Journal of Visualized Experiments

Efficient nucleic acid extraction and 16S rRNA gene sequencing for bacterial community characterization

--Manuscript Draft--

Manuscript Number:	JoVE53939R2
Full Title:	Efficient nucleic acid extraction and 16S rRNA gene sequencing for bacterial community characterization
Article Type:	Invited Methods Article - JoVE Produced Video
Keywords:	Microbiome, metagenomics, 16S, DNA extraction, RNA extraction, bacteria, human sequencing, next-generation sequencing, high-throughput sequencing, vaginal, stool, swab
Manuscript Classifications:	3.13: Female Urogenital Diseases and Pregnancy Complications; 4.13.444.308: DNA; 4.13.444.735: RNA; 5.1.370.225.875.150.125: Bacterial Typing Techniques; 5.5.393.760.319: High-Throughput Nucleotide Sequencing
Corresponding Author:	Douglas Kwon, M.D. Ph.D. Ragon Institute of MGH, MIT and Harvard Cambridge, MA UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author E-Mail:	dkwon@mgh.harvard.edu
Corresponding Author's Institution:	Ragon Institute of MGH, MIT and Harvard
Corresponding Author's Secondary Institution:	
First Author:	Melis Nuray Anahtar, Ph.D
First Author Secondary Information:	
Other Authors:	Melis Nuray Anahtar, Ph.D
	Brittany A Bowman
Order of Authors Secondary Information:	
Abstract:	There is a growing appreciation of the role of microbial communities as critical modulators of human health and disease. High throughput sequencing technologies have allowed for the rapid and efficient characterization of bacterial communities using 16S rRNA gene sequencing from a variety of sources. Although readily available tools for 16S rRNA sequence analysis have standardized computational workflows, sample processing for DNA extraction remains a continued source of variability across studies. Here we describe an efficient, robust, and cost effective method for extracting nucleic acid from swabs. We also delineate downstream methods for 16S rRNA gene sequencing, including generation of sequencing libraries, data quality control, and sequence analysis. The workflow can accommodate multiple samples types, including stool and swabs collected from a variety of anatomical locations and host species. Additionally, recovered DNA and RNA can be separated and used for other applications, including whole genome sequencing or RNA-seq. The methods described allow for a common processing approach for multiple sample types that accommodate downstream analysis of genomic, metagenomic and transcriptional information.
Author Comments:	
Additional Information:	
Question	**Response**
If this article needs to be "in-press" by a certain date to satisfy grant requirements, please indicate the date below and	

explain in your cover letter.	

TITLE:
Efficient nucleic acid extraction and 16S rRNA gene sequencing for bacterial community characterization

AUTHORS:
Anahtar, Melis N.
Ragon Institute of MGH, MIT, and Harvard
Massachusetts General Hospital
Cambridge, USA
manahtar@mgh.harvard.edu

Bowman, Brittany A.
Ragon Institute of MGH, MIT, and Harvard
Massachusetts General Hospital
Cambridge, USA
bbowman1@mgh.harvard.edu

Kwon, Douglas S.
Ragon Institute of MGH, MIT, and Harvard
Massachusetts General Hospital
Cambridge, USA
dkwon@mgh.harvard.edu

CORRESPONDING AUTHOR:
Kwon, Douglas S.
dkwon@mgh.harvard.edu

KEYWORDS:
Microbiome, metagenomics, 16S, DNA extraction, RNA extraction, bacteria, human sequencing, next-generation sequencing, high-throughput sequencing, vaginal, stool, swab

SHORT ABSTRACT:
We describe an efficient, robust, and cost effective method for extracting nucleic acid from swabs for characterization of bacterial communities using 16S rRNA gene amplicon sequencing. The methods allow for a common processing approach for multiple sample types that accommodate a number of downstream analytic processes.

LONG ABSTRACT:
There is a growing appreciation of the role of microbial communities as critical modulators of human health and disease. High throughput sequencing technologies have allowed for the rapid and efficient characterization of bacterial communities using 16S rRNA gene sequencing from a variety of sources. Although readily available tools for 16S rRNA sequence analysis have standardized computational workflows, sample processing for DNA extraction remains a continued source of variability across studies. Here we describe an efficient, robust, and cost effective method for extracting nucleic

acid from swabs. We also delineate downstream methods for 16S rRNA gene sequencing, including generation of sequencing libraries, data quality control, and sequence analysis. The workflow can accommodate multiple samples types, including stool and swabs collected from a variety of anatomical locations and host species. Additionally, recovered DNA and RNA can be separated and used for other applications, including whole genome sequencing or RNA-seq. The methods described allow for a common processing approach for multiple sample types that accommodate downstream analysis of genomic, metagenomic and transcriptional information.

INTRODUCTION:

The human lower reproductive tract, gastrointestinal system, respiratory tract, and skin are colonized by complex bacterial communities that are critical for maintaining tissue homeostasis and supporting the health of the host[1]. For instance, certain lactobacilli create an inhospitable environment for pathogens by acidifying the vaginal vault, producing antimicrobial effectors and modulating local host immunity[2-4]. The growing appreciation for the bacterial microbiome's importance has also increased interest in characterizing bacterial communities in many clinical contexts. Here we describe a method to determine the composition of the bacterial microbiome from genital swabs. The protocol can be readily modified for stool samples and swabs collected from other anatomical locations and other host species.

Due to the inherent limitations in the number of samples that can be collected and stored from a given study participant, this protocol was designed to extract DNA, RNA, and potentially even protein from a single swab using an adapted phenol-chloroform based bead-beating method[5,6]. The combination of physical disruption of bacterial cell walls with bead-beating and chemical disruption with detergents allows rapid lysis of Gram-positive, Gram-negative, and acid-fast bacteria without additional enzymatic digestion steps. To obtain high quality RNA, it is recommended to use dry swabs that were kept at or below 4 °C immediately after collection and during transport to the laboratory (if applicable), and stored long-term at -80 °C.

To determine the bacterial microbiome within a given sample, this procedure utilizes 16S rRNA gene amplicon sequencing, which is currently the most cost-effective means to comprehensively assign bacterial taxonomy and perform relative quantification. Alternative methods include targeted qPCR[7], custom microarrays[8], and whole-genome sequencing[9]. The 16S rRNA gene contains nine hypervariable regions, and there is no consensus regarding the optimal V region to sequence for vaginal microbiome studies. Here we use the 515F/806R primer set and build on the pipeline designed by Caporaso *et al.*[10-12]. Caporaso *et al.*'s 515F/806R primer set enables multiplexing of hundreds of samples on a single sequencing run due to the availability of thousands of validated barcoded primers and compatibility with Illumina sequencing platforms. Unlike the Human Microbiome Project's 27F/338R primer set[13], 515F/806R also effectively amplifies *Bifidobacteriaceae* and thus accurately captures *Gardnerella vaginalis*, an important member of the vaginal microbial community in some women. Alternatively, a 338F/806R primer pair has been successfully used for pyrosequencing of vaginal

samples[14] and a 515F/926R primer pair has recently become available for next-generation sequencing[12].

Finally, this protocol provides basic instructions to perform 16S amplicon analysis using the Quantitative Insights into Microbial Ecology (QIIME) software package[15]. Successful implementation of the QIIME commands described here yields a table containing bacterial taxonomic abundances for each sample. Many additional quality control steps, taxonomic classification methods, and analysis steps can be incorporated into the analysis, as described in detail on the QIIME website (http://qiime.org/index.html). If the analysis will be performed on an Apple computer, the MacQIIME package[16] provides easy installation of QIIME and its dependencies. Alternative software packages for 16S rRNA gene sequence analysis include Mothur[17] and UPARSE[18].

PROTOCOL:

The study protocol was approved by and followed the guidelines of the Biomedical Research Ethics Committee of the University of KwaZulu-Natal (Durban, South Africa) and the Massachusetts General Hospital Institutional Review Board (2012P001812/MGH; Boston, MA).

1. Extraction of total nucleic acid from cervicovaginal swabs

Note: Perform nucleic acid extractions in sets of 16 samples or fewer. The protocol as written below assumes samples are processed in sets of 12. If performing multiple rounds of extractions, serially number the extraction batches and record each sample's extraction batch number aa well as other sample information (include metadata such as the participant's ID number, age, date/time of swab collection, hormonal contraceptive type, sexually transmitted infection testing results, etc.) in Table 1.

1.1) Preparation of reagents and fume hood

1.1.1) Prepare a buffer comprised of 200 mM sodium chloride (NaCl), 200 mM Tris, and 20 mM edetic acid (EDTA) in 100 mL of nuclease-free water. Filter-sterilize the solution by passing it through a 0.22 µm filter. Chill an aliquot of 10 mL of buffer on wet ice.

1.1.2) Adjust the pH of the phenol:chloroform:isoamyl alcohol (IAA) (25:24:1) to pH 7.9 by adding 65 µL of Tris alkaline buffer per 1 mL phenol, shaking the mixture for 2 minutes, and allowing the two phases to separate either naturally or by centrifugation at 10,000 x g for 5 minutes at room temperature.

Caution: Phenol is toxic if swallowed, if inhaled, or in contact with skin and eyes. Do not breathe fumes. Wear impervious gloves, safety glasses with side-shields, and a lab coat.

