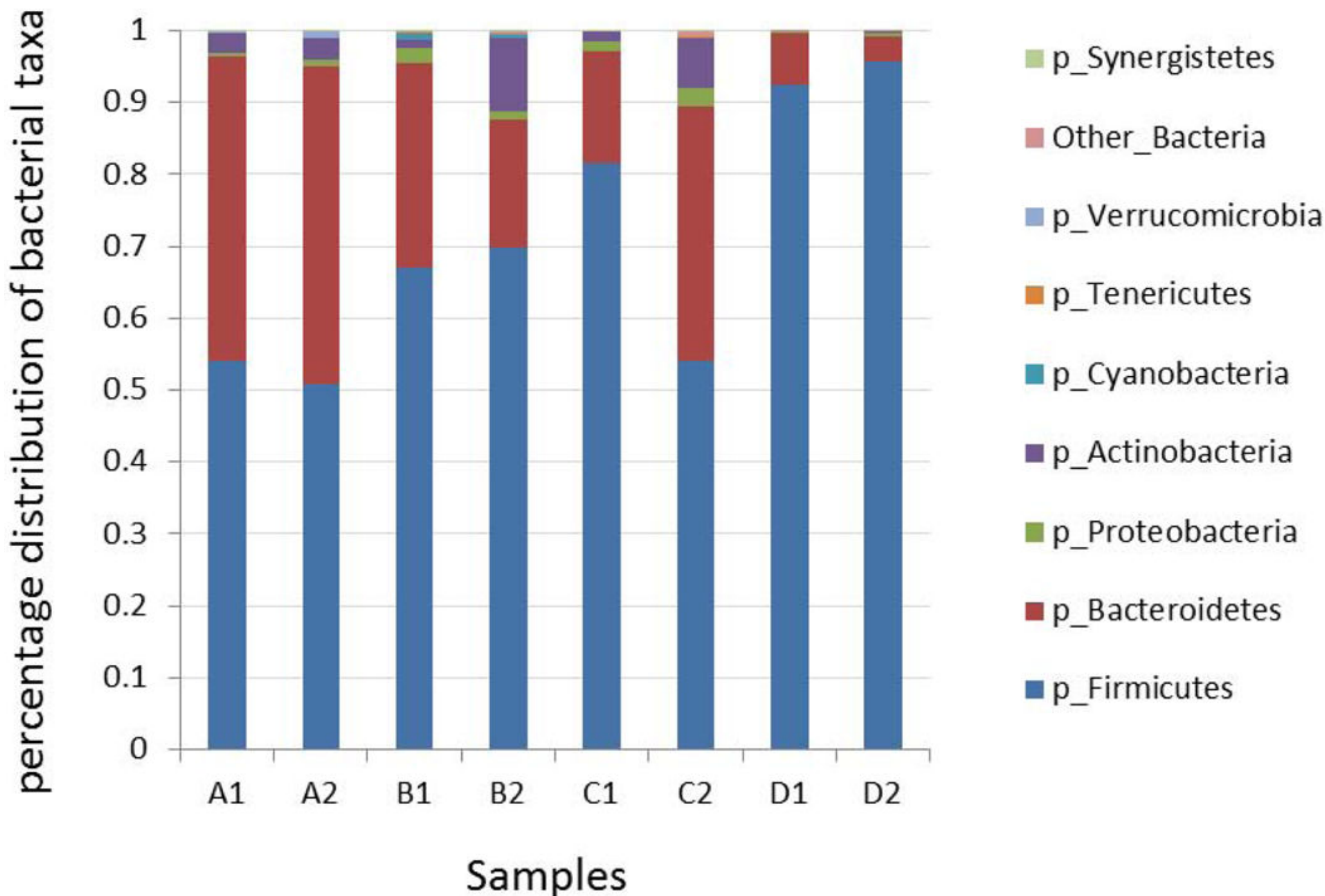


Figure 8.

