Apr 05, 2019

Working

Fuhrman Lab 515F-926R 16S and 18S rRNA Gene Sequencing Protocol

Version 2

David Needham[1], Erin Fichot[1], Alma Parada[1], Yi-Chun Yeh[1], Jed Fuhrman[1]

[1]University of Southern California

dx.doi.org/10.17504/protocols.io.vb7e2rn

Fuhrman Lab

Jed Fuhrman
University of Southern California

ABSTRACT

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Alma E Parada, David M Needham, Jed A Fuhrman (2015) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental microbiology 18 (5):1403-1414 https://doi.org/10.1111/1462-2920.13023

Needham, D. M., Fichot, E. B., Wang, E., Berdjeb, L., Cram, J. A., Fichot, C. G., & Fuhrman, J. A. (2018). Dynamics and interactions of highly resolved marine plankton via automated high-frequency sampling. The ISME Journal. https://doi.org/10.1038/s41396-018-0169-y

working_protocol_Tag_Sequencing_protocol.91216 (2).docx

PROTOCOL STATUS

Working

We use this protocol in our group and it is working

MATERIALS

NAME	CATALOG #	VENDOR
Quant-iT dsDNA Pico Green assay kit (Invitrogen)	P7589	Life Technologies
SPRIselect reagent kit	B23317	Beckman Coulter
Qubit dsDNA HS Assay kit	Q32854	Thermo Fisher Scientific
Agencourt AMPure XP beads		
Agilent High Sensitivity DNA Kit	5067-4626	Agilent Technologies

SAFETY WARNINGS

Primers

1 We have obtained primers from Operon/Eurofins (https://www.eurofinsgenomics.com/en/home/), 50 nmol scale, salt-free purified.

- Reconstitute main stocks in TE, 100 µM, store at -80° C
- Working stocks: dilute in TE, 10 µM, store at -20° C (can be stored at 4° C for several days to avoid freeze/thaw)

Suggestion: prepare primer plates containing 1:1 mix of unique combinations of 10 µM 515F:926R barcoded primers ([final] of each is 5 µM). Cover plate with sterile foil (USA Scientific, 2923-0110)

515F[2]

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCAGCGTGYCAGCMGCCGCGGTAA

Illumina adapter | Illumina seq primer | NNNN | bc | 515F primer (4Y)

926R[3]

CAAGCAGAAGACGGCATACGAGATATCACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGYCAATTYMTTTRAGTTT

Illumina adapter | index | Illumina seq primer | 926R primer

Barcodes and Indices

	5'-3' for. barcode			5'-3' rev. index
515F (4Y) 1	TCAGC		926R 1	ATCACG
515F (4Y) 2	GTATC		926R 2	CGATGT
515F (4Y) 3	GCTAC		926R 3	TTAGGC
515F (4Y) 4	ACGCA		926R 4	TGACCA
515F (4Y) 5	CGCGT		926R 5	ACAGTG
515F (4Y) 6	CTGGT		926R 6	GCCAAT
515F (4Y) 7	GCGTT		926R 7	CAGATC
515F (4Y) 8	GGAAC		926R 8	ACTTGA
515F (4Y) 9	AAGCC		926R 9	ATGTCA
515F (4Y) 10	AGCTT		926R 10	CCGTCC
515F (4Y) 11	CCCTT		926R 11	GTCCGC
515F (4Y) 12	CGCAC		926R 12	GTGAAA
515F (4Y) 13	GGTGT		926R 13	CACCGG
515F (4Y) 14	GGCAG		926R 14	CACGAT
515F (4Y) 15	TGATA		926R 15	CACTCA
515F (4Y) 16	TGTGC		926R 16	CAGGCG

6 base error-correcting barcodes as used to date; 8 base indices might be better for future work as they can yield more combinations. Additionally, using completely unique combinations of forward and reverses (no overlap in either) might reduce the amount of cross-talk between samples.

These 16 forward and reverse primers allow for 256 unique combinations for multiplexing, more are possible.

PCR

2 5 PRIME HotMasterMix

- 100 Rxns: 2200400 (VWR 10052-240) vs 1000 rxns: 2200410 (VWR 10052-242)

	[stock]	vol per rxn (µL)	[final]/rxn
PCR water*		12.5	
5' Master Mix	2.5 x	10****	0.5 U taq, 45 mM Kcl, 2.5 mM Mg2+, 200 uM each dNTP
1:1, 515F:926R primer mix**	5 µM each primer	1.5	0.3 mM each primer
DNA***	0.5 ng/µL	1	0.5 ng
total vol		25	

*VWR cat# 95043-414

**Equivalent to adding 0.75 µL of each 10 µM working primer stock; We have tested/used 0.2-0.4 mM successfully

***Can be modified according to [DNA] or DNA quality as long as you use the same amount (ng) in each reaction for a particular study. We have tested/used 100 pg – 2 ng.