1.1.3) Filter-sterilize 25 mL of 20% sodium dodecyl sulfate (SDS) through a 0.22 µm filter. Make 5 mL aliquots of the sterilized SDS.

1.1.4) Chill a 10 mL aliquot of isopropanol at -20 °C.

1.1.5) Prepare a bead beating tube for each swab to be processed by weighing out 0.3 g of glass beads into a sterile 2 mL tube that is suitable for the bead beater.

1.1.6) Obtain swabs by sampling the ectocervix with a sterile absorbent swab. Immediately after collection, place the swab into an empty and sterile cryovial, store at 4 °C for 1 to 4 hours during transport to the lab, and store for several months at -80 °C. Transfer the swabs (contained within individual tubes) to be processed to wet ice.

1.1.7) Prepare the biological safety cabinet (BSC). Use a BSC with a "thimble" connected to the building exhaust to ensure proper removal of volatile chemicals.

1.1.7.1) Remove all materials from the hood.

1.1.7.2) Clean all surfaces of the hood with bleach, followed by a decontaminant that removes RNases, DNases, and DNA from surfaces. Clean all subsequent items brought into the hood using bleach followed by nucleotide decontamination, including gloves. Use fresh RNase/DNase-free reagents, such as pipette tips, whenever possible.

1.1.7.3) Tape a sterilized chemical biohazard bag to the rear of the hood. All dry waste containing phenol or chloroform should be placed into this bag for proper disposal.

1.1.7.4) Place a sterile bottle into the hood to collect liquid waste containing phenol or chloroform.

1.2) Phenol-chloroform extraction.

1.2.1) In the hood, to each bead beating tube, add 500 µL of buffer (from step 1.1.1), 210 µL of 20% sodium dodecyl sulfate, and 500 µL of phenol:chloroform:IAA (25:24:1, pH 7.9).

1.2.2) Transfer the swab from the transport vial into the bead beating tube using a new pair of sterile forceps. Thoroughly rub the swab head against the internal walls of the bead beating tube for at least 30 seconds. Re-cap the sample when done. If performing extractions from multiple swabs, change gloves between each sample.

1.2.3) Chill the sample on ice for at least ten minutes. Remove the swab from the bead beating tube by holding the swab handle with sterile tweezers while pressing the swab head against the internal tube wall using a clean P200 tip. Discard the swabs in the dry chemical waste bag. Note: The "squeegee" action (pressing the swab head) will liberate liquid from the absorbent swab and increase the nucleic acid recovery.

1.2.4) Place the bead beating tube into the bead beater and homogenize for 2 min at 4 °C.

1.2.5) Centrifuge the bead beating tube for 3 min at 6,000 x g and 4 °C to pellet debris and separate the aqueous and phenol phases.

1.2.6) Transfer the aqueous phase (~500-600 µL) to a sterile 1.5 mL tube. Add an equal volume of phenol:chloroform:IAA. Mix by inversion and brief vortexing.

1.2.7) Centrifuge the tube for 5 min at 16,000 x g and 4 °C.

1.2.8) Transfer the aqueous phase to a new sterile 1.5 mL tube. Be conservative and do not transfer material from the interphase layer or the underlying phenol phase. Note the volume of the transferred aqueous phase. Save the phenol phase for future protein isolation.

1.2.9) Add 0.8 volume of isopropanol and 0.1 volume of 3M sodium acetate (pH 5.5). Mix thoroughly by inversion and briefly vortexing.

1.2.10) Precipitate the nucleic acid by chilling the tube at -20 °C for at least 2 hours (up to overnight).

1.3) Isopropanol precipitation and ethanol wash

1.3.1) Centrifuge the tube for 30 minutes at approximately 16,000 x g and 4 °C. Carefully use a pipette to remove the supernatant, leaving the pellet intact.

1.3.2) Add 500 µL of 100% ethanol. Dislodge the pellet with gentle vortexing or pipetting without touching the pellet. Centrifuge for 5 minutes at 16,000 x g and 4 °C.

1.3.3) Carefully discard the ethanol supernatant. Use a P10 pipet to remove as much ethanol as possible without disturbing the pellet.

1.3.4) Air dry the pellet at room temperature for 15 min.

1.3.5) Resuspend the pellet in 20 µL of ultra-pure 0.1x Tris-EDTA buffer. Allow the sample to chill on ice for 10 minutes and pipette repeatedly to ensure full resuspension. If the pellet does not dissolve, transfer the tube to a 40 °C heat block for up to 10 minutes to aid dissolution.

1.3.6) Measure the nucleic acid concentration using a spectrophotometer[19].

1.3.7) If desired, separate DNA from RNA using a column clean-up kit, following the manufacturer's protocol[20].

1.3.8) Store the nucleic acid at -80 °C or continue.

2. PCR amplification of the 16S *rRNA* gene V4 hypervariable region

Note: Perform the PCR amplification in sets of 12 samples or fewer to minimize the risk of contamination and human error. If performing multiple rounds of amplification, serially number the amplification batches and recording each sample's amplification batch number in Table 1.

2.1) Preparation of the reagents and PCR hood

2.1.1) Add the PCR amplification set information to Table 1, which will serve as the basis of the mapping file at the sequence analysis stage.

2.1.2) Remove all materials from a PCR hood and clean the internal surfaces thoroughly with bleach followed by a decontaminant that removes RNases, DNases, and DNA. Be sure to decontaminate every reagent and piece of equipment (*e.g.* pipettes) before placing them in the hood. Wear fresh gloves cleaned with a nucleic acid decontaminant prior to working in the hood.

2.1.3) If necessary, dilute the nucleic acid template to 50-100 ng/µL using DNA-free and nuclease-free water.

2.1.4) Thaw aliquots of the 5X high-fidelity (HF) buffer, dNTPs, and primers in the clean PCR hood. Gently vortex and centrifuge all solutions after thawing. To minimize freeze-thaw cycles and the risk of stock contamination, prepare aliquots of the 5X HF buffer, dNTPs, and primers.

2.1.5) Place a clean benchtop cooler rack for microcentrifuge tubes and a PCR plate cooler into the hood.

2.2) For PCR reaction, prepare the master mix by combining 15.5 µL of ultra-pure water, 5 µL of 5x HF buffer, 0.5 µL of dNTPs, 0.5 µL of 515F forward primer, 0.75 µL of 3% DMSO, and 0.25 µL of Polymerase for each reaction. Assemble all reaction components in the cooler and add the polymerase last. Mix thoroughly by pipetting. Add two extra samples to the reaction count when preparing the master mix, to account for pipetting error.

2.3) PCR reaction setup:

Note: Perform amplifications in triplicate, meaning each sample is amplified in three separate 25 µL reactions. Run a no-template water control with each primer pair. Work quickly but carefully, avoiding introduction of any contamination.

2.3.1) Label an 8-well strip with individual caps and place into a PCR cooler.

2.3.2) Pipette 90 µL of master mix into the first well.

2.3.3) Add 2 µL of the reverse primer (Supplemental File 1). Be sure to carefully note the reverse primer barcode used with each sample in Table 1.

2.3.4) Mix well and transfer 23 µL of master mix to the fourth well (the no-template control).

2.3.5) Add 2 µL of water to the fourth well.

2.3.6) Add 6 µL of the appropriate sample to the first well. Mix well and transfer 25 µL to the second well. Change tips and transfer another 25 µL from the first well to the third well. Firmly cap every well, making sure not to touch the inside of the wells or cap in the process.

2.3.7) Repeat for each sample.

2.4) Perform PCR amplification

2.4.1) Transfer the strip tubes to a thermocycler and run the following program: 30 sec at 98 °C, followed by 30 cycles of 10 sec at 98 °C, 30 sec at 57 °C, and 12 sec at 72 °C, followed by a 10 min hold at 72 °C and final hold at 4 °C.

2.4.2) Perform the following steps on a clean lab bench. Quickly spin the tubes to collect liquid from the walls. Combine triplicate PCR reactions from each sample, with a total volume of 75 µL, into a sterile labeled tube. Also transfer 25 µL of each no-template control into a separate sterile tube. Do not combine amplicons from different samples yet.

2.5) Validation of successful PCR amplification of samples by gel electrophoresis.

2.5.1) Prepare a 1.5% agarose gel (1.5 g agarose powder in 100 mL of 1x TAE buffer) with enough wells to hold each amplicon, water control, and ladder[21].

2.5.2) While the gel hardens (about 30 minutes), prepare the sample for electrophoresis: Add 1 µL of 6x loading dye to a new, labeled tube. To that tube, add 5 µL of the amplicon and mix by pipetting.

2.5.3) When the gel has set, remove the combs, place the gel in the electrophoresis tank, and fill the tank with 1x TAE buffer.

2.5.4) To the first well, add 5 µL of DNA ladder.

2.5.5) Load 5 µL of the sample amplicon to another well. Load 5 µL of the no-template amplicon to a separate well. Continue as needed for each sample.

2.5.6) When all samples have been loaded, slide the tank lid in place and turn on the power source to 120 V. Allow the gel to run for 30 – 60 minutes.

2.5.7) View the gel under UV light.

