Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An optimized protocol for high-throughput amplicon-based microbiome profiling

Daryl M. Gohl (dmgohl@umn.edu)
University of Minnesota Genomics Center, Minneapolis, MN, USA
Allison MacLean
University of Minnesota Genomics Center, Minneapolis, MN, USA
Adam Hauge
University of Minnesota Genomics Center, Minneapolis, MN, USA
Aaron Becker
University of Minnesota Genomics Center, Minneapolis, MN, USA
Dinesha Walek
University of Minnesota Genomics Center, Minneapolis, MN, USA
Kenneth B. Beckman
University of Minnesota Genomics Center, Minneapolis, MN, USA

Method Article

Keywords: Microbiome, Next-generation sequencing, PCR, amplicon

Posted Date: July 25th, 2016

DOI: https://doi.org/10.1038/protex.2016.030

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Here we describe an optimized protocol for amplicon-based microbiome profiling by next-generation sequencing \(NGS). The protocol uses a modular, two-step PCR process that provides versatility as a common set of indexing primers can be paired with many sets of marker gene specific primers. We describe primer sets that have been used to prepare libraries for variable regions V1-V3, V3-V4 V3-V5, V4, V4-V6, and V5-V6 of the 16S ribosomal RNA \(rRNA) gene, as well as for variable region V9 of the 18S rRNA gene and the ITS1 and ITS2 regions. This protocol is designed for high-throughput and incorporates the best practices for amplicon-based NGS library preparation that we have previously described. We have demonstrated that these best practices substantially increase accuracy compared to existing methods and also, importantly, minimize the dropout of taxa that have mismatches to the amplification primers.

Introduction

Next-generation sequencing \(NGS) technologies have enabled a revolution in the ability to analyze and quantify the microbiome. Two prominent experimental approaches are typically used to assess microbial diversity; marker gene surveys, where a portion of a highly conserved gene such as the 16S ribosomal RNA \(rRNA) gene is amplified and sequenced to identify the organisms that make up a microbial community, and shotgun metagenomics, where whole genome sequencing is performed on material isolated from a mixed microbial community. Amplicon-based approaches have the advantage of being less expensive \(both in terms of library preparation costs and sequencing costs), allowing much deeper and more sensitive profiling of a microbial community at a given cost. In addition, for samples such as biopsies, where there is a low abundance of microbial DNA compared to host DNA, the ability to dramatically enrich for microbial sequences using PCR make marker gene surveys much more attractive. Finally, shotgun approaches are heavily reliant on reference databases, so for eukaryotes/fungi or uncharacterized bacterial species, amplicon-based approaches allow for a broader range of detection. However, it has been widely recognized that amplicon-based marker gene surveys are prone to error and bias at many steps[1–5]. For this reason, methods for preparing microbiome amplicon libraries for NGS have proliferated. We have systematically compared a number of these methods and dissected the parameter space of the amplification process, providing insights into the mechanistic basis of several sources of error and bias[6]. Through these experiments, we have compiled a list of best practices for amplicon-based marker gene surveys, which improve accuracy, reduce PCR artifacts such as chimeras, and minimize drop-out of taxa due to mismatches between the amplification or sequencing primers and the marker gene template sequence: These recommendations are as follows: 1) Use a proofreading polymerase. 2) Use a highly processive polymerase. 3) Do not use sequencing primers that overlap with the amplification primers. 4) Optimize template concentration. 5) Minimize PCR cycle number. Here we present a protocol that incorporates these best practices. This protocol uses a two-step PCR process, where the marker gene of interest is amplified with a pair of adapter-tailed primers in a primary PCR reaction, and sample-specific dual indices and flow cell adapters are added in a subsequent indexing reaction \(Figure 1)[7]. We describe primer sets that have been used to prepare NGS libraries for variable regions V1-V3, V3-V4, V3-V5, V4, V4-V6, and V5-V6 of the 16S ribosomal RNA \(rRNA) gene, as well as for variable region V9 of the 18S rRNA gene and the ITS1 and ITS2 regions, as well as a set of dual indexing primers that is compatible with all of the marker gene primer sets. The modularity of this design makes it easy to survey new regions of interest,

since only a single pair of adapter-tailed locus-targeting primers needs to be synthesized. This protocol incorporates automation using liquid-handling robots for sample template optimization and normalization. This automation is not required to carry out this protocol, but enables this protocol to be carried out at very high throughput.

Reagents

1) 384-well plate. \(Deepwell or regular; dilution plate) \(Eppendorf/Thermo Scientific) 2) 384-well Barcoded plate. \(Thermo Scientific) 4) 96-well plates. \(BioRad) 5) Nuclease-free water. \(Fisher Scientific) 6) Microseal F \(foil) seals. \(BioRad) 7) Optical Seals \(qPCR). \(Applied BioSystems) 8) KAPA HiFi HotStart PCR Kit with dNTPs. \(KAPA) 9) Primer set for primary PCR \(marker gene variable region). \(IDT) 10) Indexing primer set for indexing PCR. \(IDT) 11) DMSO. \(Fisher Scientific) 12) ROX. \(Life Technologies) 13) SYBR. \(Life Technologies) 14) White Matrix troughs. \(Thermo Scientific) 15) Microseal B PCR Film. \(BioRad) 16) Indexing oligo plate. \(IDT) 17) 1.5mL microfuge tubes. \(Ambion) 18) AmPure XP Beads. \(Beckman Coulter) 19) Magnetic Plate. \(Life Technologies) 20) Ethanol, 200 proof. 21) 50mL Conical tubes. \(Corning) 22) PicoGreen reagents. \(Life Technologies) 23) Agilent DNA HS chips. \(Agilent) 24) Caliper LabChip XT DNA 750 chips. \(Caliper) 25) MiSeq sequencing reagents. \(Illumina)