****For trouble samples, increase master mix to 12 uL, and this seems to improve amplification (0.6 U taq, 54 mM KCl, 3 mM Mg2+, 240 uM each dNTP).

Note, we initially used more expensive Taq such as HiFi from NEB, but saw no appreciable difference in terms of sequence abundances in our mock communities and samples.

3 Use UV-Crosslinker to treat consumables. Press "Time", then "10.0" to run for 10 minutes

- PCR strip tubes (cat#)
- PCR water
- Any tubes needed to make a master mix

4 Wipe down inside of the PCR hood and all pipettes with 10% bleach. Turn the UV light on (15 min).

5 Make a master mix of PCR water and 5' Master Mix (5% extra to account for pipetting error)
- Flick-mix, spin tube
- Pipet 22.5 µL into each PCR strip tube

6 Add 1.5 µL of primer mix to each tube
- For primer plate, use a Kimwipe to wipe 10% bleach on aluminum foil cover
- Then use Kimwipe to wipe 70% EtOH on aluminum foil cover
- Wait for EtOH to evaporate before pipetting through aluminum foil.
- When done, place new foil directly on top of punctured foil.

7 Add 1 µL of template DNA to appropriate tubes
- Do not add mock community template in the clean hood, since these are pooled PCR products and could conceivably cause contamination of the hood. Instead, wait until all other sample tubes are closed, take everything out of the hood, add mock community template to appropriate tubes, then proceed
- Amplify ≥ 1 even and ≥ 1 mock community per sequencing run

8 Flick-mix tubes, then centrifuge briefly.

9 Place tubes in thermocycler:

Initial Denaturation:	95°C 120s	**Our default PCR setup is 30 cycles with 0.5ng template
30 Cycles** of:	95°C 45s 50°C 45s 68°C 90s	
Final Elongation Step:	68°C 300s	
Refrigeration:	4°C forever	

Assess Amplification

10 <u>Preparing gel</u>
1.0 g agarose
100 mL TAE buffer (or TBE)

Swirl to mix agarose and buffer

11 Microwave 1 minute, swirl. If solution is not clear, microwave in 10 sec increments and swirl. **caution, hot**

12 Add **7 µL SYBR-safe dye**, vortex.

13 Prepare casting tray, making sure it is level and that orange gaskets are arranged properly

14 Pour gel into tray

15 Add comb(s)

16 Preparing Samples

Cut strip of parafilm.

17 Pipet **2 µL loading dye** onto parafilm, one dot per sample

18 Pipet **3 µL of sample** into loading dye, pipet up and down 2 times.

- If replicate PCRs were performed, can pool before loading onto gel

19 Loading and running gel

Place solidified gel in proper orientation (samples running down toward you)

20 Pour used TAE running buffer (or TBE, if that was used to make gel) into chamber, fill until it just covers gel

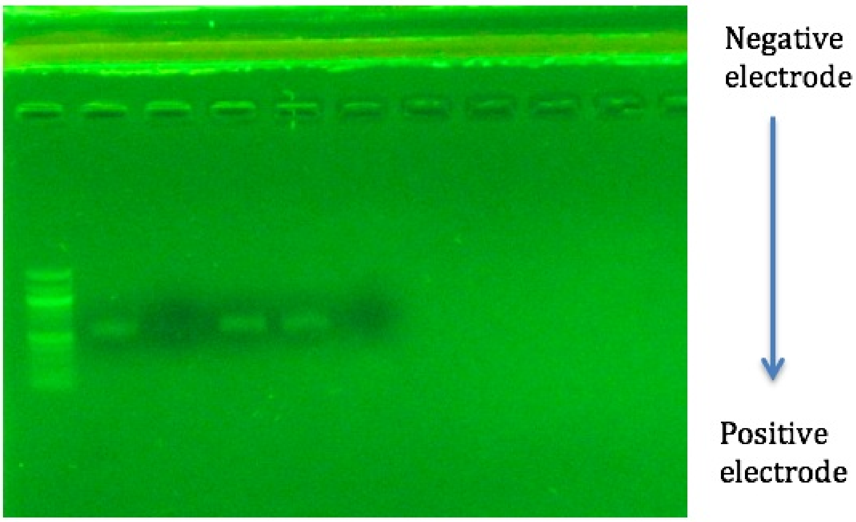
21 Remove comb

22 Pipet **5 µL of 100 bp ladder** into first well (left)

23 Pipet 5 µL of dye/sample into each remaining well

24 Plug in electrodes to power source

25 Run for **~100 V, 30-45 minutes**.



26 What to look for

- Amplification ~563 bp
- May see 2 bands if eukaryotic sequences are present (band will be about 200 bp longer)