2.5.7.1) Verify successful amplification of each sample by noting a single strong band around 380 bp. If there is a double band, re-amplify the sample with a different reverse barcode (Step 2.3). If there is no band at all, re-amplify the sample using either the same reverse barcode or a new reverse barcode (Step 2.3). If re-amplification is unsuccessful, PCR inhibitors may be present in the sample, in which case, perform a column-based DNA cleanup to remove PCR inhibitors.

Note: Successful amplification may not be possible if the bacterial DNA concentration in the original sample is insufficient (<5 ng/µL).

2.5.7.2) Verify the lack of reagent contamination by noting the absence of a band in the no-template control.

2.5.8) Store the remaining 70 µL of amplicon at -20 °C. Discard the remaining 20 µL of the no-template control, assuming it did not yield a band.

3. **Library pooling and high-throughput sequencing**

3.1) Create the amplicon pool by combining an equal volume (2-5 µL) of each amplicon into a single sterile tube. If the band from a sample looked particularly weak, add twice the volume relative to the rest of the samples.

3.2) Remove the PCR primers from the amplicon pool using a PCR Clean-up kit, following the manufacturer's instructions[22]. Perform the clean-up with multiple columns if the amplicon pool volume is over 100 µL. Note: Each column has a 100 µL capacity.

3.2.1) Store the library at -20 °C or proceed to the next step.

3.3) If applicable, combine the primer-free amplicon pools to create the final library. Determine the DNA concentration of the library using a spectrophotometer or a fluorometric system[23]. A 260/280 ratio between 1.8-2.0 is indicative of pure DNA.

3.4) Dilute the library to 20 nM. Confirm the quality of the library by visualizing a single band around 400 bp using an electrophoresis instrument. Confirm the concentration of the library using a fluorometric system[23].

3.5) Perform a final 1:10 dilution in water to dilute the library to 2 nM. Then, store the library at -20 °C indefinitely.

3.6) Send an aliquot of the final library with the three required sequencing primers (Read 1, Read 2, and Index; see Tables of Materials/Equipment) to be sequenced on an Illumina sequencer. If fewer than 300 samples have been multiplexed for sequencing, use a single-end 300 bp run and with a 12 bp index read on a MiSeq, with a final library concentration of 5 pM and a 10% denatured PhiX spike-in. See the supplemental materials of Caporaso *et al.* ISME J, 2012[10] for detailed sequencing instructions.

4. **Sequence analysis**

Note: Outlined here is a basic pipeline for sequence analysis using the QIIME 1.8.0 software package. For simplicity, the provided commands assume that the mapping file is called mapping.txt, the 12 bp index read file is called index.fastq, and the 300 bp sequencing read file is called sequences.fastq. Install QIIME or MacQIIME[16] and familiarize yourself with the basics of UNIX to execute these commands. Read the complete guide to QIIME at: http://qiime.org/index.html.

4.1) Complete the mapping file for the experiment (Table 1). Include as much metadata as possible. Note which samples have been extracted or amplified in the same batch, to determine whether there are batch effects.

4.2) Save the mapping file as a text file, *e.g.* mapping.txt. Validate the formatting of the mapping file by executing the following command:validate_mapping_file.py -m mapping.txt -o mapping_output

Note: This command uses the built-in "validate_mapping_file.py" QIIME script that makes a new folder, called "mapping_output", containing an .html file indicating the mapping file errors, if any.

4.3) Check the quality of the sequencing reads using a high-throughput sequence data quality checking program, such as FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Figure 5 demonstrates the per base sequence quality that can be expected from a successful run.

Note: The sequencer assigns each nucleotide base a Phred quality score, which corresponds to the probability that the base has been erroneously called. A Phred quality score of 10 indicates that there is a 10% chance that the nucleotide has been incorrectly assigned, 20 indicates a 1% chance, 30 indicates a 0.1% chance, and 40 (the highest possible score) indicates a 0.01% chance[24].

4.4) Using the mapping file as a key, demultiplex, quality filter the sequencing data, and save the results to a folder (in this case, called "sl_out") by executing this command[25]: split_libraries_fastq.py --rev_comp_mapping_barcodes -i sequences.fastq -o sl_out/ -b index.fastq -m mapping.txt -q 29

Note: The q flag denotes the maximum unacceptable Phred quality score, *e.g.* "-q 29" filters out any sequences with Phred scores below 30, ensuring 99.9% accuracy of the base calls.

4.5) Using the Greengenes 16S operational taxonomic unit (OTU) reference database[26] (http://qiime.org/home_static/dataFiles.html), perform open-reference OTU picking by executing this command[27]: pick_open_reference_otus.py -i sl_out/seqs.fna -r 97_otus.fasta -o ucrss/ -s 0.1

Note: The -s flag indicates the fraction of sequences that failed to align to the reference database that will be included in the *de novo* clustering. "-s 0.1" includes 10% of the

failed sequences in the *de novo* clustering. Use the -a flag to parallelize the OTU picking process and reduce the processing time from days to hours if multiple cores are available.

4.6) Create a user-friendly taxonomic abundance table by merging OTUs at the species level by executing this command[28]: summarize_taxa.py -i ucress/otu_table_mc2.biom -o summarized_otuSpecies/ -L 7

Note: The resulting table can be easily viewed in any spreadsheet software. Note that 16S rRNA sequencing does not reliably provide species level resolution.

4.7) Determine the ecological diversity within each sample by computing several alpha diversity metrics with the QIIME script alpha_diversity.py. Then, determine the diversity between pairs of samples using the QIIME script beta_diversity.py.

4.8) Visualize the data, *e.g.* by using an EMPeror[29] principal coordinates plot or heatmap.

4.9) Perform formal statistical comparisons of mapping file categories, *e.g.* with QIIME's compare_catagories.py script[30].

REPRESENTATIVE RESULTS:

The general overview of the protocol, which enables the determination of relative bacterial abundances from a swab using 16S rRNA gene sequencing, is shown in Figure 1. The protocol has been optimized for human vaginal swabs, but can be easily adapted for most mucosal sampling sites and other hosts. Figure 2 demonstrates the high-quality DNA and RNA that can be isolated using the bead-beating protocol. Figure 3 illustrates a successful PCR amplification of 12 samples, where each amplification with a sample yielded a single strong band of the correct size and each water control did not yield a band. Figure 4 illustrates the quantification of the final library pool prior to sequencing. Figure 5 shows a typical sequence quality profile after a single-end 300 bp MiSeq run.

Figure 1: Schematic overview of the protocol. First, nucleic acid is extracted from a swab by bead-beating in a buffered solution containing phenol, chloroform, and isoamyl alcohol. Variable region 4 of the 16S rRNA gene is then amplified from the resulting nucleic acid using PCR. PCR amplicons from up to hundreds of samples are then combined and sequenced on a single run. The resulting sequences are matched to a reference database to determine relative bacterial abundances. The entire protocol can be performed in approximately three days.

Figure 2: High-quality nucleic acid extracted using the phenol:chloroform bead beating method. A) DNA quality, as assessed using a spectrophotometer. An A260/A280 ratio between 1.8 and 2.0 indicates pure nucleic acid that is not contaminated with phenol or protein. **B)** After a column clean-up, this protocol can yield high-quality RNA, indicated by strong 16S and 23S rRNA peaks. **C)** RNA degradation can occur if the sample is not kept cold after collection (during transport and storage) or if RNases are present during processing.

Figure 3: Confirmation of successful 16S rRNA gene amplification using the 515F and barcoded 806R primer set. Top) Gel electrophoresis is used to confirm the presence of a single band around 380 base pairs in every sample that was amplified with template. The absence of a band indicates unsuccessful amplification; this is usually due to human error and the PCR reaction from that sample should be repeated. **Bottom**) No template (water) controls run in parallel with the same primer pair should *not* have a band present. The presence of a band in the water control indicates contaminated reagents; discard the reagents that may be contaminated and re-do the PCR amplifications of both the template and water control for that primer pair.

Figure 4: Quantification of the final library pool concentration and validation of the library size. After pooling the individual sample amplicons, the concentration of the final library pool must be determined. The library pool must then be further diluted to achieve a 2 nM concentration.

Figure 5: Representative bar plot of the sequence quality scores at each position of the read. It is normal for the sequence quality to drop after 200 base pairs, but the average quality score should remain above 30.

Table 1: Mapping file template. Creating an accurate and thorough mapping file is critical for successfully executing the protocol. The mapping file is not only required for executing QIIME, but it also enables the researcher to maintain the link between the sample barcode and metadata, to analyze the data for any systematic biases (*e.g.* batch-to-batch variation), and to determine interesting correlations between the metadata and bacterial populations. A bare-bones mapping file is provided, but users are encouraged to add as many columns containing metadata as possible. Examples of additional metadata for a vaginal swab includes the participant's age, date/time of swab collection, hormonal contraceptive type (if applicable), sexually transmitted infection testing results, etc.

Supplemental File 1: List of barcoded reverse primer sequences[10]. The first three columns can be used to complete the mapping file, and the last column provides the entire primer sequence for ordering purposes.