Equipment

1) Multichannel and single channel pipettes. 2) Rainin Liquidator 96 pipette with p20/p200 tips. \(Rainin) 3) Rainin single/multichannel pipettes with p20/p200/p1000 tips. \(Rainin) 4) BioRad Tetrad thermocycler, or equivalent. \(BioRad) 5) Agilent Bioanalyzer. \(Agilent) 6) MiSeq. \(Illumina) 7) Platform shaker. 8) Centrifuges \(for plates and tubes). 9) Plate reader \(for PicoGreen). - Optional 10) Thermo Matrix Pipette with 30 µl tips. \(Thermo Scientific) - Optional 11) ABI 7900. \(Applied BioSystems) - Optional 12) EpMotion 5075 Robot with 50 and 300 µl tips. \(Eppendorf) - Optional 13) Caliper XT. \(Caliper) - Optional

Procedure

1) Primary PCR amplification The first step of this protocol can be carried out using option A or B, depending on whether the user wishes to perform the template concentration optimization step \(A - recommended for new sample types or unknown samples) or not \(B - recommended for samples with known and consistent template concentrations). **CRITICAL STEP:** In order to minimize the potential for cross-contamination of samples, for all sample transfer steps, ensure that the liquid in the plate has been completely spun down and remove the seals slowly to avoid aerosolization. Likewise, for all sample mixing steps, make sure each individual well of the plate is well sealed. **1A.I) Creating sample dilution plate** **CRITICAL STEP:** For low abundance samples, sample dilution may result in population bottlenecking. Optimal results are achieved by minimizing PCR cycles and selecting the most concentrated samples possible. Potential bottlenecking can be assessed by sequencing multiple dilutions. i) Vortex samples to mix at 1900RPM for 15 seconds and spin-down on centrifuge. ii) Using the Liquidator or a multichannel pipet, dispense 10 µl of undiluted sample into quadrant 1 \(A01) of a 384-well deep-well plate that already has 18 µl water dispensed into quadrants 2 \(A02), 3 \(B01), and 4 \(B02). Label plate appropriately. iii) Make a 10 fold dilution series using one set of 20 µl Liquidator tips, or using a multichannel pipet. Transfer 2 µl of each sample from

quadrant 1 \(A01) to quadrant 2 \(A02). Mix by pipetting up and down 10x. **CRITICAL STEP:** Do not go to the second stop on the Liquidator trigger, as this will introduce bubbles. iv) Using the same tips, transfer 2 µl of well-mixed sample from quadrant 2 \(A02) to quadrant 3 \(B01). Mix samples in quadrant 3 by pipetting up and down 10x. v) Finally, transfer well-mixed samples from quadrant 3 \(B01) to quadrant 4 \(B02) to complete the dilution series. vi) Using the same set of tips, change the volume on the Liquidator to 3 µl, and transfer 3 µl of each dilution to a barcoded 384-well PCR plate, starting with the most dilute sample in quadrant 4 \(B02) and ending with the most concentrated sample in quadrant 1 \(A01). Label plate appropriately. vii) Cover plate with foil seal, vortex to mix and spin-down on centrifuge. **PAUSE POINT:** Store plates at -20°C or continue on to primary PCR. **1A.II) Primary PCR** i) Thaw the KAPA HiFi PCR kit reagents as well as the sample dilution 384-well plate if it has been stored in the freezer. Remember to vortex and centrifuge the plate\(s) and reagents \(when thawed) before using them. ii) Make a 2x KAPA HiFi qPCR master mix using the following recipe \(see Table 1 for list of marker-gene primer sequences): 1.2 µl 5x KAPA HiFi buffer 0.18 µl 10 mM dNTPs 0.3 µl DMSO 0.12 µl ROX \(25 µM) 0.003 µl 1000x SYBR Green 0.12 µl KAPA HiFi Polymerase 0.3 µl forward primer \(10 µM) 0.3 µl reverse primer \(10 µM). 0.48 µl Nuclease-free water **TROUBLESHOOTING:** This recipe uses the KAPA HiFi Polymerase enzyme at 1x the manufacturer's recommendation. For samples with low template abundance, a lower enzyme concentration such as 0.5x may be desirable, as we have shown that decreasing the enzyme concentration can improve sample balance and reduce adapter dimer contamination. iii) Using a Matrix or multichannel pipette, dispense 3 µl of 2x KAPA HiFi qPCR master mix in each reaction well on the 384-well plate containing sample for a final reaction volume of 6 µl. iv) Seal the plate with an optical seal, vortex at 1900 rpm for 15 seconds to mix, and spin-down in centrifuge. v) Run the following qPCR protocol on the ABI 7900. Typically, this should be run for 20 cycles and stopped at the end of the final extension phase. Cycling Conditions: 95°C – 5 minutes 15-30 cycles of: 98°C – 20 seconds 55°C – 15 seconds 72°C – 1 minute Hold at 4°C **TROUBLESHOOTING:** Samples that amplify poorly can be amplified with increased cycle numbers. It is recommended that a water blank be run to allow assessment of potential contamination in reagents. **PAUSE POINT:** Store plate at -20°C or continue on to analysis and cherry-picking and primary PCR dilution. **1A.III) Analysis and cherry-picking and primary PCR dilution** This analysis and cherry-picking protocol has been protocolized on an EpMotion 5075 Robot. Cherry-picking of optimally amplified samples could also be done using other systems or even manually, though using a robot is strongly recommended to avoid error. i) Export the "Clipped data" .txt file from the ABI 7900, open it with Microsoft Excel, and paste the contents of this file into the "paste clipped data" tab of the DI_qPCR_Analysis.xlsx template \(see Supplemental Files). ii) Examine the amplification curves for each sample in the "96-well summary" tab of the DI_qPCR_Analysis.xlsx spreadsheet. If the samples have amplified consistently, enter the number of amplification cycles and a target Rn value in the "Set thresholds" tab. This target Rn value should be the Rn at which most or all samples have at least one dilution that is in the mid-to-late exponential phase of the PCR at the final amplification cycle. Try to minimize the number of plateaued \(i.e. overamplified) samples that you are selecting. Such samples can be excluded automatically by entering "TRUE" in the "Exclude overamplified samples?" box on the "Set thresholds" tab. **CRITICAL STEP:** Double-check the cherry-picking choices in the "Calculations" tab \(to make sure that the sub-optimal choices are not being made, for instance, choosing a non-amplified samples when all of the amplified samples have plateaued). If necessary, cherry-picking choices can be manually edited by changing the entry in column "AA" in the "Calculations" tab to select the desired dilution. iii) Open the "EpMotion program" tab and save the data in this tab as a .csv file. iv) Centrifuge the 384-well qPCR plate to collect sample and remove the optical seal