- No-template-control PCR blanks, no amplification
- Primer dimer may be present, but should be removed upon clean-up (below)

Clean-Up: Agencourt AMPure XP PCR Purification

27 If cleaning only a few samples at a time, use the following protocol (Ampure Beads). Otherwise, clean PCR reactions with SequalPrep Normalization Plate (Invitrogen A10510-01).

28 Before beginning:

Warm aliquot of beads at RT for 30 minutes.

29 Spin down the PCR rxns tubes using the microcentrifuge (brief spin)

30 Bring up volume of remaining PCR product to 40 µl with TE

31 Should have 22 µL PCR product remaining, so add 18 µL TE

32 Label another set of PCR tubes for collection of cleaned amplicons

33 Label another set of PCR tubes (or a plate) for 1:5 dilution of cleaned amplicons (to be used for quantification by Pico Green assay, see below)

34 To this set of tubes, add 4 µl of TE

35 Once beads are warm, add the beads to 0.8 x ratio in PCR tubes

- 32 µl beads for 40 µl sample

36 Vortex strip tubes for several seconds

37 Allow the beads and DNA to bind by waiting 5 minutes.

38 After 5 minutes, place the tubes in magnetic separator.

39 Wait 3 minutes for full separation

40 Remove and discard the clear buffer

41 Add 200 µL of *freshly-made* 80% ethanol. (Ethanol is hygroscopic such that it will pull water out of the air and be diluted. While is this probably more of a consideration for 100% Ethanol, it is good lab practice to remake fresh 80% Ethanol.)

42 Vortex the tube for 5-6 seconds. Not all of the pellet will be dislodged, but that’s ok. Let the ethanol and beads incubate for about 3 minutes.

43 Again, place the 500 µL tube on the magnet for separation.

44 Remove the ethanol and keep tubes on magnet.

45 Repeat the addition of 200 µl of Ethanol.

46 Remove the ethanol (no incubation necessary).

47 Remove the tube and spin for several seconds to collect the remaining ethanol in the bottom of the tube. Put back on magnetic plate

48 Remove the remainder of the ethanol

49 Allow to dry on magnet with open tube for ~5 minutes. The beads should not crack, you will lose DNA, supposedly.

50 Add 10 uL of TE Buffer to the beads and pipette several times making sure to break up all of the bead pellet. The TE buffer elutes the DNA from the beads

51 Incubate for about 5 minutes, separate on the magnet.

52 Collect your DNA by pipetting off about 9.5 µL of the TE (being careful not to remove any of the beads, though, it is apparently ok if you get a tiny amount, but not ideal)

53 Add 1 µL of this to the 4 µL of TE set aside for the 1:5 dilutions to be used for quantification

- No need to dilute blanks
- Store cleaned PCR products at 4°C if using within a week. Otherwise, freeze at -20°C.

Quantification with Pico-Green dsDNA Quant-iT Assay Kit

54
- Samples need to be in the range of the standard curve (0-15 ng/µL), otherwise the quantification is not accurate
- In general, well-amplified PCR samples are around 10-40 ng/uL, so it is best to run 1:5 dilutions of cleaned PCR products to get them within the range of the standard curve
- 1:5 dil of 10 ng/µL = 2 ng/µL
-
- 1:5 dil of 40 ng/µL = 8 ng/µL
-
- Probably best not to dilute no template controls (blanks). One might get a better idea of if there might be contamination present. Note, no template controls usually do have measurable amounts of DNA, but we think this is likely from primer dimer or another source. See further details below regarding blanks.

Quantify

Invitrogen P7589 Kit Components

55
- 20X TE buffer
- pico-green dye
- DNA lambda standard (100 ng/µL)

56 Before beginning

Walk to the ice machine with ice bucket, retrieve ice.