DISCUSSION:

Here we describe a protocol for the identification and characterization of relative bacterial abundances within a human vaginal swab. This protocol can easily be adapted for other sample types, such as stool and swabs of other body sites, and for samples collected from a wide variety of sources. The extraction of nucleic acid by bead-beating in a buffered solution of phenol and chloroform allows for isolation of both DNA and RNA, which is particularly important when working with precious samples collected through clinical studies. The isolated bacterial DNA is excellent for bacterial taxonomic identification and genomic assembly, while the simultaneous collection of RNA provides the opportunity to determine functional bacterial, host, and viral contributions through RNA-seq. The described protocol uses a validated one-step primer set that has been

successfully deployed on a wide range of sample types, including human, canine, and environmental samples[10]. The availability of thousands of barcoded primers enables multiplexing of samples and tremendous savings on sequencing costs. The complete cost (including all reagents, a single sequencing run, and primers but not equipment) is about $20 per sample when 200 samples are multiplexed. Additionally, there is very high reproducibility when multiple swabs from the same sample site are processed independently through the entire pipeline. Overall, the protocol is cost efficient, flexible, reliable, and repeatable.

The nucleic acid extraction portion of this protocol is limited by the safety precautions required when working with phenol and chloroform, and the challenges of automating the pipeline to a high-throughput, 96-well plate format. Additionally, the vigorous bead beating used for mechanical lysis shears the bacterial DNA to approximately 6 kilobase fragments; if longer DNA fragments are required for downstream applications, the duration of bead beating should be shortened. The limitations of the bacterial identification portion of this protocol are inherent to any method that relies on 16S rRNA gene sequencing. 16S rRNA sequencing is ideal for bacterial identification to the genus and even species level, but rarely provides strain level identification. While the V4 variable region of the 16S rRNA gene provides robust discrimination amongst most bacterial species[11], additional computational methods such as Oligotyping[31] may need to be used to precisely identify certain species, such as *Lactobacillus crispatus*. Finally, information about the precise bacterial functional capabilities within a particular sample cannot be determined by 16S rRNA gene sequencing alone, though this protocol enables extraction of whole genome DNA and RNA that can be used towards this purpose.

The most critical step to ensuring success with this protocol is taking great care to prevent contamination during sample collection, nucleic acid extraction, and PCR amplification. Ensure sterility at the time of sample collection by wearing clean gloves and using sterile swabs, tubes, and scissors. To assess for contamination of the collection materials, collect negative control swabs by placing additional unused swabs directly into transport tubes at the time of sampling. In the lab, perform all pre-amplification steps in a sterilized hood containing only decontaminated supplies and using only molecular grade, DNA-free reagents. During nucleic acid extraction, prevent cross-contamination by using new sterile forceps and fresh gloves with each sample, and keeping all tubes closed unless in use. Processing unused swabs in parallel ensures sterility of both the sample collection and nucleic acid extraction; the unused swabs should not yield a pellet after isopropanol precipitation and ethanol washing. If a pellet does appear, perform 16S rRNA gene amplification to determine a possible source of the contamination (*e.g.* the presence of *Streptococcus* or *Staphylococcus* would indicate skin contamination). Additionally, perform PCR amplifications with no template control reactions in parallel to ensure that the PCR reagents and reactions have not been contaminated. If a band appears in a no template control, discard the reagents and repeat the amplification with fresh reagents. Taking these precautions will ensure successful sequencing of the bacteria of interest.

The PCR amplification step tends to require the most troubleshooting. Amplifying in sets of twelve samples provides a balance between efficiency and consistency. The complete absence of bands across all samples in a given amplification set indicates a systematic failure, *e.g.* forgetting to add a reagent or incorrectly programming the thermocycler. The absence of a band from a few samples is usually due to human error, and the amplifications should be re-run with the same pairing of sample and reverse primer. In the case of continued absence of a band, the sample can be re-amplified using a reverse primer with a different barcode. Repeated amplification failures with multiple reverse primers may indicate an inhibitor present in the sample. In that case, cleaning the DNA with a column will often remove inhibitors without significantly altering relative bacterial abundances. If multiple bands result after amplification, re-amplify the sample with a different reverse primer barcode.

In addition to preventing environmental contamination and ensuring amplification of a single specific product, successful sequencing relies on care when preparing the library pool. The goal is to combine equimolar amounts of each sample's amplicons to ensure approximately the same number of sequencing reads per sample. If the nucleic acid concentrations prior to amplification are comparable, simply adding equal volumes of each sample's amplicons is sufficient when creating the library pool. However, if the nucleic acid concentrations are vastly different and added in equal volume, the sample with the low nucleic acid concentration will be poorly represented with a low number of reads. In this case, it is possible to add a higher volume of the amplicons from the low concentration sample based upon the relative intensity of the gel band. Alternatively, it is possible to more rigorously remove primers from the individual amplicons, quantify individual sample's amplicon concentration using a fluorometric dsDNA quantification kit, and precisely combine equimolar amounts of each sample.

Once a well-balanced amplicon pool is generated, it becomes critical to carefully measure the pool's concentration. Subsequent careful dilution and spike-in with PhiX to increase the read complexity is critical for achieving optimal sequencing results. High-throughput sequencers that use sequencing by synthesis are very sensitive to the cluster density on the flow cell. Loading a library pool that is too concentrated will result in overclustering, with lower quality scores, lower data output, and inaccurate demultiplexing[32]. Loading a library pool that is too dilute will also result in low data output. Carefully quantifying the library pool prior to sequencing will ensure optimal results.

16S rRNA gene sequencing provides a comprehensive assessment of the bacteria present within a given sample and is an absolutely critical first step in hypothesis generation. The presence of a rich set of metadata further enables the researcher to test associations between particular bacterial species and important biological factors. Furthermore, the same 16S information can be used to infer the bacterial functions using with tools such as PICRUSt[33]. The ultimate goal is to use 16S characterization to identify novel associations that can be further tested and validated in model systems, adding to our growing understanding of the impact of the bacterial microbiome on human health and disease.

ACKNOWLEDGMENTS:
We would like to thank Elizabeth Byrne, David Gootenberg, and Christina Gosmann for critical feedback on the protocol; Megan Baldridge, Scott Handley, Cindy Monaco, and Jason Norman for sample preparation guidance and demonstrations; Wendy Garrett, Curtis Huttenhower, Skip Virgin, and Bruce Walker for protocol advice and fruitful discussions; and Jessica Hoisington-Lopez for sequencing support. This work was supported by the Bill and Melinda Gates Foundation and the NIAID (1R01AI111918). D.S.K. received additional support from the Burroughs Wellcome Fund. M.N.A. was supported by award number T32GM007753 from the NIGMS, and the Paul and Daisy Soros Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIGMS or the NIH.

DISCLOSURES:
The authors have nothing to disclose.

REFERENCES:
1 Huttenhower, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214, doi:10.1038/nature11234 (2012).

2 O'Hanlon, D. E., Moench, T. R. & Cone, R. A. Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PloS one* **8**, e80074, doi:10.1371/journal.pone.0080074 (2013).

3 Aldunate, M. *et al.* Vaginal concentrations of lactic acid potently inactivate HIV. *The Journal of antimicrobial chemotherapy* **68**, 2015-2025, doi:10.1093/jac/dkt156 (2013).

4 Anahtar, M. N. *et al.* Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. *Immunity* **42**, 965-976, doi:10.1016/j.immuni.2015.04.019 (2015).

5 Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. & Gordon, J. I. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 20236-20241, doi:10.1073/pnas.1319470110 (2013).

6 Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical biochemistry* **162**, 156-159, doi:10.1006/abio.1987.9999 (1987).

7 Srinivasan, S. *et al.* Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PloS one* **5**, e10197, doi:10.1371/journal.pone.0010197 (2010).

8 Dols, J. A. *et al.* Microarray-based identification of clinically relevant vaginal bacteria in relation to bacterial vaginosis. *American journal of obstetrics and gynecology*

204, 305 e301-307, doi:10.1016/j.ajog.2010.11.012 (2011).

9 Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature methods* **9**, 811-814, doi:10.1038/nmeth.2066 (2012).

10 Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal* **6**, 1621-1624, doi:10.1038/ismej.2012.8 (2012).

11 Caporaso, J. G. *et al.* 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* **108 Suppl 1**, 4516-4522, doi:10.1073/pnas.1000080107 (2011).

12 Earthmicrobiome Project. 16S rRNA Amplification Protocol. (2015).

13 Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America* **108 Suppl 1**, 4680-4687, doi:10.1073/pnas.1002611107 (2011).

14 Srinivasan, S. *et al.* Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PloS one* **7**, e37818, doi:10.1371/journal.pone.0037818 (2012).

15 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7**, 335-336, doi:10.1038/nmeth.f.303 (2010).

16 Werner, J. MacQIIME. (2015).

17 Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology* **75**, 7537-7541, doi:10.1128/AEM.01541-09 (2009).

18 Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* **10**, 996-998, doi:10.1038/nmeth.2604 (2013).

19 Thermo Fisher Scientific. *NanoDrop 2000/2000c Spectrophotometer, V1.0 User Manual*, <http://www.thermoscientific.com/content/dam/tfs/ATG/CAD/CAD Documents/Product Manuals & Specifications/Molecular Spectroscopy/UV Visible Spectrophotometers/Spectrophotometer Systems/NanoDrop/NanoDrop-2000-User-Manual-EN.pdf> (2009).

