Chapter 7

Making and Sequencing Heavily Multiplexed, High-Throughput 16S Ribosomal RNA Gene Amplicon Libraries Using a Flexible, Two-Stage PCR Protocol

Ankur Naqib, Silvana Poggi, Weihua Wang, Marieta Hyde, Kevin Kunstman, and Stefan J. Green

Abstract

Deep sequencing of polymerase chain reaction (PCR)-amplified small subunit (16S or 18S) ribosomal RNA (rRNA) genes fragments is commonly employed to characterize the composition and structure of microbial communities. Preparing genomic DNA for sequencing of such gene fragments on Illumina sequencers can be performed in a straightforward, two-stage PCR method, described herein. The protocol described allows for up to 384 samples to be sequenced simultaneously, and provides great flexibility in choice of primers.

Key words Polymerase chain reaction, Microbiome, Next-generation amplicon sequencing, 16S ribosomal RNA, Illumina MiSeq

1 Introduction

In molecular surveys performed in the field of molecular microbial ecology, the small subunit (SSU; also referred to as the 16S or 18S) ribosomal RNA (rRNA) gene is most frequently targeted. A common approach to a microbial community study is to extract genomic DNA from multiple samples, PCR-amplify this DNA using primers targeting microbial rRNA genes, pool the samples in equimolar ratio, and perform high-throughput sequencing of all samples on a next-generation sequencing (NGS) platform, such as the Illumina MiSeq (e.g., [1]). PCR amplicons from each sample must have a unique identifier sequence (called a barcode or multiplex identifier, MID) to allow samples to be pooled prior to sequencing and then allow sample-specific sequences to be separated after sequencing. Multiple strategies can be employed to incorporate such barcodes, including ligation or PCR amplification. PCR

Nalini Raghavachari and Natàlia Garcia-Reyero (eds.), *Gene Expression Analysis: Methods and Protocols*, Methods in Molecular Biology, vol. 1783, https://doi.org/10.1007/978-1-4939-7834-2_7,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

incorporation is by far the dominant approach, due to the robustness of the method, ease of use, and low cost. We describe here a simple strategy to generate sequencer-ready SSU rRNA gene amplicons using a two-stage PCR protocol, sometimes called targeted-amplicon sequencing ("TAS") [2, 3]. The two-stage protocol is suitable for researchers that wish to multiplex heavily, limit optimization, and retain flexibility in choice of primer set. The described wet-lab protocol is not specific to SSU rRNA gene amplicons, and can be adapted as needed for many different amplification products.

2 Materials

Materials for conducting PCR and NGS sequencing are listed. However, the reader should be aware that in many cases, there are alternate products of similar quality that can be exchanged with no effect on the workflow. For example, there are many varieties of PCR mastermixes that can be used; the method is not wedded to the mastermix described herein. All reaction mixtures should be created in a clean environment (PCR hoods are recommended) as microbial 16S rRNA gene amplification reactions are easy to contaminate with ambient DNA or previous amplicons. In addition, the method described here is to be used explicitly for Illumina sequencing platforms, but similar approaches can be used to generate amplicons for Ion Torrent and Pacific Biosciences sequencers. Finally, the method described below begins with genomic DNA extracted from a sample (soil, water, feces, swabs, etc.), but DNA extraction is not explicitly covered in this chapter. Nonetheless, inhibitors present in DNA extracts can negatively affect the workflow and it is important to start with high quality genomic DNA. There is no explicit range of genomic DNA quantity that is required for the method; however, samples with low DNA input are more readily affected by background contamination. Samples containing very high levels of nontarget host DNA (e.g., human cellular DNA) in addition to microbial DNA can also create problems, often through the generation of nonmicrobial amplification products. Dealing with such issues is beyond the scope of this work. All reagents should be DNA and DNAse free. Large volume reagents (e.g., mastermixes) should be aliquoted to limit the number of freeze–thaw cycles, and to insulate against reagent contamination.

2.1 General Molecular Laboratory Equipment

1. AirClean® Systems AC600 Series PCR Workstations [not required, but advised; alternatives available].
2. Pipettors and disposable tips with filters [many options available].
3. PCR tubes/plates and optical covers [many options available].
4. PCR thermocycler [many options available].

5. Real-time PCR instrument [many options available].
6. TapeStation2200 and D1000 ScreenTape (Agilent) [alternatives available].
7. Qubit 3.0 Fluorometer (ThermoFisher Scientific) [alternatives available].
8. E-Gel Electrophoresis System (ThermoFisher Scientific) with 2% precast E-Gels and UV imager multiple [alternatives available].

2.2 Assessment of Quality and Quantity of Genomic DNA Prior to PCR Amplification

1. Qubit dsDNA BR or Qubit dsDNA HS (high-sensitivity) Assay Kit (ThermoFisher Scientific) [alternatives available].

2.3 Stage 1 and 2 PCR Reagents

1. MyTaq HS Mix, 2× (PCR mastermix, Bioline) [Alternatives available].
2. Access Array Barcode Library for Illumina (Fluidigm).
3. PCR primers containing CS1 and CS2 linkers—e.g., CS1_341F (ACACTGACGACATGGTTCTACACCTACGGGAGGCAGCAG) and CS2_806R (TACGGTAGCAGAGACTTGGTCTGGACTACHVGGGTWTCTAAT). Underlined regions represent linker sequences, while the remainder of the primer represents the template-specific region of the primer. Primers 341F and 806R have been frequently used to amplify the V3–V4 variable region of the microbial small subunit (SSU or 16S) ribosomal RNA gene [4–6]. Standard synthesis is acceptable. The gene-specific region of each primer (the nonunderlined regions) can be replaced with any primer sequence desired, but the underlined regions must be retained.
4. PCR water.

2.4 DNA Purification and Normalization Reagents

1. AMPure XP (Agencourt) PCR purification reagent (Beckman Coulter) [Alternatives available] [Magnet required].
2. SequalPrep Normalization Plate Kit, 96-well (ThermoFisher Scientific) [Alternatives available].
3. 1× TE buffer (a concentrated buffer comes with the SequalPrep kit).
4. Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher Scientific) [Plate Reader or fluorometer required] [Alternatives available].
5. Pippin Prep DNA Size Selection System [Sage Science] with 2% agarose cassette [alternatives available].

