**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:

* 1. Visualize the data, *e.g.* by using a principal coordinates plot (e.g. EMPeror) or heatmap.
  2. Perform formal statistical comparisons of mapping file categories, *e.g.* with QIIME’s compare\_catagories.py script[29](#_ENREF_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 protocol20”, 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.