20 Qiagen. *AllPrep DNA/RNA Mini Kit*, <https://http://www.qiagen.com/us/shop/sample-technologies/rna-sample-technologies/dna-rna-protein/allprep-dnarna-mini-kit/ - orderinginformation> (

21 Lee, P. Y., Costumbrado, J., Hsu, C. Y. & Kim, Y. H. Agarose gel electrophoresis

for the separation of DNA fragments. *Journal of visualized experiments : JoVE*, doi:10.3791/3923 (2012).

22 MoBio. *UltraClean PCR Clean-Up Kit Instruction Manual*, <http://www.mobio.com/images/custom/file/12500(1).pdf> (2013).

23 Invitrogen. *Quant-iT PicoGreen dsDNA Reagent*, <https://tools.thermofisher.com/content/sfs/manuals/mp07581.pdf> (

24 Illumina. *Quality Scores for Next-Generation Sequencing*, <http://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf> (2011).

25 QIIME. *split_libraries_fastq.py*, <http://qiime.org/scripts/split_libraries_fastq.html> (

26 DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* **72**, 5069-5072, doi:10.1128/AEM.03006-05 (2006).

27 QIIME. *pick_open_reference_otus.py*, <http://qiime.org/scripts/pick_open_reference_otus.html> (

28 QIIME. *summarize_taxa.py*, <http://qiime.org/scripts/summarize_taxa.html> (

29 Vazquez-Baeza, Y., Pirrung, M., Gonzalez, A. & Knight, R. EMPeror: a tool for visualizing high-throughput microbial community data. *GigaScience* **2**, 16, doi:10.1186/2047-217X-2-16 (2013).

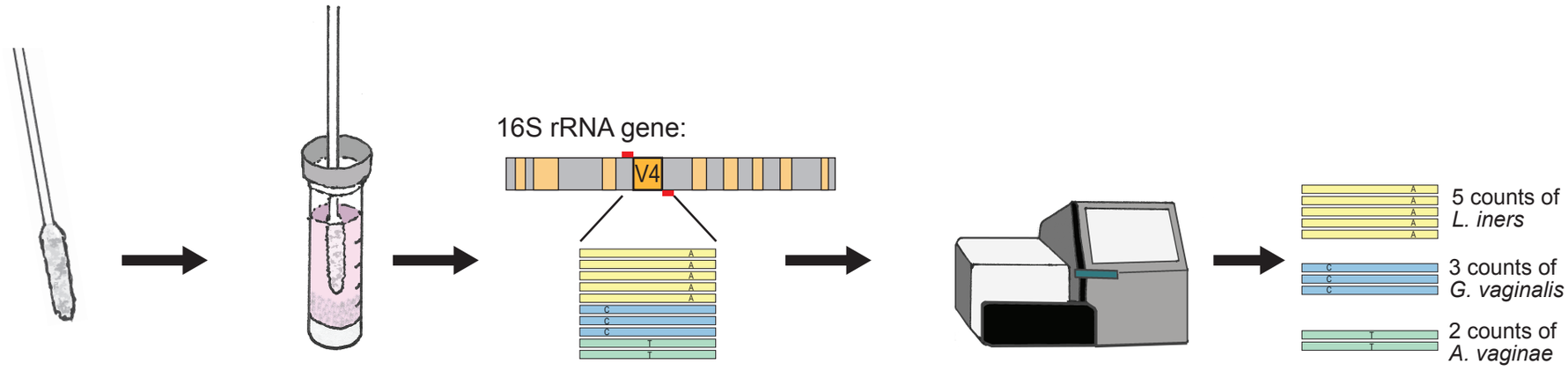
30 QIIME. *Comparing categories*, <http://qiime.org/tutorials/category_comparison.html> (

31 Eren, A. M. *et al.* Exploring the diversity of Gardnerella vaginalis in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. *PloS one* **6**, e26732, doi:10.1371/journal.pone.0026732 (2011).

32 Illumina. *Diagnosing and preventing flow cell overclustering on the MiSeq system*, <http://support.illumina.com/content/dam/illumina-marketing/documents/products/other/miseq-overclustering-primer-770-2014-038.pdf> (2015).

33 Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology* **31**, 814-821, doi:10.1038/nbt.2676 (2013).

Figure 1
Click here to download Figure: JoVE_Figures_2015July11_Part1.pdf



Sample collection

DNA (and RNA) isolation

16S rRNA gene PCR amplification of variable 4 region

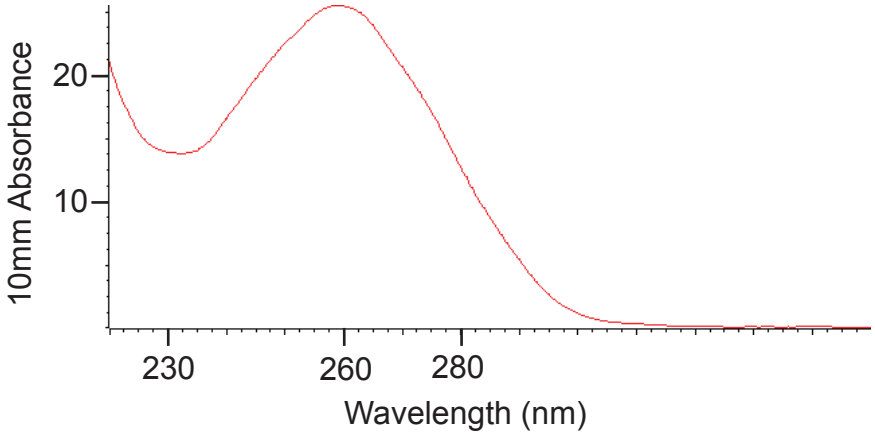
Amplicon pooling and Illumina sequencing

