Friday, January 24, 2014

Sephorah Zaman, Ph.D.

Science Editor

JoVE

17 Sellers St., Cambridge, MA 02139

Subject: Manuscript re-submission

Dear Dr. Zaman,

Following reviewer’s comments, we have substantially modified our manuscript JoVE 51709 (see point by point revision below) and we would like to re-submit this manuscript entitled “Next-generation sequencing of 16S ribosomal RNA gene amplicons” and co-authored by Sylvie Sanschagrin and Etienne Yergeau for consideration in the Journal of Visualized Experiments. Ms. Sanschagrin optimized the protocols in the laboratory and wrote the protocol section, while I supervised her in the laboratory, conceptualized the protocols and wrote the remaining of the manuscript.

This protocol is ideally suited for JoVE’s unique multimedia format as it is a seemingly straightforward protocol, but with a lot of tricky steps. Most researchers also never have the opportunity to visualize the sequencing process itself as it is often carried out by an external sequencing center. In the manuscript we also mention that 16S rRNA sequencing is the gold standard for determining community composition in environmental microbiology, which should make our protocol interesting to a wide audience.

Most sincerely,



Etienne Yergeau, PhD

National Research Council of Canada,

6100 Royalmount Ave.

Montréal (Qc)

H4P 2R2, Canada

514-255-5584

Etienne.Yergeau@cnrc-nrc.gc.ca

**Editorial comments:**  
  
1) All of your previous revisions have been incorporated into the most recent version of the manuscript. Please download this version of the Microsoft word document from the "file inventory" to use for any subsequent changes.

Done.

2) Please do not underline any text in the protocol. Bold lettering, italics or a different font size may be used instead.

Corrected.  
  
3) Editor removed the company weblink from step 3.3.5 since JoVE is unable to publish manuscripts containing commercial sounding language. The calculator may instead be listed in the Table of Reagents/Equipment with the weblink (<http://ioncommunity.lifetechnologies.com/>) included in the “Comments” column of the table.

Ok.

4) Please specify the centrifugation speed in steps 3.6.4 and 3.6.5.

This centrifugation is performed in a micro-centrifuge in which it is impossible to set the centrifugation speed.

5) Please simplify step 3.1 of the protocol so that individual steps contain 2-3 actions per step and no more than 4 sentences per step. We recommend that you split your longer steps into sub-steps numbered accordingly, i.e. step 1 is followed by sub-step 1.1 then 1.1.1 as necessary.

Done.

6) After adding spaces between steps, the protocol length is over our 3 page limit for filmable content. Please use yellow highlighting to identify a total of 2.75 pages of protocol text (which includes headings and spaces) to identify which portions of the procedure are most important to include in the video, i.e. which steps should be visualized to tell the most cohesive story of your protocol steps. See JoVE's instructions for authors for more clarification and remember that the non-highlighted protocol steps will remain in the manuscript and therefore will still be available to the reader.

Done.

7) Please update the title of your article in the "License to Publish" form.

Done.

8) Please disregard the comment below if all of your figures are original.  
If you are re-using figures from a previous publication, you must obtain explicit permission to re-use the figure from the previous publisher (this can be in the form of a letter from an editor or a link to the editorial policies that allows you to re-publish the figure). Please upload the text of the re-print permission (may be copied and pasted from an email/website) as a Word document to the Editorial Manager site in the "Supplemental files (as requested by JoVE)" section. Please also cite the figure appropriately in the figure legend, i.e. "This figure has been modified from [citation]."

All our figures are original.

**Reviewer #1:**

*Manuscript Summary:*   
N/A  
  
*Major Concerns:*  
Most of the methods are not provided in sufficient detail to replicate, and most of the method that is described is already provided by the instrument manufacturer.

We disagree. We re-read carefully the methods and we find that most steps are sufficiently detailed. We could not reproduce protocols from the manufacturers and summarized most steps with enough details to be reproduced.

The choice of primer selection and bioinformatics methods are not adequately explained.

Our goal was to give an example of the analytical methods used and provide the reader with potential alternatives for primers, bioinformatic pipelines and even sequencing platforms. As such we don’t feel the need to justify extensively the choice of the methods, but we added references in the discussion for the readers to make their own choice of primer, sequencing technology and analysis pipeline.

*Minor Concerns:*  
N/A  
  
**Reviewer #2:**   
*Manuscript Summary:*   
The manuscript presents a workflow for analyzing bacterial community composition using 16S rRNA gene sequencing on the Ion Torrent DNA sequencing platform.   
  
*Major Concerns:*  
1. In the introduction, it would be helpful to include a few references that show some of the issues with PCR-based analysis of communities. This would be particularly helpful for researchers who are not yet familiar with the approach and the caveats.

We added two recent references related to PCR issues.

2. In the introduction, the authors suggest that depth of sequencing is not important for bacterial community analysis. This may be accurate if one is only interested in the most abundant taxonomic groups, but the I am sure the authors are aware that there is an exceptionally high diversity of less abundant species in all ecosystems and that deep sequencing has provided much insights into the composition of this "rare biosphere". The authors must add a couple references here (at least Sogin's original paper) and perhaps a more recent example (ie Gibbons et al. PNAS, Evidence for a persistent seed bank throughout the global ocean.