2.5 DNA Library Q/C

1. KAPA Illumina Library Quantification kit [alternatives available].
2. KAPA DNA standard 0.

2.6 Next-Generation Sequencing

1. Illumina MiSeq sequencer [Prepared libraries can also be sequenced externally at core facilities].
2. Illumina version 2 (v2) reagent kit (500 cycles) or Illumina version 3 (v3) reagent kit (600 cycles), and reagents therein (e.g., HT1).
3. 1 N NaOH.
4. Illumina PhiX Control v3.
5. Tris–HCl (pH 8.5 with 0.1% Tween 20)
6. CS1 sequencing primer containing locked nucleic acids (LNA) nucleotides. The sequence for primer CS1 is A + CA + CTG + ACGACATGGTTCTACA. The "+" precedes a LNA base. LNA primers can be ordered from Exiqon (www.exiqon.com/orderlna-oligos).
7. CS2 sequencing primer containing LNA nucleotides. The sequence for primer CS2 is T + AC + GGT + AGCAGAG ACTTGGTCT.
8. CS2rc (reverse complement) index sequencing primer contain LNA nucleotides. The sequence for primer CS2rc is A + GAC + CA + AGTCTCTGCTACCGTA.

3 Methods

The standard workflow for preparing samples for next-generation amplicon sequencing contains two independent PCR steps, which together generate PCR amplicons containing Illumina sequencing adapters, a single-index barcode sequence, and the region of interest. An overview of the two PCR stages is shown in Fig. 1. Briefly, the first stage of PCR amplification is used to generate amplicon products in the region of interest, and simultaneously, incorporating linker sequences (here, the CS1 and CS2 linker sequence from Fludigm) at the 5′ ends of all molecules. These linker sequences are essential to the subsequent barcoding strategy. After verification of successful PCR amplification, an aliquot of the first stage PCR product is used as a template for the second stage of PCR amplification. In most cases, no purification or removal of first stage primers is necessary. In the second PCR stage, we leverage the Fluidigm Access Array barcoded primers. These primers are sold in 4 × 96-well plates, and a total of 384 unique primer sets are provided. These Access Array primers provide flexibility, as they do

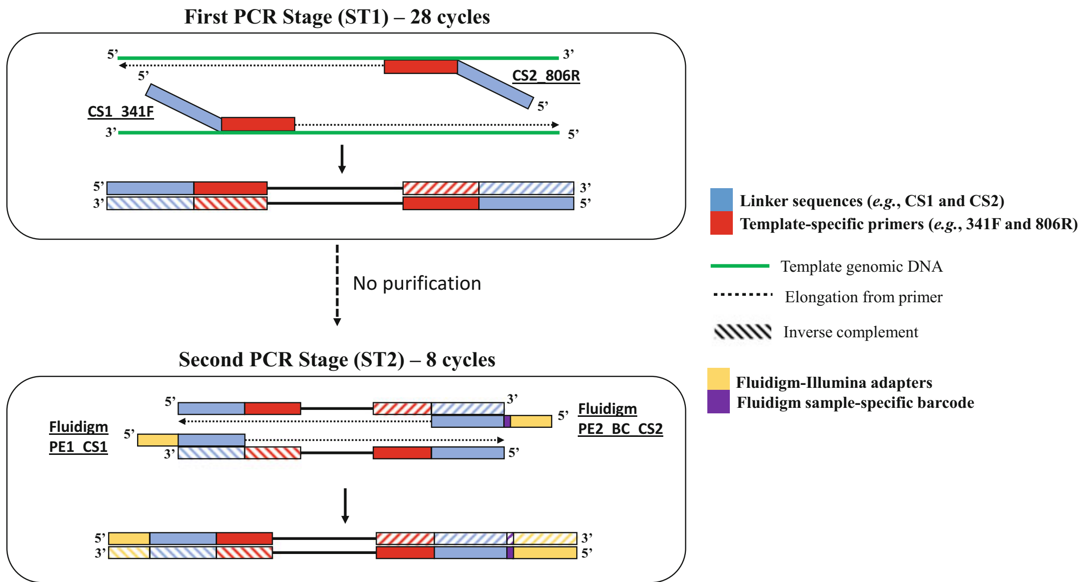

Fig. 1 Standard workflow for generating Illumina sequencer-ready PCR products. In the first stage of PCR (ST1), target-specific primers (here, primers 341F/806R, targeting the V3–V4 region of the bacterial 16S rRNA gene) which contain 5′-end linker sequences (i.e., CS1 and CS2 linkers) are used to generate amplification products in standard PCR reactions. Twenty-eight cycles are performed to generate a robust yield from the reaction. Subsequently, after verifying successful amplification, a portion of the amplification product is transferred to a second stage PCR reaction in which different primers are used. In the second stage, primers (here, AccessArray primers from Fluidigm) anneal to the linker regions that are present in all samples. Primers contain Illumina sequencing adapters, and the reverse primer also contains a sample-specific barcode of ten bases in length. Eight additional cycles are performed in ensure that the vast majority of amplicons contain Illumina sequencing adapters

not contain any gene-specific information; rather, they contain the CS1 and CS2 linker sequences at the 3′-ends (Fig. 1). Thus the first stage PCR products, containing CS1/CS2 linkers at the 5′-ends of the molecules, serve as suitable templates for amplification with the Access Arrays primers. Only a limited number of cycles of PCR amplification are required (typically 8 cycles are used), as the first stage of PCR is where the primary exponential amplification is performed, while the second stage is primarily used to incorporate sequencing adapters and sample-specific barcode into the final amplicons. One final note: the term “library” is used to mean DNA fragments containing Illumina sequence adapters and sample-specific barcode sequences together with the region of interest.