Data analysis

Figure 2
Click here to download Figure: JoVE_Figures_2015July11_Part2.pdf

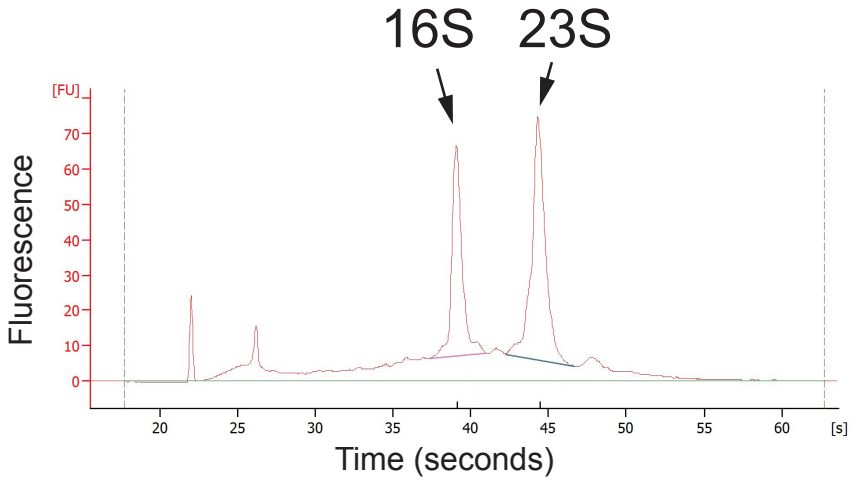
A

A260/A280:
1.8-2.0



B

High-quality RNA
RIN: 8.6



C

Degraded RNA

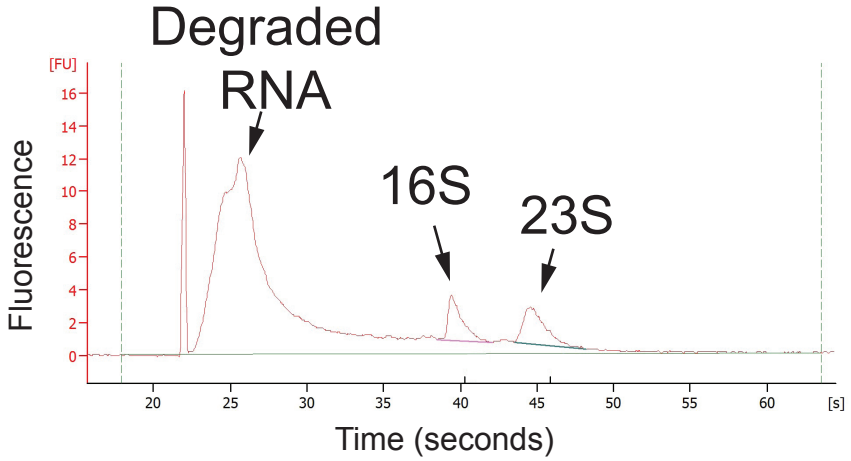


Figure 3
Click here to download Figure: JoVE_Figures_2015July11_Part3.pdf

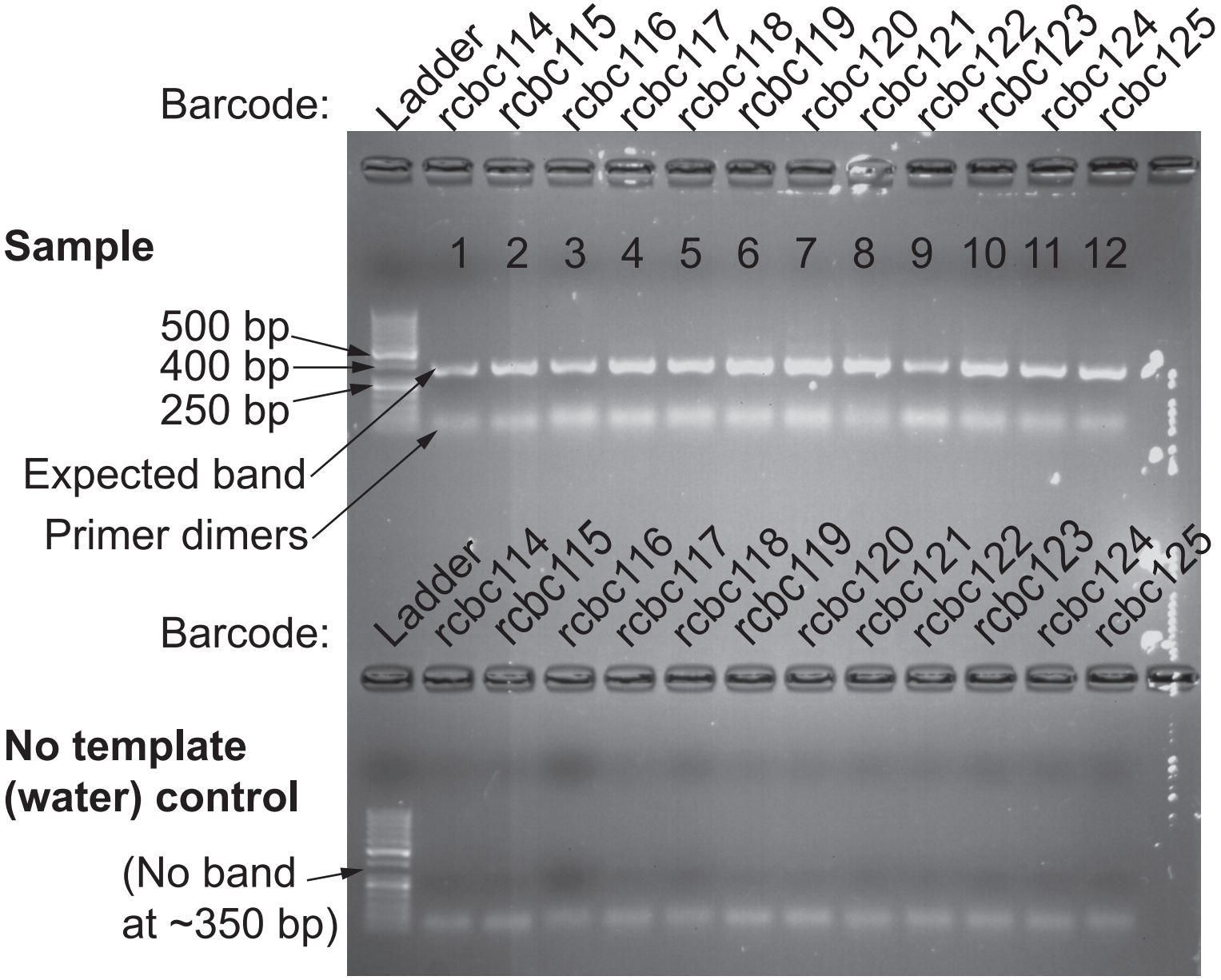


Figure 4
Click here to download Figure: JoVE_Figures_2015July11_Part4.pdf

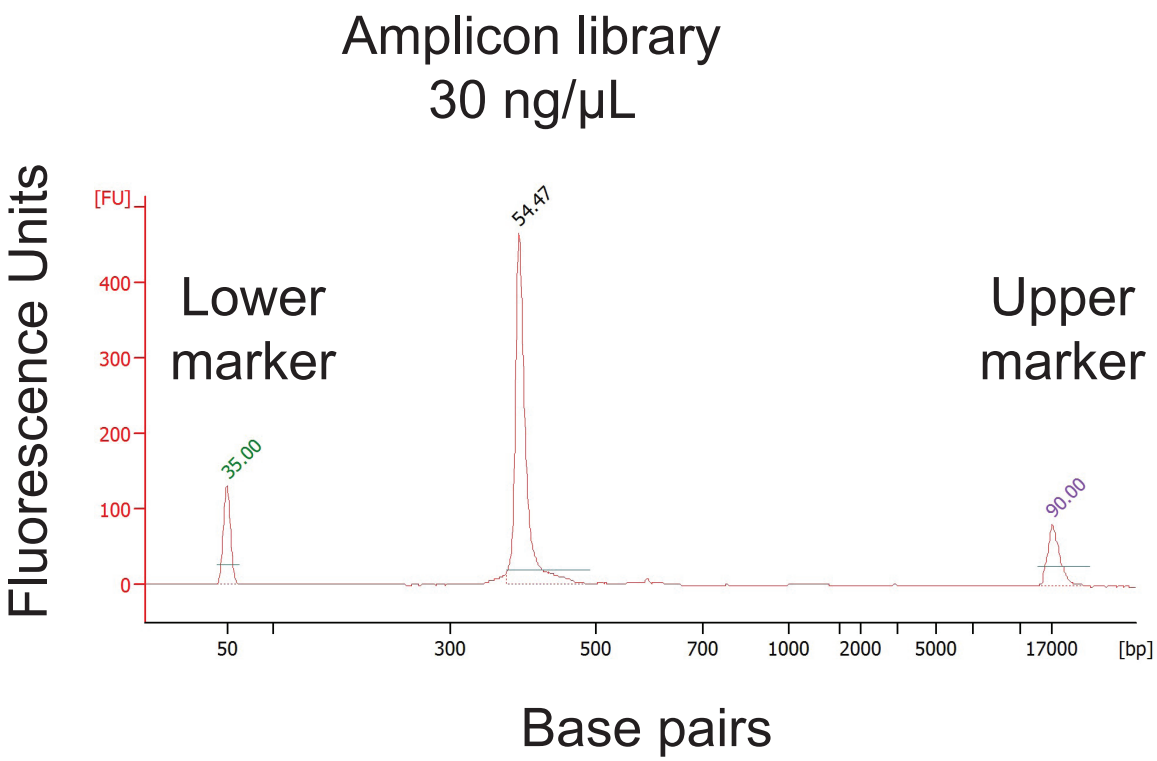


Figure 5
Click here to download Figure: JoVE_Figures_2015July11_Part5.pdf

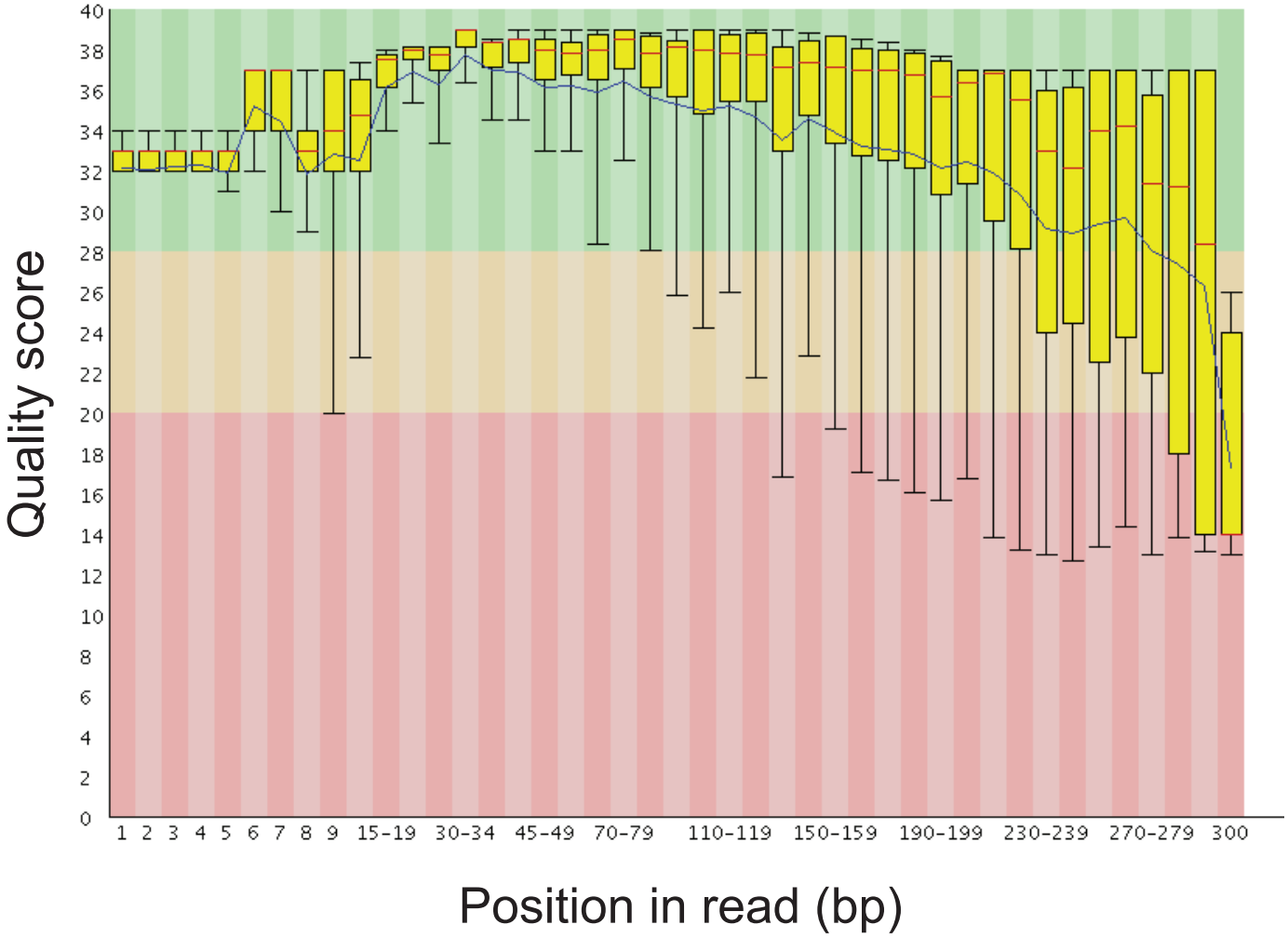


Table 1
Click here to download Table: _Table1_Mappingfile.xlsx

#SampleID	BarcodeSequence	LinkerPrimerSequence	rcbcPrimer
#An example mapping file can be found at: http://qiime.org/_static/Examples/File_Formats/Example			
AG2350	TCCCTTGTCTCC	CCGGACTACHVGGGTWTCTAAT	rcbc000

SampleType	ExtractionBatch	AmplificationPlate	Description
_Mapping_File.txt			
CervicalSwab	1	A	

Excel Spreadsheet- Table of Materials/Equipment

Click here to download Excel Spreadsheet- Table of Materials/Equipment: _Materials.xlsx

Equipment:	**Company**	**Catalogue Number**	**Comments/Description**
Mini-Beadbeater-16	BioSpec	607	
PCR workstation			Any PCR hood can be used, *e.g.* , the AirClean 600.
Thermocycler			Any thermal cycler with a heated lid can be used, *e.g.* , MJ Research PTC-200.
Electrophoresis system			Any electrophoresis system can be used, *e.g.* the Thermo Scientific Owl EasyCast B1 Mini Gel Electrophoresis system.
Nanodrop	Thermo Scientific	2000C	Any other DNA quantification method will be sufficient
Bioanalyzer	Agilent	2100	An alternative is the Agilent 2200 TapeStation Instrument. Not absolutely necessary but very helpful.
MiSeq or HiSeq	Illumina		
Materials:	**Company**	**Catalogue Number**	**Comments/Description**
Catch-All Sample Collection swab	Epibio	QEC89100	Other swabs can be used but the Catch-All swab is recommended by the Human Microbiome Project.
ELIMINase	Fisher	04-355-31	
SteriFlip 50 mL filtration device (0.22 µm)	EMD Millipore	SCGP00525	
0.1 mm glass beads	BioSpec	11079101	
2 mL screw-cap tubes	Sarstedt	72.694.006	For bead beating
UltraPure 5M NaCl	Life Technologies	24740-011	Molecular Biology Grade
1 M Tris-HCl	Ambion (Invitrogen)	AM9856	Molecular Biology Grade
0.5 M EDTA	Ambion (Invitrogen)	AM9260G	Molecular Biology Grade
Sodium Dodecyl Sulfate, 20% Solution	Fisher	BP1311-200	Molecular Biology Grade
UltraPure DNase/RNase-free distilled water	Ambion	10977-015	Molecular Biology Grade, for buffer preparation
2-Propanol BioReagent, for molecular biology, ≥99.5%	Sigma	I9516-500ML	Molecular Biology Grade
Phenol:Chloroform:IAA, 25:24:1	Invitrogen	AM9730	Warning: Toxic
3 M Sodium Acetate, pH 5.5	Life Technologies	AM9740	Molecular Biology Grade
Disposable sterile polystyrene forceps, PS	Cole Parmer	EW-06443-20	
1.5 mL, clear, PCR clean tubes	Eppendorf	22364120	
PCR grade water	MoBio	17000-11	For PCR
Phusion High-Fidelity DNA Polymerase	New England Biolabs	M0530S	
dNTP mix	Sigma	D7295-0.5mL	
0.2 ml PCR 8-tube with attached clear flat caps, natural	USA Scientific	1492-3900	Any 8-tube strips that are DNase, RNase, DNA, and PCR inhibitor free will work
Agarose	BioExpress	E-3121-25	
50X TAE buffer	Lonza	51216	
DNA gel stain	Invitrogen	S33102	
6X DNA Loading Dye	Thermo (Fisher)	R0611	
50bp GeneRuler Ladder	Thermo (Fisher)	SM0373	
AllPrep DNA/RNA kit	Qiagen	80284	
UltraClean PCR Clean-up Kit	MoBio	12500-100	
Quant-iT PicoGreen dsDNA Assay Kit	Thermo Fisher Scientific	P11496	An alternative is Qubit Fluorometric Quantification (Life Technologies)
Primers:	**Company**	**Catalogue Number**	**Comments/Description**
515F (forward primer) 5'-AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTGCCAGCMGCCGCGGTAA-3'			Order at 100 nmole; Purification: Standard Desalting. Resuspend at 100 µM. ****Critical: primers must be resuspended with MoBio PCR Grade Water (see above) in a hood to avoid contamination.****