Bar graph depicting bacterial taxa at the phylum level. The eight samples (A1, A2, B1, B2, C1, C2 and D1, D2) are shown. The figure shows a taxa summary chart showing distribution of top ten bacterial taxa (at phylum level) for 8 human fecal samples.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

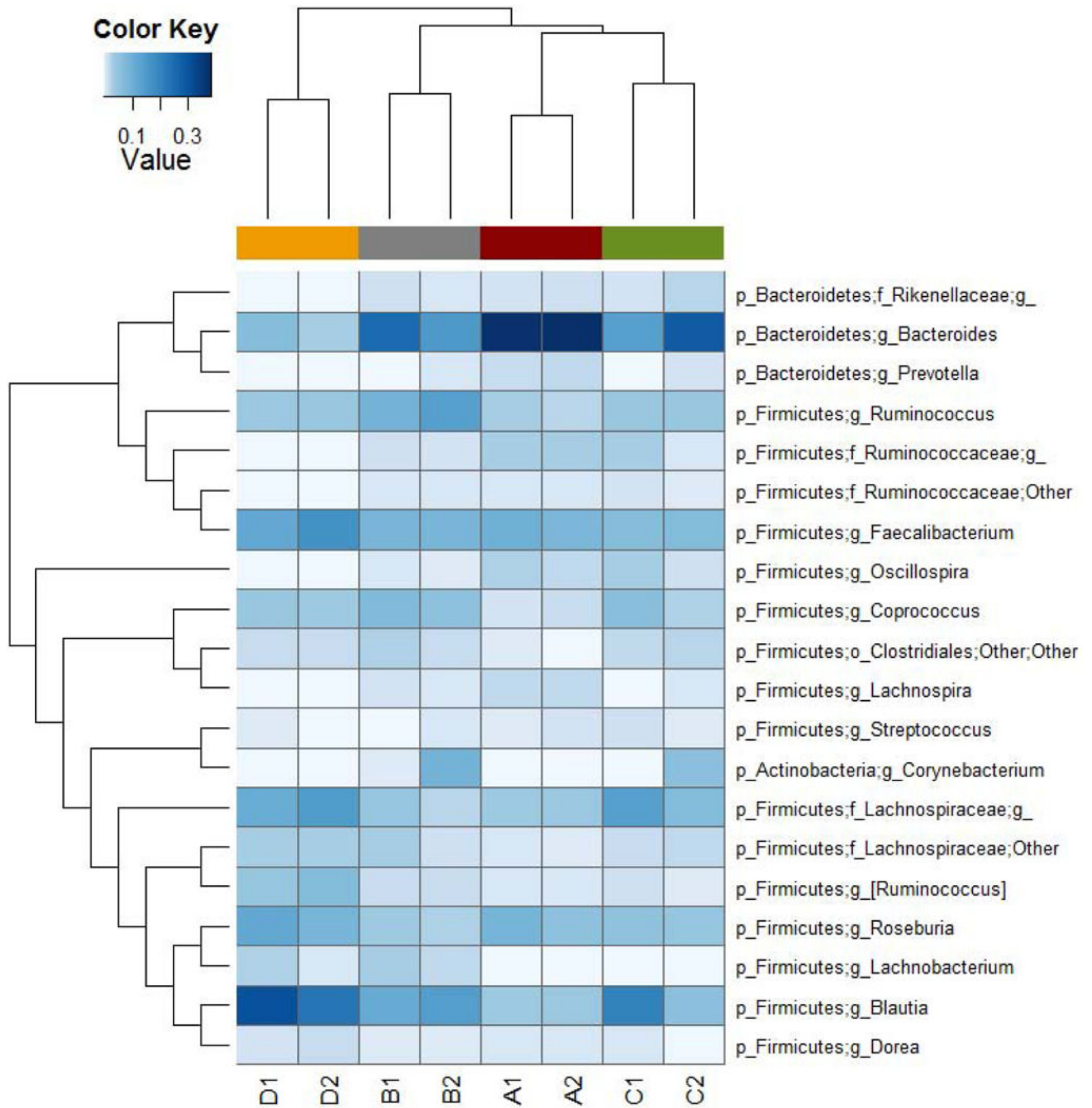


Figure 9.