3.1 Stage One (ST1) PCR Amplification

1. Thaw reagents for standard PCR amplifications (*see* **Note 1**). Using the protocol herein, this will include 2× PCR master-mix, DNA-free water, and 10 μM solutions of gene-specific primers contain 5′-linkers (e.g., CS1_341F and CS2_806R) (*see* **Note 2**).

2. PCR amplifications are performed in 10 μl reaction volumes (*see* **Notes 3** and **4**). Per sample, the mastermix should contain the following:

   5 μl MyTaq HS Mix, 2×.

   0.5 μl 10 μM Forward Primer (e.g., CS1_341F).

   0.5 μl 10 μM Reverse Primer (e.g., CS2_806R).

   3 μl DNA-free water.

3. After the mastermix has been created for the appropriate number of samples, 9 μl of mastermix should be pipetted into individual wells of a 96-well plate or individual strip tubes. Next, place 1 μl of DNA-free water into the negative control reaction tube or well, and seal if possible. Subsequently, 1 μl of genomic DNA template from each sample should be added to a separate well. (~1–50 ng) (*see* **Notes 5** and **6**).
4. ST1 PCR cycling conditions are shown below (*see* **Note 7**).
   - 95 °C for 5′ [initial denaturation].
   - 28 cycles of the following:

     95 °C for 30″.

     50 °C for 30″.

     72 °C for 60″.
   - 72 °C for 7′ [final elongation].
5. Samples should be assessed for amplification yield using agarose gel electrophoresis (2% gels) and appropriate size standard (*see* **Note 8**). The expected band size is 510 bp for *E. coli* (GenBank: AB269763), including the CS1 and CS2 linkers. Size may vary slightly among different bacterial taxa. Expected yield should be reasonably robust, but not extremely strong (*see* **Note 9**). Negative controls should have no visible band of the correct size, but may contain a primer front or in some cases, primer dimers (*see* **Note 10**).
6. In most cases, no further processing of samples is necessary before proceeding to the next amplification step (*see* **Note 11**).

3.2 Stage Two (ST2) PCR Amplification

1. After generating ST1 PCR amplicons, and verifying clean negative controls, investigators can proceed to the second stage of library preparation. ST2 amplification is necessary to ensure that the PCR products in the reaction contain the PE1 and PE2 adapters that are necessary for cluster formation on Illumina sequencers. Standard PCR "best practices" should continue to be employed.
2. ST2 PCR amplifications are performed in 20 μl reaction volumes (*see* **Note 12**). Per sample, the mastermix should contain the following:

10 μl MyTaq HS Mix, 2×.

4 μl DNA-free water.

3. 14 μl of mastermix should be pipetted into individual wells of a 96-well plate or individual strip tubes.
4. 4 μl of Fluidigm Access Array primer pools (unique primer pair for each sample; primers are named PE1_CS1 and PE2_BC_CS2, where the BC represents a unique (384 total) barcode or index) is added to each well (*see* **Note 13**).
5. Two microliter of ST1 PCR amplification products (from Subheading 3.2, above) are added to each well (*see* **Notes 14–16**).

   PCR cycling conditions are as follows:

   - 95 °C for 5′ [initial denaturation].
   - 8 cycles of the following:

     95 °C for 30″.

     60 °C for 30″.

     72 °C for 30″.
   - 72 °C for 7′ [final elongation].
6. Samples should be assessed for amplification yield using agarose gel electrophoresis (2% gels) and appropriate size standard. The expected band size is 569 bp for *E. coli* (GenBank: AB269763), including the PE1 and PE2 adapters and sample-specific barcode (*see* **Notes 8**, **17**, and **18**).

3.3 Preparing Libraries for Sequencing

3.3.1 Purification, Equalization, and Concentration of ST2 PCR Amplicons

1. After the ST2 amplification has been completed, the samples are now ready to be purified and pooled together before sequencing on the Illumina MiSeq instrument. Purification is required to remove unincorporated primers as well as removing spent PCR reagents. Although many approaches can be used to purify PCRs, we have found that the 96-well SequalPrep normalization plate kits are relatively inexpensive, and have the added advantage of reducing variability in the concentration of purified amplification product. Each well in these plates has a limited but similar binding capacity, and as long as the total input DNA amount exceeds 250 ng, the plates should yield final concentrations within two to threefold of each other (*see* **Note 13**). Purifications should be performed according to the manufacturer's instructions, but no shakers, magnets or vacuum are required, and the kit comes with all necessary reagents. No specific size selection is performed by this purification step as the kit captures fragments from 100 bp to 20 kb (*see* **Note 19**).
2. In this second stage of normalization, PicoGreen quantification is performed on each sample purified using the SequalPrep protocol (*see* **Note 20**), using the Quant-iT PicoGreen dsDNA Assay Kit (*see* **Note 21**).

3. A reaction mix is prepared by diluting the PicoGreen reagent at a ratio of 1:400 using prepared 1× TE (component B (20× TE) is diluted with nuclease free water).
4. A standard dilution series is generated from a 100 ng/μl Lambda DNA stock solution (component C). The 100 ng/μl stock solution is serially diluted using 1× TE to working concentrations of 10, 5, 2.5, 1.25, 0.625, 0.3125, 0.156 and 0.078 ng/μl.
5. Each well of a microtiter plate is filled with 98 μl of this reaction mix and then 2 μl of either the SequalPrep purified DNA from above, or from Lambda DNA dilutions. Reactions are mixed by vortexing and incubated in the dark for 5 min. Plates are then read in a plate spectrofluorometer (excitation at 480 nm, emission at 520 nm), and the resulting fluorescent values are plotted for a standard curve and the unknowns compared to the curve to determine concentrations for each sample. In general, no replicates are performed. However, duplicates can help identify spurious pipetting errors that lead to improper equalization.
6. After DNA concentrations are measured for each sample going into the final pool, the investigator is ready to pool samples together (*see* **Note 22**) (*see* Fig. 2). This can be done by hand,