from the plate. v) Using the EpMotion, transfer 2 µl of sample of the appropriate dilution from the 384-well qPCR plate to a 96-well plate containing 18 µl of water, making a 1:10 dilution of the initial PCR. Label this plate appropriately. vi) Cover 1:10 dilution plate with foil seal, vortex at 1900 RPM for 30 seconds and spin down in centrifuge. vii) Transfer 5 µl of the sample to a 96-well plate containing 45 µl of water to make a 1:100 dilution. Label this plate appropriately. **PAUSE POINT:** Store plates at -20°C or continue on to Indexing PCR. **1.B.I) Creating sample plate** i) If desired, normalize samples to a consistent concentration in a 96-well plate. Label plate appropriately. ii) Transfer 3 ul of sample from each well to a 96-well run plate. Seal the plate with a foil seal. Label plate appropriately. iii) Vortex samples to mix at 1900RPM for 15 seconds and spin-down in a centrifuge. **PAUSE POINT:** Store plates at -20°C or continue on to primary PCR. **1.B.II) Primary PCR** i) Thaw the KAPA HiFi PCR kit reagents as well as 96-well plate if it has been stored in the freezer. Remember to vortex and centrifuge the plate\(s) and reagents \(when thawed) before using them. ii) Make a 2x KAPA HiFi PCR master mix using the following recipe \(see Table 1 for list of marker-gene primer sequences): 1.2 µl 5x KAPA HiFi buffer 0.18 µl 10 mM dNTPs 0.3 µl DMSO 0.12 µl KAPA HiFi Polymerase 0.3 µl forward primer \(10 µM) 0.3 µl reverse primer \(10 µM). 0.6 µl Nuclease-free water **TROUBLESHOOTING:** This recipe uses the KAPA HiFi Polymerase enzyme at 1x the manufacturer's recommendation. For samples with low template abundance, a lower enzyme concentration such as 0.5x may be desirable, as we have shown that decreasing the enzyme concentration can improve sample balance and reduce adapter dimer contamination. iii) Using a Matrix or multichannel pipette, dispense 3 µl of 2x KAPA HiFi PCR master mix in each reaction well on the 96-well plate\(s) containing sample for a final reaction volume of 6 µl. iv) Seal the plate\(s) with Microseal PCR film, vortex at 1900 rpm for 15 seconds to mix, and spin-down in centrifuge. v) Run the following PCR protocol on the BioRad Tetrad \(or equivalent PCR machine). Typically, this should be run for 20 cycles and stopped at the end of the final extension phase. Cycling Conditions: 95°C – 5 minutes 15-30 cycles of: 98°C – 20 seconds 55°C – 15 seconds 72°C – 1 minute Hold at 4°C **TROUBLESHOOTING:** Samples that amplify poorly can be amplified with increased cycle numbers. It is recommended that a water blank be run to allow assessment of potential contamination in reagents. **PAUSE POINT:** Store plate at -20°C or continue on to Primary PCR dilution. **1.B.III) Primary PCR dilution** i) Transfer 2 µl of sample of each well of the 96-well Primary PCR plate to a 96-well plate containing 18 µl of water, making a 1:10 dilution of the initial PCR. Label this plate appropriately. ii) Cover 1:10 dilution plate with foil seal, vortex at 1900 RPM for 30 seconds and spin down in centrifuge. iii) Transfer 5 µl of the sample to a 96-well plate containing 45 µl of water to make a 1:100 dilution. Label this plate appropriately. **PAUSE POINT:** Store plates at -20°C or continue on to Indexing PCR. **2.I) Indexing PCR: Picking an indexing scheme** i) Determine an i5 and i7 dual-indexing scheme, ensuring no index overlap between samples to be pooled for sequencing \(see Appendix 1 on making indexing primer plates). The final scheme used will be dependent on how many samples are to be pooled together for sequencing. For low numbers of samples, follow the Illumina's guidelines for low-plex sample pooling \
(http://www.illumina.com/documents/products/technotes/technote_nextera_low_plex_pooling_guidelines.pdf) See Table 2 and attached MiSeq setup sheets for index primer sequences. **CRITICAL STEP:** Be absolutely certain that no two samples have the same index combination. ii) Copy and paste the sample names, sample plate number, into the appropriate MiSeq setup template \(see Supplemental Files – MiSeq setup templates). Save this file as a .csv file. **2.II) Indexing PCR** i) Thaw the KAPA HiFi PCR kit reagents as well as the Primary PCR 1:100 dilution plate if it has been stored in the freezer. Remember to vortex and centrifuge the plate\(s) and reagents \(when thawed) before using them. ii) Make sure there is a corresponding 5 µM oligo