Heat map. The four samples (A1, B1, C1 and D1) are shown. The darker color in the map represents greater abundance of the microbe. On the right the microbe at the phylum level is identified.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

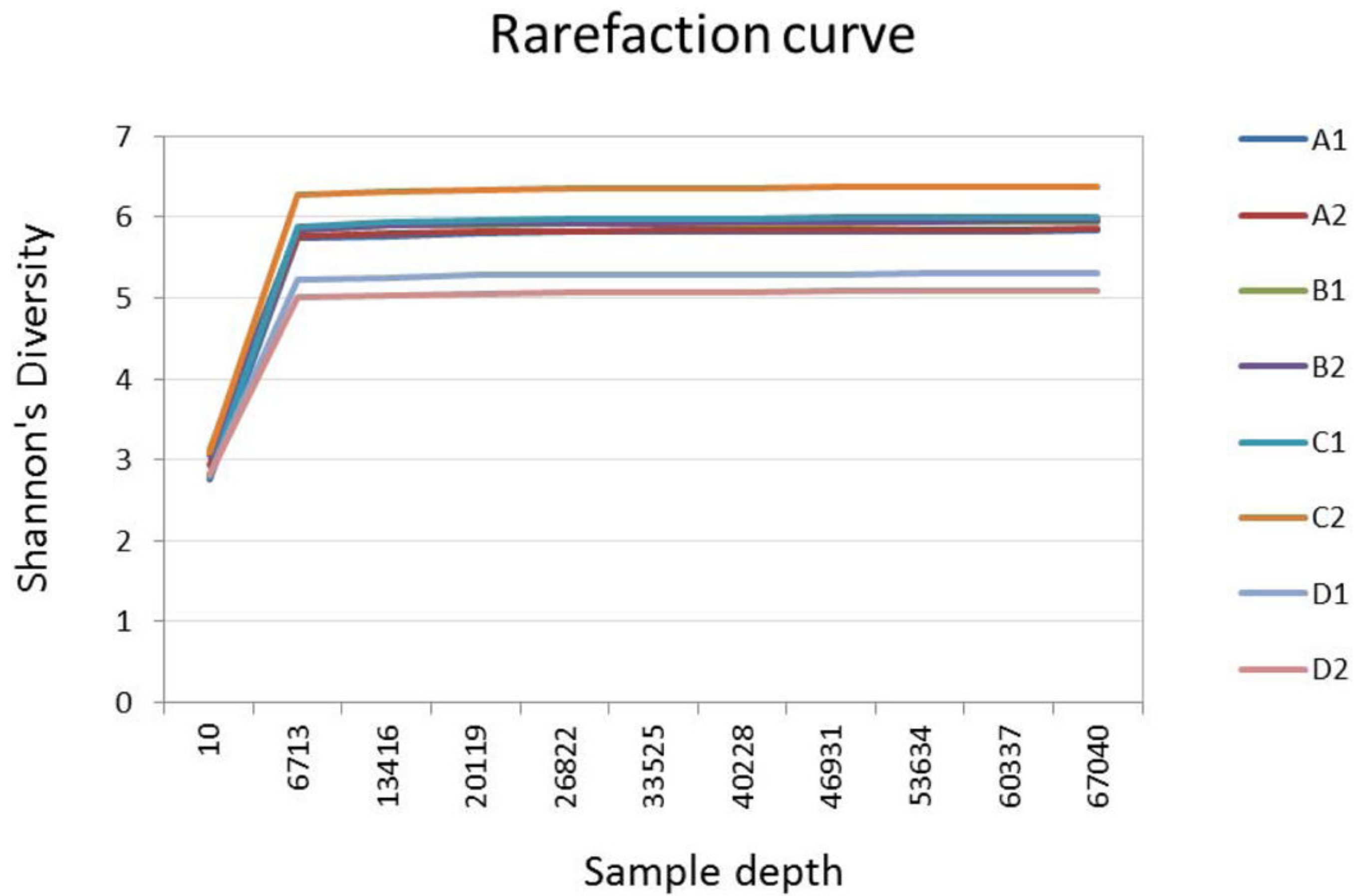


Figure 10.