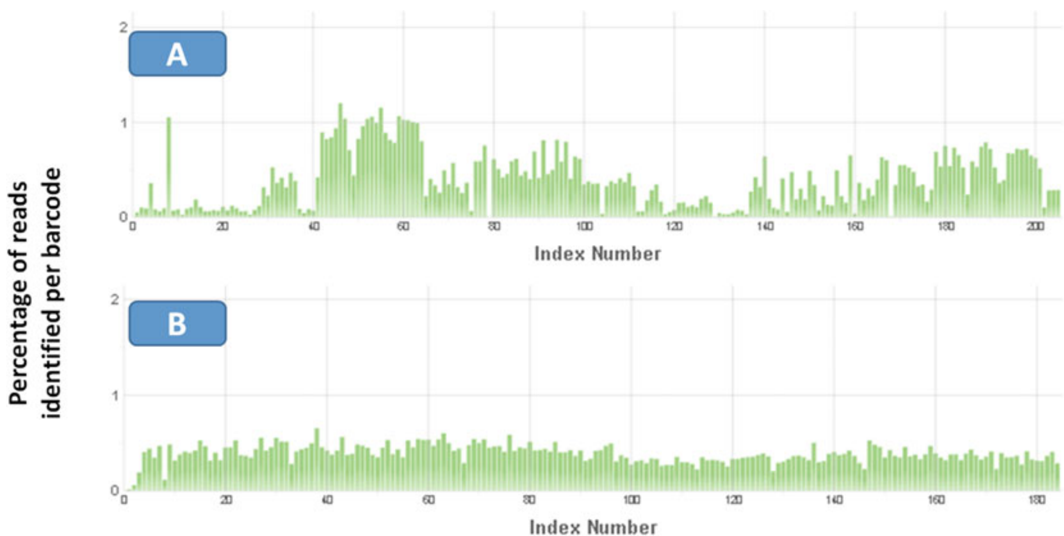

Fig. 2 Example of uneven (**a**) and even (**b**) distribution of reads among samples on two different MiSeq sequencing runs. Uneven distribution can result from pooling amplification products of different sizes and poor equalization of input DNA concentration from ST2 amplification products into the final library pool. Amplification products with nonspecific bands, strong primer dimers, and other artifacts can also contribute to measured concentrations, leading to uneven sequence distribution. Evenly distributed libraries, using equalization strategies described in the text, allow for more samples to be run on a single run, thereby decreasing per-sample costs, and improving robustness of downstream analyses

by pipetting samples of purified PCR products in volumes inversely proportional to their DNA concentration (*see* **Note 23**). For samples without any detectable DNA, the entire volume (~18 μl) may be used in an attempt to generate some sequence data from that sample (*see* **Note 24**). The total volume of the final pool will depend upon the number of samples pooled. The final DNA concentration of the pool is expected to be in the range of 1–2 ng/μl (i.e., the expected output concentration from the SequalPrep plates). For high-throughput pooling, a liquid handling robot is suggested (*see* **Note 25**).

7. The final pool generated in **step 3** can be quite a large volume (100–1000 μl), has a relatively dilute DNA concentration, and may contain nonspecific bands, including primer dimers (*see* **Note 26**). Purification of the pool using AMPure XP beads is necessary to address these three features, and is performed according to the manufacturer's instructions, with two modifications. First, a bead–sample ratio of 0.6× is used, instead of the standard 1.8× ratio (*see* **Note 27**), and this purification is performed twice in succession (*see* **Note 28**).
8. For the first cleanup, a volume of 200 μl from the full pool is used for purification with AMPure XP beads, using a bead-–sample ratio of 0.6× (i.e., 120 μl of XP reagent combined with 200 μl of pooled ST2 PCR products) (*see* **Note 29**). After the purification step is complete, the DNA is eluted in 50 μl of water.
9. The purification step is repeated, maintaining the 0.6× bead ratio (i.e., 30 μl of XP reagent combined with 50 μl of eluate from **step 5**). The final elution is again performed with 50 μl of water.
10. The final purified pool is then analyzed using gel electrophoresis (e.g., using an Agilent TapeStation with a D1000 ScreenTape; Fig. 3). There may be low levels of nonspecific amplification products visible (e.g., peaks at ~390 and 797 bp, Fig. 3) (*see* **Note 30**).

3.3.2 Quantification of the Final Library Before Sequencing

1. Prior to sequencing, the final purified pool of amplicons should be quantified using real-time, quantitative PCR (qPCR). Quantification is performed using the KAPA Library Quantification Kit for Illumina sequencers, according to the manufacturer's instructions, with several modifications.
2. First, sequencing libraries (i.e., pooled and concentrated ST2 amplicons) are prediluted to approximately 3 nM based on Qubit and TapeStation measurements. DNA concentration in nM is determined through Tapestation analysis by defining the entire amplicon peak in the Tapestation analysis software. The resulting peak area is calculated for ng/μl and nM concentrations.