plate for each index scheme being used \(see Appendix 1). iii) Make a 3.33x KAPA HiFi Indexing PCR master mix using the following recipe: 2 µl 5x KAPA HiFi buffer 0.3 µl 10 mM dNTPs 0.5 µl DMSO 0.2 µl KAPA HiFi Polymerase iv) With a multichannel pipette, dispense 3ul of mix into each well of a 96-well PCR plate. Label this plate appropriately. v) Using the Liquidator, or a multichannel pipet, pull 5 µl from the Primary PCR 1:100 dilution plate and dispense into the corresponding 96-well PCR plate that already contains 3 µl of the Indexing PCR master mix. vi) Using the Liquidator, or a multichannel pipet, pull 2 µl from the appropriate 5 µM indexing scheme plate and dispense into the corresponding 96-well PCR plate that contains the indexing PCR master mix and sample. vii) Optional – If you are preparing more than 96 samples, with the tips still submerged in the sample, change the Liquidator volume knob to 10 µl and draw up the 10 µl reaction dispense \(starting in quadrant 1, A01) into a 384-well plate. Label this plate appropriately. viii) Repeat steps iii-vii using a new set of Liquidator tips if there is more than one plate to set up. ix) Cover the PCR plate\(s) with Microseal PCR film, vortex at 1900 RPM for 30 seconds and spin down in a centrifuge. x) Amplify PCR plate\(s) for 10 cycles on the BioRad Tetrad \(or comparable PCR machine) using the following cycling conditions: 95°C – 5 minutes 10 cycles of: 98°C – 20 seconds 55°C – 15 seconds 72°C – 1 minute Hold at 4°C xi) After the PCR program is complete, centrifuge the plate to collect the sample. **PAUSE POINT:** Store plates at -20°C or continue on to Normalization and Pooling. **3) Normalization and pooling** The normalization and pooling step of this protocol can be carried out using option A or B, depending on availability of equipment and reagents, whether or not size selection is required, and scale \(A - PicoGreen-based normalization - recommended for smaller-scale projects or when doing size selection; B - SequalPrep–based normalization - recommended for larger-scale projects). **TROUBLESHOOTING:** Note that the presence of adapter dimers can throw off mass-based normalization calculations and may lead to sample pool imbalance if different samples have different amounts of adapter dimer contamination \(particularly if libraries are size selected). **3.A.I) PicoGreen quantification of indexed samples** i) Determine the individual sample concentrations by PicoGreen \(following manufacturer's protocol). **3.A.II) Normalization calculations** i) Determine a desired concentration for normalization. This concentration can be adjusted depending on how concentrated or dilute the samples are. ii) For each sample, calculate the amount of water that needs to be added to a fixed volume of indexing PCR sample \(typically 2 µl, 4 µl, or 6 µl) in order to get the desired concentration. **3.A.III) Sample Normalization** i) Use the EpMotion to transfer the appropriate amount of water to each well of a new 96-well plate. Label this plate appropriately. ii) Use the Liquidator or a multichannel pipette to add the previously determined fixed amount of indexing PCR sample to each well on the normalization plate. iii) Cover plate with foil, vortex to mix and spin-down in the centrifuge. **TROUBLESHOOTING:** We have observed some systematic differences in the efficiency of amplification and ultimate sample balance associated with specific indexing primers. At this point it is unknown whether this is due to intrinsic sequence-specific properties of the indexing primers or due to the specific synthesis. However, whatever the cause, we have found that sample balance can be improved by adding empirically determined index-specific correction factors to the normalization calculations. **PAUSE POINT:** Store plates at -20°C or continue on to Pooling and clean-up. **3.A.IV) Pooling and clean-up** i) Dispense an equal volume of each sample to be pooled into a trough. Mix well and transfer from the trough to a 1.5 ml microfuge tube. **CRITICAL STEP:** If doing the optional size selection step, make sure not to pool more than 1 µg \(Which is the maximum input for the Caliper cut). The amount of each sample to pool depends on the number of samples and the normalized sample concentration. ii) Use the Speedvac to concentrate the sample pool down to 20-100 µl and purify the sample pool using 1X AmPureXP beads \(Appendix 2). If you will be size selecting the material on the Caliper XT, elute in 12.5 µl of