Rarefaction curve for 8 fecal samples using Shannon's diversity. The samples with the highest Shannon values corresponding to samples with most diversity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

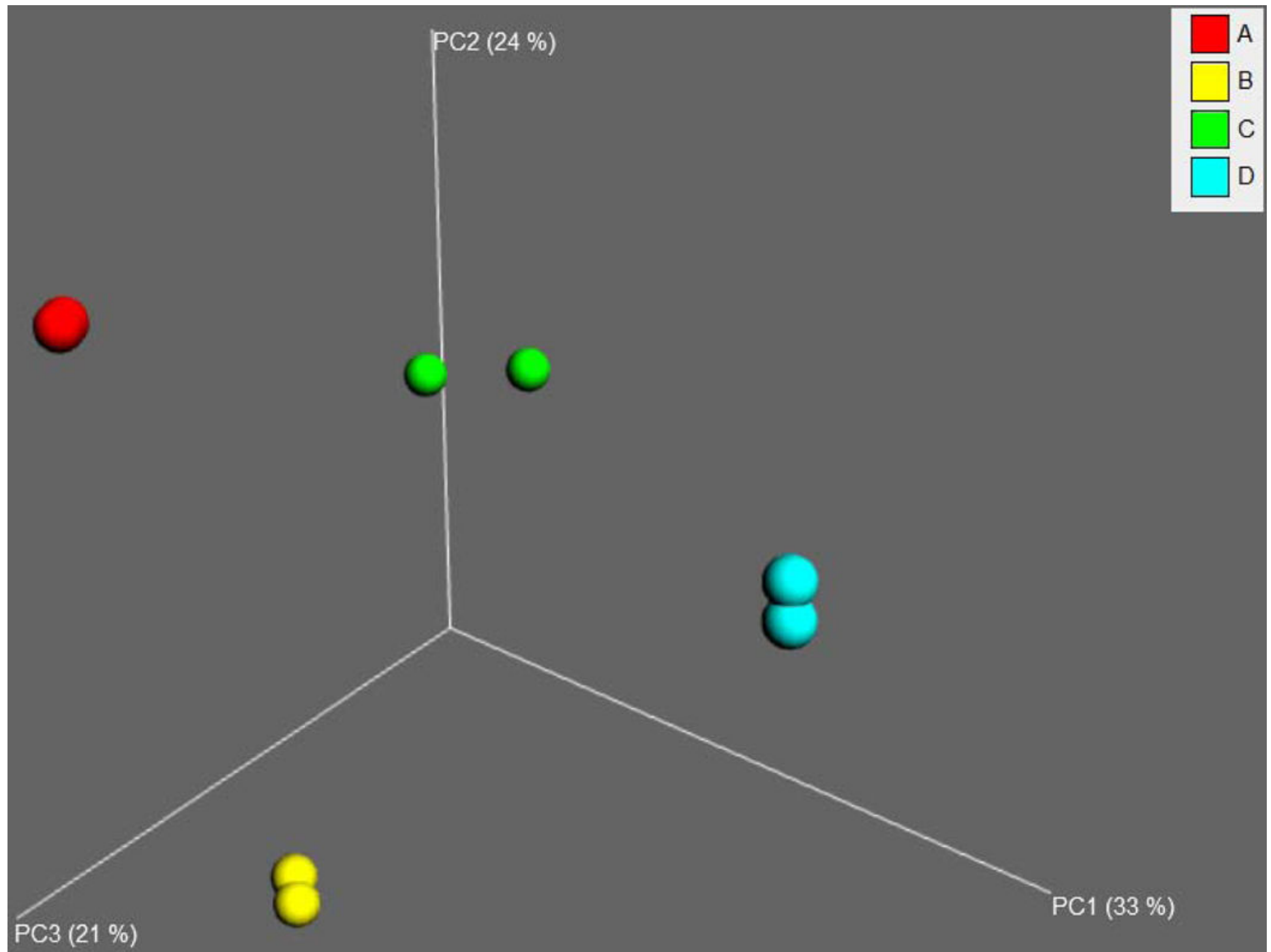


Figure 11.

Principal Coordinate Analysis plot. The eight samples (A1, A2, B1, B2 etc.) are presented in the PCoA analysis; note that the sample designated as A consists of A1 and A2.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Generation of Standard Curve in 500μl

Final concentration of λ standard μg/ml	Volume of λat 5 μg/ml (μl)	Volume of 1× TE (μl)
2	200	300
1.5	150	350
1	100	400
0.5	50	450
0.2	20	480
0.15	15	485
0.05	5	495
0.02	2	498

Samples need to be diluted to fall into the range of the standards (0.02–2μg/ml).