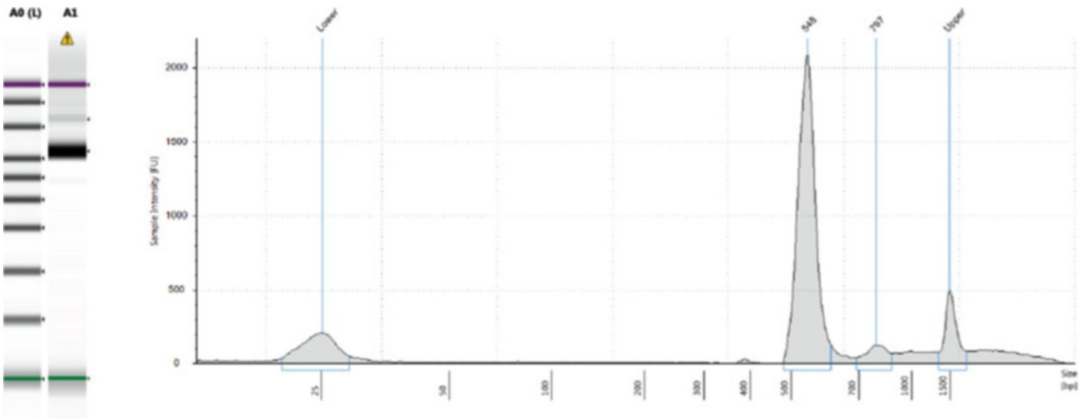

Fig. 3 Quality control analysis of a pooled amplicon library prior to Illumina MiSeq sequencing. The shown TapeStation electrophoresis analysis was performed using a D1000 Screen Tape. The primary peak, indicated as 548 bp by the TapeStation, represents the pool of amplicons generated from ST1 reactions using CS1_341F and CS2_806R primers, followed by ST2 reactions with Fluidigm Access Array primers. Peaks at 25 and 1500 bp represent spike-in standards. Smaller peaks at 797 and ~390 bp represent nonspecific amplification products. No additional size selection of this library was performed, due to the high ratio of target to nonspecific amplification products. Other library pools may favor further size selection and purification before sequencing. In some cases target and nontarget bands are too close for physical size selection, and bioinformatics strategies may be required for removal of unwanted sequences

3. This diluted library is then further diluted at 1:500, 1:1000, and 1:2000 in a serial manner. Each dilution is analyzed in the KAPA qPCR assay in duplicate and all values are used in the back calculation of the "stock" library concentration.
4. A standard curve utilizing a twofold dilution is prepared (instead of the order-of-magnitude dilutions as indicated in the KAPA instructions). We dilute the DNA standard 0, available from KAPA as a separate item, to make a twofold dilution standard curve between 20 and 0.3125 pM. In this manner, we can be assured that at least one of the three dilutions we prepare for each sample will fall within the standard curve even if the "stock" library concentration is substantially divergent from 3 nM. Library concentrations are calculated from the Ct obtained for each dilution and normalized using the size of each library as obtained from the TapeStation analysis (*see* **Note 31**). A final dilution is made to the stock library from above to a working concentration of 2 nM as calculated from the qPCR assay.

3.4 Loading the Sequencer

Loading of the sequencer should be performed according to the manufacturer's instructions, with some exceptions, described below. Sequencing of the 16S rRNA gene amplicons on Illumina sequencers is more difficult than standard sequencing (e.g., whole

genome sequencing). This is due to low sequence complexity in such amplicons, which creates problems for Illumina sequencers during cluster identification and focusing (e.g., [7, 8]). To address this concern, two approaches are taken. First, the pool of PCR amplicons is spiked with a prepared high complexity library (i.e., Illumina phiX viral DNA library) at a volume ratio of 10–30%. Second, the amount of total library loaded onto the sequencer is reduced, thereby generating fewer clusters but with higher quality. Alternative approaches are also possible, but beyond the scope of this study [7]. Finally, we will note that the MiSeq is relatively inflexible for high loading densities—and a total failure of the sequencing run is the result of overloading the sequencer, as opposed to simply lower quality sequence data. We typically aim for cluster densities in the range of 500–700K clusters per square mm (*see* **Note 32**). As mentioned previously, many core facilities have MiSeq instruments and will sequence prepared libraries as a service. Be sure to provide specific details of the prepared library to the core facility, ensure that they are aware of the custom sequencing primers needed (see below), and notify the facility of the low complexity of the amplicon pool. Below, an overview of the steps taken to load MiSeq sequencers is shown; these steps apply to both Illumina MiSeq V2 and V3 kits.

1. *Denature and dilute library of pooled PCR products.*
   (a) Prepare a fresh solution of 0.2 N NaOH. Mix 200 μl 1 N NaOH with 800 μl of DNA-free water.
   (b) Combine 5 μl of 0.2 N NaOH with 5 μl of 2 nM pooled library. [Note: 2 nM concentration should be determined using quantitative PCR, as described above.]
   (c) Incubate at room temperature for 5 min.
   (d) Add 990 μl of chilled hybridization buffer (HT1; provided with MiSeq sequencing reagents) to the 10 μl volume from step c. This creates a concentration of 10 pM.
   (e) Further dilute denatured DNA, as necessary, to a desired final working concentration (generally from 4.5 to 8 pM for amplicon libraries). We recommend starting at the lower end of this concentration range, and increasing slowly over several runs to get to the desired loading density. Individual MiSeq instruments vary, and no single concentration is appropriate for all libraries and sequencers (*see* **Note 33**). A total volume of 600 μl is needed (e.g., for a 5.5 pM loading, add 330 μl of the 10 pM denatured library with an additional 270 μl of chilled HT1 buffer).

2. *Denature and dilute phiX library.*
   (a) Prepared phiX libraries, purchased from Illumina, are diluted and denatured.
   (b) Combine 3 μl of 10 mM Tris–HCl (pH 8.5 with 0.1% Tween 20) with 2 μl of 10 nM phiX stock library to obtain 5 μl at 4 nM.
   (c) Combine 5 μl of 0.2 N NaOH with the 5 μl 4 nM phiX and Tris–HCl (from **step b**)
   (d) Incubate at room temperature for 5 min.
   (e) Add 990 μl of chilled hybridization buffer (HT1; provided with MiSeq sequencing reagents) to the 10 μl volume from **step d**. This creates a concentration of 20 pM. Such a dilution can be stored at −20 °C for up to 3 weeks; after this, fresh dilutions should be prepared.
   (f) The phiX library is then further diluted to 12.5 pM by combining 375 μl 20 pM denatured phiX library with 225 μl of chilled HT1.
   (g) The diluted phiX library is then mixed with the denatured pool of PCR products (amplicon libraries) at an appropriate ratio. In general, we aim to achieve 15–20% phiX volume ratio, though lower (or higher) phiX levels can be used. For example, in our standard approach, 90 μl of the 600 μl of denatured library (**step 1e**, above) is removed and replaced with 90 μl of diluted phiX library generated in **step 2f**. Please note that the exact amount of phiX will vary depending upon the size of the amplicon and the concentration determined in **step 2d**, above (*see* **Note 34**).
3. *Loading the MiSeq sequencer.*
   (a) Loading of the MiSeq v2 or v3 kits should be performed according to the manufacturer's instructions, with modifications of the sequencing primers. PCR products generated with Fluidigm Access Array primers, as detailed in this protocol, require custom sequencing primers. For Access Array sequencing, the LNA-containing primers CS1, CS2, and CS2rc are required for the forward read, reverse read, and index read, respectively (*see* **Note 35**).
   (b) Sequencing primers specific for the CS1, CS2, and CS2rc must be combined with the default Illumina sequencing primers contained within the MiSeq reagent cartridge. The default primers contained in the kit are required to initiate sequencing of the phiX library, and the CS1, CS2, and CS2rc primers are required to initiate sequencing of the custom amplicon libraries generated in ST2. Stocks of each of the CS1, CS2, and CS2rc primers are prepared at a