nuclease-free water. If the sample will not be size selected, or if you have pooled >1.5 µg of sample, elute in 25 µl of nuclease-free water or EB \(10 mM Tris-Cl, 1 mM EDTA, pH 8.0). **PAUSE POINT:** Store samples at -20°C or continue to Library QC or optionally size selection using the Caliper XT. **3.B.I) SequalPrep plate-based normalization** i) Purify the 10 µl indexing PCR reactions using a SequalPrep normalization plate, following the manufacturer's instructions. **CRITICAL STEP:** For new sample types, make sure that there is sufficient input material post-indexing PCR to use the SequalPrep plates. This can be assessed using the PicoGreen assay. If samples are expected to be consistently amplified, it may suffice to spot-check several samples. **3.B.II) Pooling and clean-up** i) Dispense 10 µl of each sample to be pooled into a trough. Mix well and transfer from the trough to a 1.5 ml microfuge tube. For projects with large numbers of samples, the volume of pooled material can be adjusted downward. **TROUBLESHOOTING:** We have typically observed yields that are ~5-10 fold lower than those listed in the SequalPrep manual. Keep this in mind when pooling to ensure that you have enough material in the final library to sequence. The eluted material, which is typically in the range of 0.1-0.2 ng/ul, will need to be concentrated 5-10x in order to get the library into the 1-2 nM range required for Illumina sequencing. ii) Use the Speedvac to concentrate the sample pool down to 20-100 µl and purify the sample pool using 1X AmPureXP beads \(Appendix 2). Elute in 25 µl of nuclease-free water or EB \(10 mM Tris-Cl, 1 mM EDTA, pH 8.0). **PAUSE POINT:** Store samples at -20°C or continue to Library QC. **OPTIONAL: Size selection using the Caliper XT** i) Remove the Caliper LabChip XT DNA 750 kit reagents from the 4°C and the AmPureXP beads and let them come to room temperature. The Caliper reagents are light sensitive- so keep them in the dark. ii) Load 10 µl of the concentrated pool on a single Caliper LabChip DNA 750 chip lane and cut depending on the region being amplified, following the manufacturer's protocol. The sizes used for size selecting different variable regions are as follows: 16S V1-V3 = 662 bp +/- 20% 16S V3-V4 = 550 bp +/- 20% 16S V3-V5 = 722 bp +/- 20% 16S V4 = 427 bp +/- 20% 16S V4-V6 = 685 bp +/- 20% 16S V5-V6 = 416 bp +/- 20% 18S V9 = 260 +/- 20% ITS1, ITS2: Size selection is not recommended for these regions, as they have highly variable size distributions **CRITICAL STEP:** Do not overload the chip \(> 1 µg), as this can affect migration of the DNA and the accuracy of sizing. ii) Purify sample with 1.8X AmPureXP Beads and elute in 20 µl of water. **PAUSE POINT:** Store samples at -20°C or continue to library QC. **4) Library QC** i) Perform library QC by running pooled library on an Agilent DNA HS chip, following the manufacturer's protocol and verifying that the size distribution is in the expected range \(see Anticipated Results section for examples and troubleshooting information). ii) Determine the concentration of the pooled library by PicoGreen \(following manufacturer's protocol). iii) Dilute the pooled library to 2 nM in EB using the concentration determined by the PicoGreen assay, and the expected size for the amplicon in the case of the 16S or 18S rRNA genes \(see above), or the average size determined by the Agilent DNA HS chip in the case of ITS1 and ITS2. **5) Sequencing** i) Thaw MiSeq reagent kit \(typically either a 500 cycle v2 or 600 cycle v3 kit). ii) Generate a MiSeq sample sheet using the Illumina Experiment Manager program and paste in the sample and barcode information. Load the sample sheet on the MiSeq. iii) Follow Illumina's protocol for preparing the instrument and flow cell. iv) In a 1.5 ml microfuge tube, denature 10 µl of 2 nM library by incubating with freshly diluted 0.2 N NaOH for 5 minutes at room temperature. v) Add 980 µl of Illumina's HT1 buffer to bring the sample to 20 pM. vi) Dilute to 8 pM by mixing 400 µl of 20 pM library and 600 µl of Illumina's HT1 buffer in a clean 1.5 ml microfuge tube. vii) Prepare 8 pM PhiX. viii) Remove 150 µl of the 8 pM library and discard. Add 150 µl of 8 pM PhiX to the remaining 850 µl of 8 pM library \(15% PhiX spike). ix) Add 600 µl of the 8 pM \(15% PhiX) library to the sample well of the MiSeq cartridge and initiate sequencing. Libraries generated with this protocol are sequenced using standard Illumina sequencing primers \(which are already present in the MiSeq reagent

cartridge). **TROUBLESHOOTING:** Sequence quality will be dramatically affected by cluster density and amount of PhiX spike. Accurate clustering will depend on the accuracy of library quantification. The amount of library clustered and may need to be varied to account for differences in library quantification methodology or technique between groups. In addition, for longer amplicons such as V1-V3, V3-V5, and V4-V6, it is desirable to spike in 25% PhiX, rather than 15% \(see Anticipated Results section). **APPENDIX 1 – Making 5 µM indexing scheme primer plates** **CRITICAL STEP:** In order to avoid potential contamination, we recommend making up batches of indexing scheme plates designed to be used a small number of times \(≤5). **1) Making a 5 µM master plate** i) Starting with the 100 µM stock plates \(see Table 2 for primer sequences and plate layouts), first make 10 µM daughter plates containing the desired forward and reverse indexing schemes. Note: forward indexing primers are arrayed in columns \(and given numerical prefixes as names) and reverse indexing primers are arrayed in rows \(and given letter prefixes as names). ii) In a deep-well 96-well plate, for the forward indexing scheme, add 270 µl of nuclease-free water to each well of the column\(s) corresponding to the indices that you will be using in the 100 µM stock plate. Next transfer 30 µl of each forward index to the column\(s) containing water. iii) In a deep-well 96-well plate, for the reverse indexing scheme, add 180 µl of nuclease-free water to each well of the row\(s) corresponding to the indices that you will be using in the 100 µM stock plate. Next transfer 20 µl of each forward index to the row\(s) containing water. iv) Using a multichannel pipet, add 20 ul of the 10 µM forward primers column-wise to a new 96-well plate. Label the plate appropriately \(i.e. Indexing scheme 1B \(5 µM)). v) Using a multichannel pipet, add 20 ul of the 10 µM reverse primers row-wise to the 96-well plate made in step ii. vi) Using a multichannel pipet or the liquidator, Add 40 µl of water to each well create a 5 µM master plate containing 80 µl of indexing primer mix. vii) Seal the plate with a foil seal, vortex at 1900 rpm for 15 seconds to mix, and spin-down in centrifuge. viii) Repeat for additional indexing schemes if desired **PAUSE POINT:** Store master plates at -20°C or continue to Aliquoting limited use indexing plates. **2) Aliquoting limited use indexing plates** i) Thaw the indexing scheme master plate\(s) if it has been stored in the freezer. Remember to vortex and centrifuge the plate\(s) \(when thawed) before using them. ii) Carefully remove the foil seal from the indexing scheme master plate and use the liquidator or a multichannel pipet to dispense 12 µl of primer mix from each well into each of 6 96-well plates. Each plate will have enough primer mix for 5 indexing reactions, thus each batch makes enough material for 5 x 6 x 96 = 2880 samples per indexing scheme. iii) Store indexing scheme plates at -20°C. **APPENDIX 2 – AmPure XP bead purification** i) Vortex AmPure XP beads until they are well mixed. Calculate the amount of beads needed to clean-up the samples \(typically a ratio of 1.8x beads:sample is used, though 1x or even 0.8x beads:sample ratios can be used in cases where it is desirable to remove material that is <200 bp, such as adapter dimers). ii) Add appropriate volume of well-mixed AmPure XP beads using a multichannel pipette to each well \(if using a plate-based magnet) or tube \(if using a tube-based magnet). Allow the sample and beads to incubate for 5 minutes at room temperature. iii) Place the plate \(or tubes) on the magnetic stand at room temperature for 2 minutes, until the liquid appears clear. Remove and discard the supernatant, being careful to not disturb the magnetic bead pellet. iv) With the magnetic plate still on the stand, wash the bead pellet 2X for 30 seconds with 200-500 µl of fresh 80% ethanol, also while not disturbing the pellet. Make sure that 80% ethanol completely covers the beads during the wash. v) Remove all traces of the 80% ethanol and let the plate/tubes dry at room temperature for 10 minutes. vi) Resuspend the beads in designated amount of nuclease free water and pipette gently to mix. Allow beads/sample to incubate at room temperature for 2 minutes. vii) Place the plate/tubes back on the magnetic stand at room temperature for 2