We already mentioned that environmental microbiology studies GENERALLY don’t need an extreme depth of sequencing. The rare biosphere is an interesting concept, but is far from the preoccupation of most microbial ecology studies, as it is not well defined what the role of these rare organisms might be. We therefore maintain our statement that extreme depth of sequencing is generally not required.

We prefer not referring to the Sogin paper as the rare biosphere detected there was subsequently shown to be caused by sequencing artefacts, see:

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12:** 118-123.

Huse SM, Welch DM, Morrison HG, Sogin ML (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12:** 1889-1898.

3. It would be helpful if the authors explained which region of the 16S rRNA gene they are targeting, and why this region was selected. Okay, I see that this issue is addressed in the discussion, but perhaps it should be moved to the introduction and a reference should be added that compares the usefulness of the different regions of the 16S. Pat Schloss has a reference that may be useful here.

We added the suggested reference and several sentences in this paragraph, but decided to keep this section in the discussion.

4. Other than the PCR, the protocol is based on kits. It would be helpful if the authors were clear about which kit was being used when. For example, 3.3 could say "Perfrom the library quantification using Qubit and DNA quantification kit XXX" Also, I don't see these reagents listed at the end of the manuscript! They should be included.   
  
[**Editorial Comment:** Please keep JoVE's protocol requirements in mind as you address the above comment as well as some of Reviewer 2 comments. No commercial sounding language should be added to the protocol, including company and brand names. All commercial products should be sufficiently referenced in the table of materials/equipment.]

The Qubit quantification was performed using custom-made oligos. As for the rest of the protocol, it is only the emulsion PCR and sequencing steps that rely on kits, the PCR amplification, purification quantification and data analysis are customized. All the kits used are mentioned in the reagent table.

*Minor Concerns:*  
na  
  
*Additional Comments to Authors:*  
na  
  
  
**Reviewer #3:**   
Review of Sanschagrin and Yergeau - "Next-generation sequencing of 16S rRNA gene amplicons"  
  
This chapter is a straightforward description of the workflow for generating 16S rRNA gene amplicon sequence libraries using the Ion Torrent Personal Genome Machine (PGM) platform. Details are provided for primer design, PCR amplification and sequencing. Some discussion of the 16S rRNA gene amplicon conceptual approach for microbial community characterization is also provided. The manuscript is acceptable, but not particularly novel; there are many such manuscripts describing this style of workflow. Specific comments are provided below.

We agree that similar worflows are detailed elsewhere. The novelty resides in the use of video for demonstrating this particular method.

L25: "...methods now ALLOW for the characterization..."

Corrected.

L27: Spell out ribosomal RNA the first time; also probably in the title.

Done.

L31-32: "...standards is to sequence 16S rRNA gene amplicons generated by domain-level PCR reactions amplifying from genomic DNA. Traditionally, this was performed by cloning and Sanger (capillary electrophoresis) sequencing of PCR amplicons."

Corrected.

L36: Perhaps remove personal pronouns from the text

Corrected.

L43: "An example analysis where the reads were classified with a taxonomy-finding algorithm within the software package Mothur...

Corrected.

L49: I'm a bit picky about the use of the word metagenomic for domain-level amplicon sequencing (i.e. I don't use it).

Recent reviews on metagenomics include amplicon sequencing.

L52: Not sure why amplicon sequencing is more straightforward that shotgun metagenomics. It's faster, for sure. Also, another advantage of amplicon sequencing (particularly 16S) is that the data analysis workflow is also really standardized.

We replaced straightforward by fast and added a sentence about standardization of methods.

L58: "...traditionally, 16S rRNA gene AMPLICON sequencing was..."

Corrected.

L60: "..sequencing on isolated plasmids; consequently, most studies analyzed FEWER than 100 clones per sample."

Corrected.

L68: "...Roche 454 pyrosequencing..."

Corrected.

L73: replace "less" with "fewer". Also "bases" instead of "bp"

Corrected.

L76-77: I think this is quite debatable; certain applications in microbial ecology demand very high sequencing output - for example, shotgun metagenomics. Here, I think small benchscale sequencers cannot match the output and price from the larger sequencers. 5-10 Gb per sample might easily be used for environmental metagenome.

We re-phrased that to specifically refer to amplicon, small genomes and low diversity metagenome sequencing.

L81-82: Refer to Ion Torrent PGM and Ion Torrent Proton. Proton is now generating 10-15 Gb with the PI chip v2; MiSeq is probably closer to 10 Gb, assuming 50 M reads of 200 bases (realistically, the 300 base reads generate 200 bases of quality data).

Corrected.

L86: I think "recently" is not the right word; this has been done for years already.

We disagree: benchtop sequencers are on the market for only 2-3 years.

L86: "a wide variety of environments from deep ocean sediments to human microbiome (REFs). More recently, the Ion Torrent PGM has also been used for community analyses of diverse environments as well."