concentration of 100 μM. The Illumina sequencing primers provided in the reagent cartridge are found in positions 12, 13, and 14. These primers are removed from the reagent cartridge and added to individually labeled Eppendorf tubes. To the contents of position 12, add 4 μl of 100 μM CS1 primer (read 1), and mix well. To the contents of position 13, add 4 μl of 100 μM CS2rc primer (index read), and mix well. To the contents of position 14, add 4 μl of 100 μM CS2 primer (read 2), and mix well. Finally, the entire contents of the prepared primer tubes are added back to their original positions in the reagent cartridge.

(c) Finally, the prepared library (pooled ST2 and phiX) is loaded onto the MiSeq reagent cartridge, according to the manufacturer's instructions. All other standard loading procedures should be followed. A sample sheet must be provided to the sequencer for running and demultiplexing of the barcoded reads. An example sample sheet is shown in Fig. 4, with the minimum required information. The exact index sequences for all 384 unique indices are provided by Fluidigm.

(d) Once sequencing is complete, data are available in Illumina's cloud storage system, BaseSpace. Alternatively, data can be stored locally and transferred using mobile external hard-drives prior to bioinformatics analyses. In our standard bioinformatics pipeline, data are demultiplexed in

[Header]							
Investigator Name	multi						
ProjectName	MiseqXXXX						
Experiment Name	MiseqXXXX						
Date	3/28/2018						
Workflow	GenerateFASTQ						
Assay	null						
Chemistry	Default						
[Reads]							
301							
301							
[Settings]							
[Data]							
Sample_ID	Sample_Name	Sample_Plate	Sample_Well	I7_Index_ID	index	Sample_Project	Description
Sample1_Name	Sample1_Name			FLD0001	GTATCGTCGT	MiSEqXXXX_Project_Name	
Sample2_Name	Sample2_Name			FLD0002	GTGTATGCGT	MiSEqXXXX_Project_Name	
.	.			.	.	.	
.	.			.	.	.	
.	.			.	.	.	
Sample384_Name	Sample384_Name			FLD0384	ACGGTGCTAG	MiSEqXXXX_Project_Name	

Fig. 4 An example sample sheet for Illumina MiSeq sequencing. Shown is a sample sheet for v3, 2 × 300 base sequencing, with up to 384 samples pooled together

BaseSpace, leading to the production of two FASTQ files per sample (forward read and reverse read). In some strategies, all samples will be pooled together and the sequencer will generate a forward read file, a reverse read file, and an index read file. Both data output strategies can be handled using the software package QIIME [9].

4 Notes

1. Standard PCR "best practices" should be employed. PCR mastermixes should be made in a PCR hood or otherwise clean environment. Reaction mastermixes should be made with enough volume for all samples, positive and negative controls, and several excess reactions to allow for slight pipetting errors. A fresh, unopened tube of DNA-free water should be used. Likewise, it may be helpful to aliquot PCR mastermixes and primers into single-use volumes to reduce the potential for reagent contamination.
2. Primers can be ordered "LabReady" from IDT, and will arrive dissolved at 100 μM concentration. This is quite convenient, and also can reduce the possibility for contamination of the primers with ambient microbial genomic DNA.
3. We find that 10 μl reaction volumes are sufficient for the first stage of PCR, since we use a small aliquot of this PCR yield for the second stage of PCR. However, if you have optimized amplification conditions using other volumes, then you may continue to use those volumes with no adverse effects. We have found that the addition of CS1/CS2 linkers does not usually alter any of the standard PCR conditions. In some cases, however, PCR conditions may need to be adjusted when adding the linker sequences. In general, annealing temperature can be altered—generally increased; here, we recommend running a temperature gradient. For primers that have degeneracies, we recommend using the lowest annealing temperature that does not generate spurious bands when visualized by agarose gel electrophoresis.
4. Other additives, such as BSA or $MgCl_2$ can be added as needed, and the amount of water adjusted accordingly. We prefer to avoid additives, if possible, and attempt to get amplification either through dilution of hard-to-amplify templates (i.e., diluting out PCR inhibitors) or additional purification of samples.
5. The exact amount of DNA can vary substantially from sample-to-sample, and the DNA concentration is frequently not equalized among samples. We note, however, that some

studies have shown that genomic DNA input into PCR reactions can affect the observed microbial community structure [10]. Other considerations should be taken into account. Some sample types are extremely rich in bacterial DNA (e.g., feces) while others may have substantial host-associated or other nontarget DNA in the sample (e.g., lung tissue). Thus, 10 ng of genomic DNA does not always represent roughly the same number of target gene copies. Ultimately, the ratio between target and nontarget is important, and in samples where nontarget DNA is extremely abundant, double-banding is often observed due to the generation of nonspecific amplicons generated from nontarget DNA. There are various strategies (beyond the scope of this work) for removing this unwanted amplification product, but readers may consider (a) size selection to remove unwanted bands, (b) host DNA depletion through removal of methylated DNA (e.g., NEBNext® Microbiome DNA Enrichment Kit), (c) blocking primers (e.g., Earth Microbiome 18S Illumina Amplicon Protocol, with Mammalian blocking oligonucleotide—http://www.earthmicrobiome.org/protocols-and-standards/18s/) [11, 12], and selective restriction digests [13]. Finally, complementary DNA generated from total RNA using a 16S rRNA gene-specific primer will likely contain an extraordinarily high proportion of gene targets and should be titrated to avoid overly strong PCR amplification. In our protocol, we prefer modest amplification yields from the first stage of PCR amplification—this can reduce the formation of chimeric sequences [1]. For fecal samples, we also generally dilute 1/10 the input genomic DNA because (a) the feces are highly enriched in microbial DNA with a large number of targets, and (b) even samples extracted with modern extraction kits still tend to retain some PCR inhibitors.