Reverse primers, see the Supplemental Code File and: ftp://ftp.metagenomics.anl.gov/data/misc/EMP/SupplementaryFile1_barcoded_primers_515F_806R.txt	IDT is recommended		If ordering large sets of primers, order as a 96-well plate at the 100 nmole scale. Resuspend at 100 µM. Full directions for primer ordering and resuspension at http://www.earthmicrobiome.org/files/2013/04/EMP_primer_ordering_and_resuspension.doc. ****Critical: primers must be resuspended with MoBio PCR Grade Water (see above) in a hood to avoid contamination.****
Read 1 Sequencing Primer 5'-TAT GGT AAT TGT GTG CCA GCM GCC GCG GTA A-3'			25 nmole; Purification: Standard Desalting. Resuspend at 100 µM.
Read 2 Sequencing Primer 5'-AGT CAG TCA GCC GGA CTA CHV GGG TWT CTA AT-3'			26 nmole; Purification: Standard Desalting. Resuspend at 100 µM.
Index Sequencing Primer 5'-ATT AGA WAC CCB DGT AGT CCG GCT GAC TGA CT-3'			27 nmole; Purification: Standard Desalting. Resuspend at 100 µM.
PhiX Control v3	Illumina	FC-110-3001	Required if performing the sequencing in-house. If the sequencing will be performed by a third-party sequencing center, they will already have PhiX.

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: _____

Author(s): _____

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via: ☐ Standard Access ☐ Open Access

Item 2 (check one box):

☐ The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name:

Department:

Institution:

Article Title:

Signature:  Date:

Please submit a signed and dated copy of this license by one of the following three methods:

1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

Rebuttal Comments
Click here to download Rebuttal Comments: _Responses to editor and reviewers_DSK.docx

JoVE Submission 53939

Title: Efficient nucleic acid extraction and 16S rRNA gene sequencing for bacterial community characterization

Response to editor's comments:

Editorial comment 1: 3.4: Please provide a citation for measuring DNA concentration using a fluorometric system.

Author response: We have added a citation.

Editorial comment 2: Please provide a short title, followed by a short description for Supplemental File 1.

Author response: We have added a short title and description.

Editorial comment 3: Please remove trademark and registered trademark symbols (TM/R) from the Table of Materials/Equipment.

Author response: We have removed TM/R from the table.

Editorial comment 4: There is unnecessary branding (Illumina) throughout, which should be removed:
-Introduction 1x -3.6 - "Illumina sequencer" - Use generic term in place.
-Figure 5 "Illumina" mentioned.

Author response: We removed the mention of Illumina from Figure 5 and one of the two references in 3.6. However, we believe it's important to mention it once in 3.6 as the primer set is only compatible with Illumina platforms.

Editorial comment 5: JoVE reference format requires that DOIs are included, when available, for all references listed in the article. This is helpful for readers to locate the included references and obtain more information. Please note that often DOIs are not listed with PubMed abstracts and as such, may not be properly included when citing directly from PubMed. In these cases, please manually include DOIs in reference information.

Author response: DOIs are listed where available.

Editorial comment 6: Prior to peer review, the highlighted portion of your protocol is close to our 2.75 page highlighting limit. If, in response to peer review, additional details are added to the protocol, please adjust the highlighting to identify a total of 2.75 pages of protocol text (which includes sub-headings and spaces) that should be visualized to tell the most cohesive story of your protocol steps. The highlighting should include complete statements and not portions of sentences. See JoVE's instructions for authors for more clarification.

Author response: We thank the editor for this reminder. We have not made any additions to the highlighted portion.

Responses to Reviewer #1's comments:

Reviewer Comment 1: In section 1 of the protocol, why is it only necessary to record sample details if performing multiple rounds of extractions? Isn't this necessary regardless?