We modified this sentence.

L101: "PGM"

We added the name of the sequencer used for the protocol.

L101: "After DNA extraction, 16S rRNA genes are amplified using domain-level bacterial primers that contain sequencing adapters and unique, sample-specific sequences (barcodes)."

Modified.

L103: "The DNA fragments in the pooled samples are then clonally amplified in an emulsion PCR, and then sequenced, Resulting sequences are then analyzed using publically available BIOINFORMATIC tools..."

Modified.

L111: Should indicate that this approach is called the fusion PCR method by Ion Torrent.

We modified the title of Step 1 to indicate this fact.

L114: Should explain what the sequencing key is. Also, reference is need for HotStartTaq; is this a proofreading polymerase?

Modified.

L111-116: Mention table of barcodes. I would recommend if using Ion Torrent barcodes to using those of the same length (if possible). Ion Torrent has 10,11 and 12 base barcodes, and this can be annoying for some methods to demultiplex. You may also want to recommend Golay-12 barcodes as recommended by Caporaso. Also, any barcode can be entered into the PGM or Proton system, and demultiplexing can be performed on instrument. Also, indicate what type of purification is used for the primers and approximate cost. What is the target group of the primer set? Do they pick up Archaea? Are there any known taxa that are definitely missed? Is there a reference for this primer set?

The different length of the barcode doesn’t cause problem in Mothur. We don’t recommend the primers of Caporaso for the Ion Torrent as they were not designed to be different in the Ion Torrent flow space as were the Ion Torrent recommended barcodes. We added some information in the discussion about the primers used. The cost for primer is negligible, as we get enough primer for ~1000 reactions at a cost of approximately $10. It is included in the PCR cost mentioned in the discussion.

L131: I personally think this is an insane approach - that is - to do gel purification for each sample. This might be tolerable for 10 samples, but what if you have 96? Or 500? So, not really feasible when you try to increase the throughput to take full advantage of the PGM using a 318 chip, for example. AMPure beads (or similar), for example, can be used in 96-well throughput. Particularly if the size of the PCR product is larger, then different concentrations of AMPure beads can be used to remove all fragments below 300 bp (0.6X buffer), for example. In addition, you might use other size selection of all samples pooled together in roughly equal quantities and then do gel purification together. This will lead to some variability in the number of reads per sample, but this appears to be unavoidable in any case.

We agree with the reviewer, and in fact we are using bead purification in the laboratory when sample numbers exceed 48. We added a sentence in the discussion to mention this alternative.

L141-144: We have found that the recommended concentrations of libraries are excessively high, and that we get very high polyclonals when running at the recommended concentrations. Our best runs have had much lower concentrations (at least ½ or less). We also typically use qPCR to accurately quantitate the library before emulsion PCR using KAPA kits.

In our hands, the concentrations indicated give adequate results.

L165: Need to indicate which lifetech kit is used; this isn't in the reagent table.

This is not a kit, we use custom made oligos.

L247-255: Should really define what all the parameters are.

The parameters are already defined in steps 4.2.1 to 4.2.3.

L289-L297: My feeling is that this is an underestimate of the cost of associated with this method. For one, the service contract on the instrument should be added. In addition, the price for primer synthesis (probably around $1/sample) should be added. Costs should be broken down more clearly - including the emulsion PCR, sequencing reagents, library quantification, etc. Also, if you use the 400 bp kit (for example - 515F/806R primer set) this will increase cost.

Several factors were left out of the equation to calculate costs: service contract, technician salary, laboratory space usage, instrument depreciation, etc. We wanted to indicate the cost in term of reagents, not in term of total cost, as this will vary widely from institution to institution. The price for primers is included in the price for PCR and we can get enough primer for 100-1000 reactions for around $10, which makes it negligible. We changes “library preparation” for emulsion PCR and added “sequencing reagents” to clarify the cost breakdown.

L299-308: I would indicate that actually Illumina MiSeq is the leader for amplicon sequencing at the moment. This appears to be the ability to handle hundreds of samples simultaneously, separate index read that is integrated with QIIME, increasing read length and paired-end reads.

We added some text related to this point in the discussion. In term of numbers of paper published, 454 is clearly the leader, but we agree that MiSeq might be a dominant player in the near future.

L317-329: This method should also be acknowledged to really only provide genus-level and higher classification; species is basically out.

This is now part of the discussion. We also mention that other genes might be better for species level information.  
  
Table 2: Maybe more useful than any details about a single run, you might put the ranges you have gotten from multiple runs.

We choose this run because it was a good average run. We mentioned in the representative results section that these are typical results that we get out of 16S rRNA gene amplicon runs.

Other thoughts:  
(1) Since this is specific to Ion Torrent, perhaps the title should be "Ion Torrent sequencing of 16S rRNA gene amplicons"; either that or include details of how to adapt this fusion method for Illumina and Roche 454 pyrosequencing approaches.

The title was modified following previous editorial comments. We added details about adapting this approach for Illumina and 454 in the discussion.