6. We can also recommend additional controls, if necessary. These can include (a) additional negative control reactions, (b) multiple (3–5) technical replicates of a single sample to assess technical reproducibility of the methodology, and (c) mock community DNA (e.g., ZymoBIOMICS™ Microbial Community Standards from Zymo Research or standards in development—*see* [14]).

7. We recommend that no more than 28 cycles are used during the first stage (ST1). However, under some circumstances, it may be difficult to get sufficient yield during the ST1, and additional cycles are necessary. However, please note that eight additional cycles will be performed during the second stage (ST2), and thus, weak yields are generally not problematic. Intense bands generated with 28 cycles of amplification are not preferred in the protocol, and we recommend either

(a) diluting the input genomic DNA before PCR with 28 cycles, or (b) reducing the number of cycles until a modest band is achieved. There is no specific minimum number of cycles (other than 2; *see* [2]), but it often is comforting to be able to see some band!

8. For projects with large numbers of samples with similar types of genomic DNA, a random selection of samples may be sufficient for assessing success.
9. Nonspecific bands in samples may indicate too high input DNA, cycle number, or magnesium concentration in PCR buffer. Testing a small number of samples prior to a full-scale experiment is recommended for any new primer set or sample type.
10. Please note that even negative controls that do not produce a band in agarose gel electrophoresis are likely to produce some sequence data. This is an unfortunate feature of NGS-amplicon sequencing. On the bright side, this can also sometimes mean that samples that have a very low yield can still generate enough sequence data to be meaningful.
11. In some cases, where extremely strong primer dimers or nonspecific amplicons are produced, purification of the first PCR product may be necessary. If nonspecific amplicons smaller than 300 bp are produced, and the amplicon of interest is greater than 300 bp, then a 0.6× AMPure XP cleanup can be performed. Under these conditions, DNA fragments smaller than 300 bp are not bound well, and are removed, while the longer amplicons are bound to the beads and retained for down-stream amplification steps.
12. Larger amplification reaction sizes are used to ensure sufficient PCR product for the normalization steps employing SequalPrep plates (below, in Subheading 3.4). Smaller reaction sizes can be used if extremely strong PCR amplification yields are observed, but this is not recommended. Other normalization approaches may not require large amounts of DNA; therefore, smaller PCR reaction volumes can be employed. The total mastermix should contain extra reactions for reagent loss and for reaction negative controls.
13. Multichannel pipettors or 96-well head pipettors (e.g., Eppendorf Ep96) can greatly increase the processing speed. Pipette tips must be replaced—crossover contamination of barcoded primers will create mixed signals that can be very hard to detect. Extra care should be taken when pipetting barcoded primers.
14. Strong amplification products from ST1 can be diluted before adding to the ST2 reactions; this may reduce chimera formation.

15. In some cases, additional cycles can be performed for ST2 to increase the final yield from these reactions. When performing normalization using SequalPrep plates, approximately 250 ng of total DNA is required. Thus, additional cycles or larger volume reactions may be needed for some samples to generate sufficient DNA. However, it is not advised to run different number of cycles for samples from the same project, if it can be avoided.
16. Negative control samples from ST1 should also be used as templates for ST2 reactions with unique barcodes. In addition, separate control ST2 reactions (i.e., ST2 reactions without any input DNA) should be generated with unique barcodes.
17. In general, the ST2 reaction is highly robust. When failing to generate sufficient PCR product, the problem usually originates during ST1. Since the input PCR product into ST2 usually has a high concentration, it is rare that the reaction fails completely. The yield from the ST1 reaction should be investigated when trying to troubleshoot failed reactions.
18. In general, we recommend that for sequencing on Illumina MiSeq V3 chemistry (600 cycles), the final PCR product from ST2 amplification should not exceed approximately 620 bp if merging of forward and reverse reads is required (including CS1 and CS2 linkers). For v2 chemistry (500 cycles), the final PCR product from ST2 amplification should not exceed approximately 570 bp, if merging of forward and reverse reads is required.
19. Although the SequalPrep normalization plates improve the overall evenness of DNA concentration per well, our experience is that this normalization is not sufficient for heavily multiplexed sequencing efforts where more than 96 samples are processed or where amplicons generated from different primer sets are pooled together. Thus, further equalization is performed. Please also be aware that each 96-well SequalPrep plate can behave differently from other plates. Thus, after SequalPrep plate cleanup, samples from within single plates tend to be very similar in concentration while between plates, there tends to be substantial variability. Samples from two different plates should not be mixed without additional quantification of final DNA concentration.
20. Even if the SequalPrep protocol is not used, ST2 PCR products must be purified prior to quantification using PicoGreen (or any other dye-based or absorbance-based quantification protocol). We have found that SequalPrep is the cheapest and easiest protocol for performing this in high-throughput.