> **Author Response:** We apologize for the confusion. We agree that recording sample details is necessary regardless, but serially numbering the extraction batches is only necessary if performing >1 batch. We have modified the language to say:
>
> "The protocol as written below assumes samples are processed in sets of 12. If performing multiple rounds of extractions, we also recommend serially numbering the extraction batches. Record each sample's extraction batch number and other sample information in **Table 1**."

Reviewer Comment 2: In section 4, Sequence Analysis, what version of QIIME is the text described against?

> **Author Response:** We have added the QIIME version to the text:
>
> "Note: Outlined here is a basic pipeline for sequence analysis using the QIIME 1.8.0 software package. For simplicity, the provided commands assume that the mapping file is called mapping.txt, the 12 bp index read file is called index.fastq, and the 300 bp sequencing read file is called sequences.fastq. Install QIIME or MacQIIME and familiarize yourself with the basics of UNIX to execute these commands. Read the complete guide to QIIME at: http://qiime.org/index.html."

Reviewer Comment 3: In section 4, Sequence Analysis, the motivation to track batch details is so that the analyst can identify _if_ there are batch effects, not to prevent them as the analysis is after the fact.

> **Author Response:** We agree with the reviewer and apologize if this was not clear. We hope our revision clarifies this point:
>
> "Note which samples have been extracted or amplified in the same batch, to determine whether there are batch effects."

Reviewer Comment 4: In section 4.6, it should be noted that 16S does not reliably provide species level resolution.

> **Author Response:** We thank the reviewer for their suggestion and have added this note.

Reviewer Comment 5: Was a bit surprised to see there wasn't any mention of statistical tests or visualizations that likely should be performed with QIIME. Would the authors consider adding that, or citing material where readers could find more information?

Author Response: We appreciate the reviewer's suggestion and have added the following steps:

4.7) Visualize the data, *e.g.* by using a principal coordinates plot (e.g. EMPeror) or heatmap.
4.8) Perform formal statistical comparisons of mapping file categories, *e.g.* with QIIME's compare_catagories.py script[29].

Reviewer Comment 6: Would the authors consider adding a mention of the primary differences between the DNA extraction protocol in their manuscript and the EMP DNA Extraction Protocol (found under Download Links here http://www.earthmicrobiome.org/emp-standard-protocols/16s/)? Or, if this is not applicable, a sentence indicating why would be great to include.

Author Response: We appreciate the reviewer's suggestion. Indeed, there are many similarities with the EMP protocol given the use of the same primer sets. The primary differences are:

- The EMP DNA Extraction Protocol is optimized to only recover DNA from samples using a commercially available kit. Our protocol allows for the recovery and purification of both DNA and RNA.
- The EMP Protocol does not detail how to specifically accommodate the recovery of nucleic acid from genital swab samples, which is the specific application that is the focus of our manuscript.
- The EMP Protocol uses a commercially available kit for DNA extraction, which is twice as expensive per sample than the protocol we have described.

Reviewer Comment 7: Please expand the first use of QIIME to Quantitative Insights into Microbial Ecology.

Author comment: We have expanded the first use of QIIME, which is in the last paragraph of the introduction.

Reviewer Comment 8: In section 1 of the protocol, the period between the right parenthesis and "Perform" should be replaced with "to".

Author comment: Could the reviewer clarify the location of this typo? Unfortunately we cannot seem to find it!

Reviewer Comment 9: In section 4, Sequence Analysis, it is a very good idea to compile this information at the time of sample collection.

Author comment: We agree with the reviewer and have moved the information to Section 1, Note #2:

"Note #2: Record each sample's extraction batch number and other sample information in **Table 1**. For example, for vaginal swabs, include metadata such as the participant's ID number, age, date/time of swab collection, hormonal contraceptive type, sexually transmitted infection testing results, etc."

Reviewer Comment 10: In section 4, Sequence Analysis, if the authors are not aware, there is a QIIME script called "core_diversity_analyses.py" which may be of interest to look at.

> **Author comment:** We thank the reviewer for noting the absence of diversity metric calculations. We have added the following step: "Determine the ecological diversity within each sample by computing several alpha diversity metrics with the QIIME script alpha_diversity.py. Then, determine the diversity between pairs of samples using the QIIME script beta_diversity.py."

Reviewer Comment 11: Figure 3 caption indicates "around 380 base pairs" whereas the protocol indicates 350.

> **Author response:** We thank the reviewer for noting this discrepancy and have changed the protocol to say 380 base pairs.

Reviewer Comment 12: In the discussion, there is a mention of cost. Would the authors consider including a ballpark estimate of the cost (both monetary and personal hours) per sample for the protocol?

> **Author response:** We thank the reviewer for their comment and have addressed it by adding the following text to the discussion: "The complete cost (including all reagents, a single sequencing run, and primers but not equipment) is about $20 per sample when 200 samples are multiplexed."

Responses to Reviewer #2's comments:

Reviewer Comment 1: The manuscript is written very well, but I am not sure what it adds to the literature although having the detailed protocol published would be useful.

> **Author response:** We appreciate the reviewer's opinion and agree that there is at least one written 16S protocol available from the Earth Microbiome Project. However, we believe our major contributions with this protocol include:
>
> **1)** Providing a complete and detailed pipeline from sample to data analysis, including representative data, critical steps, and troubleshooting suggestions,
> **2)** Enabling the simultaneous extraction of DNA and RNA,
> **3)** The video portion of this protocol, as proper technique is critical for success of this protocol.

Reviewer Comment 2: The only difference being that they used phenol: chloroform extraction - they say the advantage is to isolate RNA as well as DNA but do not go on to prove that the RNA is of sufficient quality for downstream analysis.

Author response: We believe that Figure 2 clearly demonstrates that the RNA is of sufficient quality for downstream analysis. RNA Integrity Numbers above 8 are considered to be very high quality.

Reviewer Comment 3: They mention isolating both RNA and DNA in the abstract but don't discuss how to separate RNA/DNA in the protocol or how to protect RNA from degradation after sampling.

Author response: We thank the reviewer for their thoughtful comment. We have added the following step: "1.3.7) If desired, separate DNA from RNA using a column clean-up kit, following the manufacturer's protocol[20]", with manufacturer information given in the reference and materials list.

Regarding the protection of RNA from degradation after sampling, we have not needed to add RNase inhibitors to the samples, as the phenol and chloroform effectively inhibits ribonucleases. The main protection steps are inherent to the protocol: decontamination of surfaces, using fresh and clean gloves, working in a sterile hood, using only RNase free reagents, and keeping reagents cold.

Reviewer Comment 4: They don't mention using lysozyme or something to lyse gram+ve cell walls, maybe I missed something?

Author response: We thank the reviewer for noting this important point. We have found that the combination of dry freeze-thaw, lysis buffer, phenol, chloroform, and bead beating is effective for lysing gram positive cell walls. Our yields were not improved by pre-treating with proteinase K for 1-2 hours. This is consistent with the literature, *e.g.* Liu, Dongyou. "Handbook of Nucleic Acid Purification," (2009) CRC Press, p. 103.

Reviewer Comment 5: I know they're focusing on the lab techniques in this paper but it would be helpful for the user to have more detailed info for the sequencing analyses part

Author response: We thank the reviewer for their suggestion. We have added two more steps to the analysis section to assist with data visualization and statistical analysis.

Responses to Reviewer #3's comments:

Reviewer Comment 1: Why is the bead beating extraction method superior to other options? In general it is at times difficult in the manuscript to know how novel these methods are and how they compare to other potential protocols that are in use elsewhere.

Author response: We appreciate the reviewer's thoughtful question and have added the following clarification to the introduction: "The combination of physical disruption of bacterial cell walls with bead-beating and chemical disruption with detergents allows rapid lysis of Gram-positive,

Gram-negative, and acid-fast bacteria without additional enzymatic digestion steps."

Reviewer Comment 2: The protocol is described as 'cost efficient, flexible, reliable and repeatable', but little of this is shown or discussed. Have they run replicates on the same samples and acquired similar data? What is the cost and how does it compare to alternatives?

Author response: We thank the reviewer for their comment and have addressed it by adding the following text to the discussion: "The complete cost (including all reagents, a single sequencing run, and primers but not equipment) is about $20 per sample when 200 samples are multiplexed. Additionally, there is very high reproducibility when multiple swabs from the same sample site are processed independently through the pipeline."

Reviewer Comment 3: It is unclear why dry swabs are being used.

Author response: Because this protocol is culture independent and does not require bacteria to remain viable, dry swabs are the most practical solution when collecting samples at a clinical site. Dry swabs have been used by others for microbiome analyses (*e.g.* Lauber *et al. FEMS Microbiology Letters*, 2010.) Placing the swabs into culture media or buffers introduces a contamination risk.

Reviewer Comment 4: The discussion around contamination is useful. Would the finding of identical (or near-identical) sequences in samples from 2 individuals also be suspect of contamination?

Author response: We thank the reviewer for their question. We would not suspect contamination if near-identical sequences were found in two samples. Because the vaginal microbiome is often colonized by a single bacterial species and the16S V4 region being amplified is about 300 base pairs, it is not unusual to find the same sequence in multiple samples.

Supplemental File 1

Click here to download Supplemental code file (if applicable): Supplemental File 1.xlsx