minutes, until the liquid appears clear. viii) Transfer supernatant \(containing the eluted DNA) to a new tube or plate.

Anticipated Results

1) Primary PCR Anticipated results for the primary PCR \(using the template optimization qPCR protocol) are shown in Figure 2. Example traces for three samples amplified with the 16S rRNA V4 primers are shown in Figure 2A-C. The trace for the water blank that was run in parallel is shown in Figure 2D. The samples that were selected for subsequent cherry-picking and indexing are shown in Figure 2E-H). **2) Indexing PCR** At this stage, we do not routinely check individual samples unless the qPCR data suggested a problem. If desired, indexed amplicons can be checked to see if they are the correct size or if there is adapter dimer contamination by running on an agarose gel or Agilent Bioanalyzer. **3) Normalization and pooling** If doing PicoGreen-based normalization, you should see PicoGreen values that are clearly above those of the water blanks. The values will depend on starting template concentration and the amount of amplification. Most typically we see values between 2 ng/µl and 20 ng/µl. **4) Library QC** For the 16S rRNA amplicons, a high quality library will have a single clear peak of the expected size \(Figure 3A). For ITS1 and ITS2, size distributions are typically more variable and a high quality library will typically have several peaks. Possible problems with library preparation include adapter dimer contamination \(Figure 3B) and \(rarely) off-target amplification of host DNA \(Figure 3C). Both of these problems are typically seen for marginal samples that have very low microbial template abundance. For bacterial samples, these undesired products can be removed by size selection of the library, though this may affect pool balance \(see Normalization and pooling section above). If a large amount of adapter dimer contamination is seen, it may be desirable to optimize the PCR reaction by adjusting the concentration of the KAPA HiFi polymerase \(see Primary PCR section above). **5) Sequencing** Due to the imaging-based nature of Illumina sequencing, the low cycle-to-cycle sequence diversity of amplicon libraries presents special challenges for sequencing[8,9]. Failure to cluster the flow cell at an appropriate cluster density can lead to lower sequence quality or run failure. In addition, spiking in a more diverse library \(typically a PhiX control) improves sequence quality. Ultimately, there is a trade-off between yield \(number of reads) and quality \(Figure 4). We have also found that there is a strong correlation between 16S rRNA variable region and sequence quality. Longer amplicons such as V1-V3 and V4-V6 are systematically associated with lower sequence quality across a range of clustering concentrations and PhiX spike-in concentrations than shorter amplicons such as V4 and V5-V6 \(Figure 4). Note that in order to avoid confounding effects of instrument-to-instrument variation \(which we have observed), all the data shown here was acquired using the same MiSeq instrument.

References

1. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 12, 87 \(2014). 2. Weiss, S. et al. Tracking down the sources of experimental contamination in microbiome studies. Genome Biol. 15, 564 \(2014). 3. Brooks, J. P. et al. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. BMC Microbiol. 15, 66 \(2015). 4. Pinto, A. J. & Raskin, L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS One 7, e43093 \(2012). 5. Sinha, R., Abnet, C. C., White, O., Knight, R. & Huttenhower, C. The

microbiome quality control project: baseline study design and future directions. Genome Biol. 16, 276 \(2015). 6. Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T. J., Clayton, J. B., Johnson, T. J., Hunter, R., Knights, D., Beckman, K. B. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nat. Biotechnol. \(2016). 7. 16S Metagenomic Sequencing Library Preparation. Illumina Tech. Note 15044223 Rev. A 8. Schirmer, M. et al. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res. 43, e37– \(2015). 9. D'Amore, R. et al. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics 17, 55 \(2016).

Acknowledgements

We thank the staff of the University of Minnesota Genomics Center for helpful discussions and technical support. This work was supported by a grant from the Minnesota Partnership for Biotechnology and Medical Genomics \(MNP IF #14.09).