21. For large sample datasets, a plate reader is a huge boon; otherwise, one-by-one analysis of DNA concentrations is extremely laborious.
22. Here, we use a strategy in which the DNA concentration is quantified in each sample, and then samples are pooled based on measured concentration. This strategy does not completely remove variability in the number of reads generated per sample, but greatly increases the overall evenness. Examples of poorly balanced and well-balanced pools of libraries are shown in Fig. 2.
23. In general, volumes of 2–10 μl are used per sample, to ensure pipetting accuracy. For samples with high DNA concentrations, dilution is used to bring the concentration into a more reasonable range, where 2–10 μl are needed for incorporation into the final pool. Pipetting smaller volumes will lead to more variability due to increased pipetting error at small volumes. We also note that pooling based on concentration is effective when all amplicons are the same size. When mixing amplicons of different size, or mixing amplicons with multiple bands (e.g., fungal internal transcribed spacer (ITS) amplicons), equal concentration does not always yield equal sequence output. Unfortunately, adjusting the ratio of different sized amplicons needs to be tested empirically. Larger amplicons will need a higher DNA concentration than shorter amplicons, but the relative yield of each sample will be influenced by the size of the other amplicons sequenced at the same time.
24. We do not recommend that the entire volume of negative controls (i.e., blanks from ST1 propagated through ST2, and ST2 blanks) be included in the final pool. Using the entire volume from blanks in the final pool can lead to situations where a substantial number of reads are produced, leading to the appearance of high negative control contamination. We recommend using only 2 μl of negative control samples as part of the final pool.
25. For example, our laboratory uses an epMotion5075 liquid handling instrument (Eppendorf) to transfer PCR products from 96-well plates to a single 1.5 ml microfuge tube. However, this can also be performed by hand.
26. Please note that samples with low concentration are used up more rapidly, and this can be an issue if you need to make a new pool in the future. Thus, when possible, use less than 10 μl of sample (as suggested)—thereby leaving 8 μl remainder for future pools. High concentration samples are never limiting.
27. The 0.6× ratio should only be used when the desired amplicon has a size larger than 300 bp. Other purification strategies can be used when the desired fragment is smaller than 300 bp

(PippinPrep size selection, different AMPure XP concentration, agarose gel electrophoresis and band excision, etc.). To remove nonspecific bands larger than the band of interest usually requires gel excision or PippinPrep automated size selection.

28. Although slightly annoying, the double purification is generally helpful to remove as much of the residual small nonspecific amplification products, primer dimer, and unincorporated primers as possible. This generally has the effect of improving sequence quality and yield.
29. In general, the entire volume of the pool yields a lot more DNA than is needed for sequencing. Thus, only a portion of the final pool is used for purification–thereby saving on AMPure XP reagents. In addition, when possible, 0.6× AMPure XP ratio is preferred, both due to the removal of unwanted small DNAs, but also for the reduced reagent cost.
30. Low levels, such as seen in Fig. 3, do not contribute substantially to the final sequencing output, and can be reasonably ignored. Fragments larger than the amplicon of interest tend to contribute least of all to the sequencing output, while fragments smaller than the amplicon of interest can be problematic when present at a more elevated level. If the nonspecific amplicons appear to reach a level of concern, additional purification strategies may be required (Pippin Prep size selection, gel excision, etc.). The exact level of concern will, unfortunately, need to be determined empirically.
31. Experienced users may be able to use Qubit quantification results directly for loading, without need for qPCR. However, this approach should only be used once a clear relationship between Qubit concentration measurements and MiSeq loading density has been made for specific amplicons. For example, we have developed a formula for amplicons of approximately 400 bp in size (the Earth Microbiome Project 515F-806R primers). Briefly, the Qubit concentration in ng/μl is multiplied by 1830.77 to generate an estimated pM concentration. The library concentration is adjusted to 2 nM and then the standard sequencing workflow is performed. Please note that formula may not work for your amplicons and sequencer, and should be determined empirically from multiple sequencing runs in which both qPCR and Qubit analyses are performed.
32. Anecdotally, we have observed that the Illumina MiniSeq is more tolerant of high loading densities, and can be loaded in similar fashion to nonamplicon libraries. This instrument, however, can only run paired-end 150 base sequencing (i.e., 300 cycle chemistry). Thus, it is mostly suitable for short amplicons, such as the 515F/806R primer set described previously [5].

33. In general, smaller amplicons (<400 bp) should be loaded at lower densities than larger amplicons.
34. Empirical testing should be performed to adjust the ratio of phiX spike-in to the level of 15–20%; initial efforts should aim for higher phiX to guarantee success of the sequencing run. Subsequently, the amount of phiX can be decreased.
35. If alternate barcoding strategies are used—such as Illumina's Nextera barcoding strategy—custom primers may not be required.

Acknowledgments

We have been assisted over the years in sequencing amplicons on Illumina MiSeq instruments by Chris Wright and Dr. Hernandez Alvaro of the High-Throughput Sequencing and Genotyping Unit at the University of Illinois at Urbana-Champaign, and we gratefully acknowledge their assistance.

References

1. Ionescu D, Overholt WA, Lynch MD, Neufeld JD, Naqib A et al (2016) Microbial community analysis using high-throughput amplicon sequencing. Manual of environmental microbiology, fourth edition. American Society of Microbiology, Boston, p 2.4. 2-1-2.4 2-26
2. Green SJ, Venkatramanan R, Naqib A (2015) Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches. PLoS One 10:e0128122
3. Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL et al (2011) Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biol Evol 3:1312–1323
4. Muyzer G, De Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol 59:695–700
5. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA et al (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci 108:4516–4522
6. Bartram AK, Lynch MD, Stearns JC, Moreno-Hagelsieb G, Neufeld JD (2011) Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. Appl Environ Microbiol 77:3846–3852
7. Wu L, Wen C, Qin Y, Yin H, Tu Q et al (2015) Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. BMC Microbiol 15:125
8. Krueger F, Andrews SR, Osborne CS (2011) Large scale loss of data in low-diversity illumina sequencing libraries can be recovered by deferred cluster calling. PLoS One 6:e16607
9. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336
10. Kennedy K, Hall MW, Lynch MD, Moreno-Hagelsieb G, Neufeld JD (2014) Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. Appl Environ Microbiol 80:5717–5722
11. Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of

small-subunit ribosomal RNA genes. PLoS One 4:e6372

12. Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples—a case study on prey DNA in Antarctic krill stomachs. Front Zool 5:12
13. Green SJ, Minz D (2005) Suicide polymerase endonuclease restriction, a novel technique for enhancing PCR amplification of minor DNA templates. Appl Environ Microbiol 71:4721–4727
14. Tighe S, Afshinnekoo E, Rock TM, McGrath K, Alexander N et al (2017) Genomic methods and microbiological technologies for profiling novel and extreme environments for the extreme microbiome project (XMP). J Biomol Tech 28:31–39