Figures

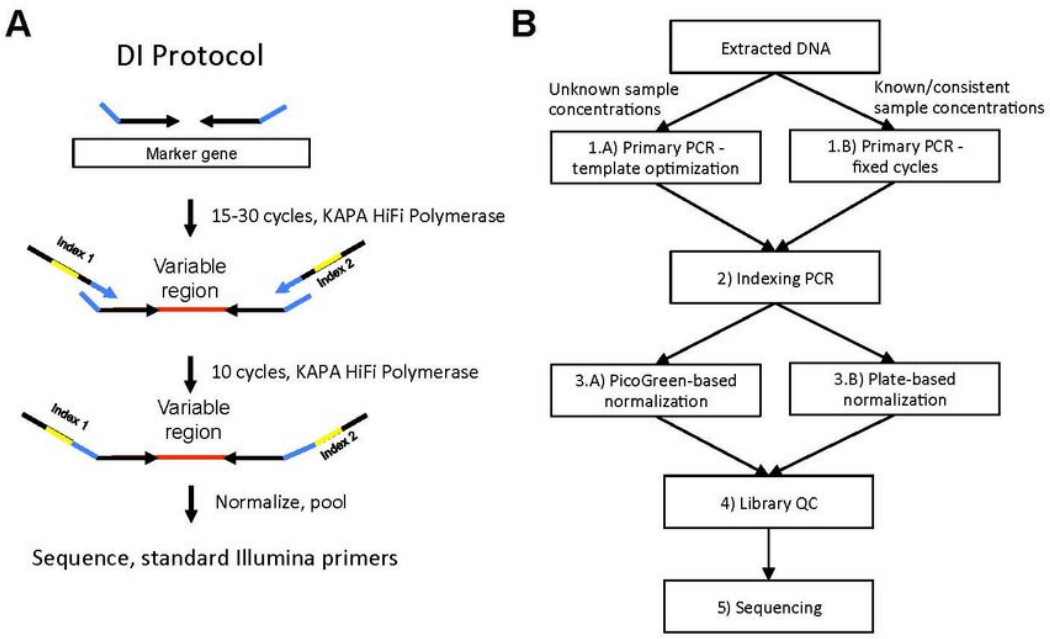

Figure 1

Figure 1

Outline of Dual-Indexing Method A) Dual-indexing protocol, in which the variable region of a marker gene is first amplified with primers containing common adapter sequences. Then, the Illumina flow cell adapters and dual-indices are added in a secondary amplification. B) Flow chart showing the discrete steps of the procedure and branch points where there are multiple options.

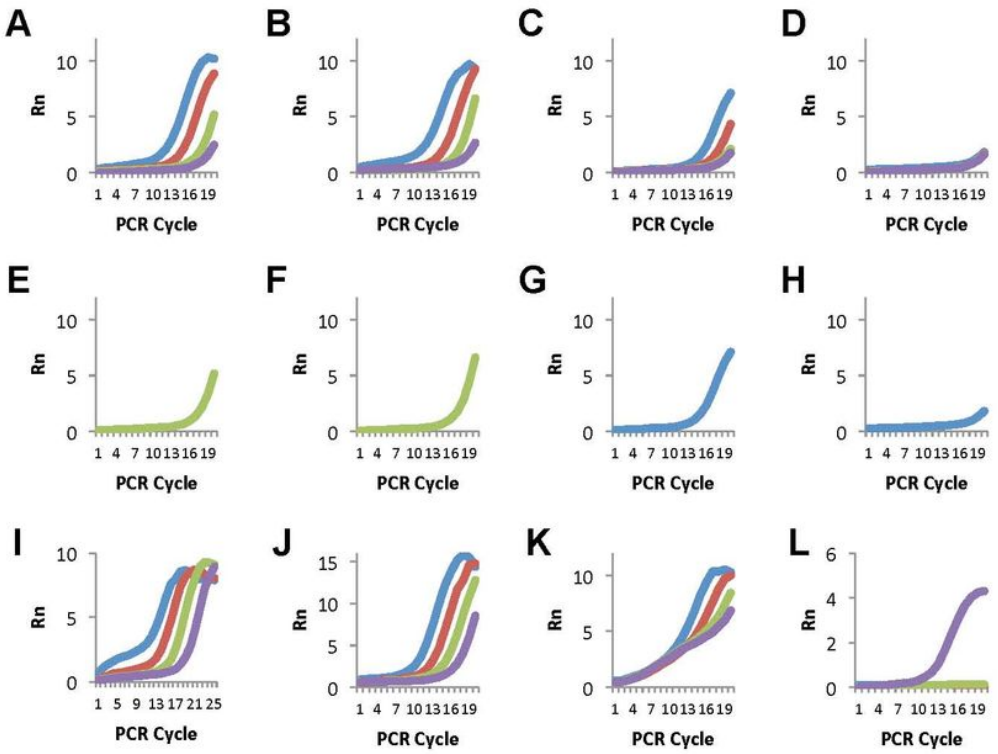

Figure 2

Figure 2

Examples of qPCR amplification traces For all panels, blue trace is the most concentrated sample, red trace is 1:10 dilution, green trace is 1:100 dilution, purple trace is 1:1000 dilution. A-D) Example 16S rRNA V4 qPCR traces for 3 microbial community DNA extracts (A-C) and a water blank control (D). E-F) Samples selected for cherry-picking and indexing PCR from the qPCR traces shown in A-D based on a targeted Rn of 6.3. I) Example qPCR traces for 16S rRNA V5-V6. J) Example qPCR traces for 16S rRNA V4-V6. K) Example qPCR traces for 16S rRNA V1-V3. L) Example qPCR trace for a 16S rRNA V5-V6 sample with evidence of PCR inhibitors. In this case, amplification is only seen in the most dilute sample.

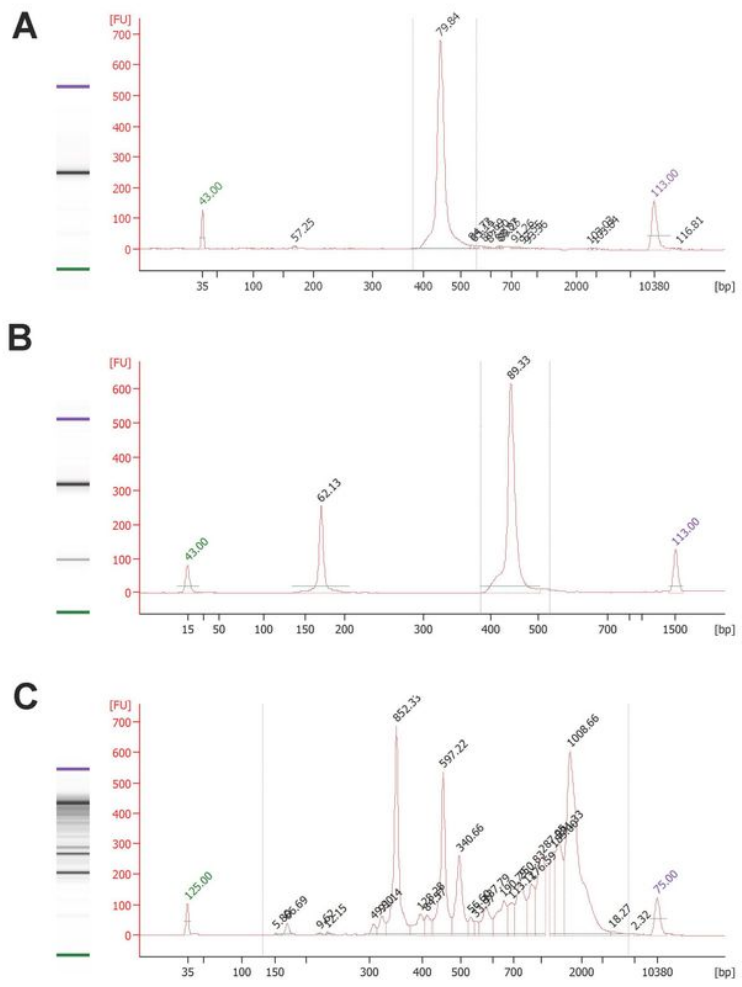

Figure 3

Figure 3

Example Agilent Bioanalyzer profiles of high quality and aberrant libraries A) Example high quality 16S rRNA V4 library. A single clear peak of the expected size (427 bp) is seen. B) Example 16S rRNA V4 library with significant adapter dimer contamination. In this case, further optimization of the PCR, or a gel cut or Caliper XT cut would be recommended. C) Example 16S rRNA V4 library from a sample with an extremely low bacterial DNA and high abundance of host DNA. In addition to the expected peak at 427 bp, additional off-target amplification is observed.

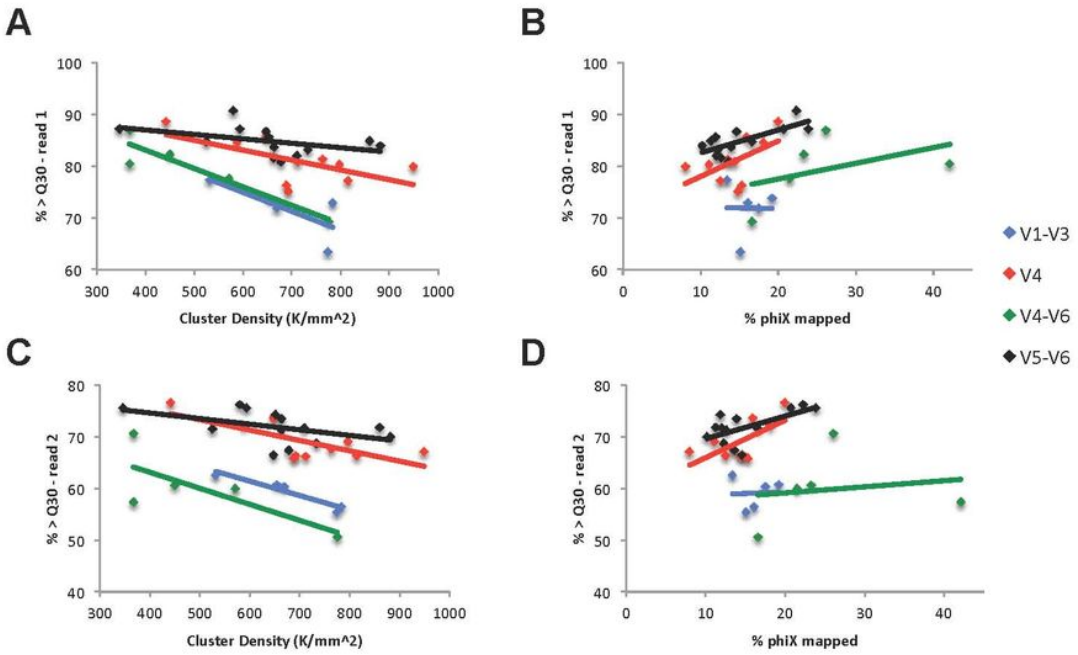

Figure 4

Figure 4

The effect of 16S variable region, cluster density, and PhiX spike-in on sequence quality Data in this figure was generated on the same MiSeq instrument using 600 cycle v3 sequencing reagents. A) Comparison of cluster density and quality score of read 1 for 4 different 16S rRNA variable region amplicon library types. B) Comparison of percent PhiX and quality score of read 1 for 4 different 16S rRNA variable region amplicon library types. C) Comparison of cluster density and quality score of read 2 for 4 different 16S rRNA variable region amplicon library types. D) Comparison of percent PhiX and quality score of read 2 for 4 different 16S rRNA variable region amplicon library types.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement0.xlsx
- supplement0.zip
- supplement0.xlsx