57 Locate reagents in fridge and place on ice. Keep Pico Green fluorescent reagent in the dark.

58 Aseptically remove 96-well plate from bag and place optically-clear strip caps over all wells

59 Cut plate so that only the number of wells that are needed needed, are used (keeping other capped wells for later).

60 Procedure

Turn on Stratagene ≥ 20 min before running

1. Open software
2. Select "quantification plate"
3. Make sure lamp is warming up (light bulb icon will be yellow or green, not red)

61 Dilute TE from kit according to the number of samples to be quantified.

	# samples →	10	20	30	40	50	60	70	80	90	100
	water (µL)	760	2x522.5	2x665	2x807.5	2x950	3x760	3x823.3	3x918.3	4x760	4x831.2
TE	20X stock (µL)	40	55	70	85	100	120	130	145	160	175
	total volume (mL)	**800**	**1100**	**1400**	**1700**	**2000**	**2400**	**2600**	**2900**	**3200**	**3500**
	tube size	2 mL	2 mL	2 mL	2 mL	2 mL	5 mL	5 mL	5 mL	5 mL	5 mL

62 Dilute pico-green dye from kit according to the number of samples to be quantified

- Keep in the dark until needed

	# samples →	10	20	30	40	50	60	70	80	90	100
	1x TE (µL)	398	547.25	696.5	845.75	995	2x597	2x646.8	2x721.4	2x796	2x870.6
pico	200X stock µL)	2	2.75	3.5	4.25	5	6	6.5	7.25	8	8.75
	total volume (µL)	**400**	**550**	**700**	**850**	**1000**	**1200**	**1300**	**1450**	**1600**	**1750**
	tube size	2 mL	2 mL	2 mL	2 mL	2 mL	2 mL	2 mL	2 mL	2 mL	2 mL

63 Dilute DNA stock

- Do serial dilutions, briefly vortexing and spinning down after each dilution
- Make in 0.5 mL LoBind tubes
- Keep on ice

	10 ng/µl	1 ng/µl	0.1 ng/µl

1x TE (µL)	18	36	18
volume of stock (µL)	2	4	2
total volume (µL)	**20**	**40**	**20**
stock (ng/µL)	100	10	1

64 Make standard curve

- Do in duplicate (columns 1+2 on plate)
- Keep on ice until ready to run

96-well row # →	H	G	F	E	D	C	B	A
[final]	0	0.1	0.5	1	5	7.5	10	15
1x TE (µL)	15	14	10	14	10	7.5	14	13.5
volume of stock (µL)	0	1	5	1	5	7.5	1	1.5
total volume (µL)	15	15	15	15	15	15	15	15
stock [] ng/µL	---	0.1	0.1	1	1	1	10	10

65 Load 14 µL of 1x TE to each well to be used for sample measurement

66 Load 1 L of sample to appropriate well

- Write down which wells contain what samples

Ex:

H	**G**	**F**	**E**	**D**	**C**	**B**	**A**	
std-0	std-0.1	std-0.5	std-1	std-5	std-7.5	std-10	std-15	**1**
std-0	std-0.1	std-0.5	std-1	std-5	std-7.5	std-10	std-15	**2**
sample 1	sample 2	sample 3	sample 4	sample 5	sample 6	sample 7	sample 8	**3**
sample 9	sample 10	sample 11	sample 12	sample 13	sample 14	sample 15	sample 16	**4**
								5
								6
								7
								8
								9
								10
								11
								12

67 Add 15 µL of diluted pico-green dye to each well

68 Briefly vortex and spin, and keep in the dark for ≥ 5 minutes

69 Read fluorescence on stratagene (typical SYBR-like fluorescence)

1. Mark wells used for standards and type in actual concentration, click on SYBR
2. Mark wells used for samples as "unknown"s, click on SYBR
3. Check Gain settings—typically use 4 but can use 8 if you need more light
4. "Run"

- Prompted to save file

70 What to look for after completion:

- Std curve
- Duplicates should be close to each other (can remove if necessary)
- R^2 should be ≥ 0.95
- Should be linear (can remove higher concentrations if necessary to keep linearity)

- Samples
- Values should fall within the corrected standard curve
- If not, they need to be re-diluted and quantified again
- To calculate actual concentration, need to multiply quantification by dilution factor
- For example, if pico value for 1:5 dilution gave a concentration of 5.7 ng/µL, 5.7 x 5 = actual undiluted concentration of 28.5 ng/µL
- Blanks typically have between 0 and 0.6 ng/µL in them after cleanup, likely from primer dimer

Dilute

71 Based upon quantification, dilute each sample to equimolar concentrations with TE, 1 ng/µL.

72 Calculate how much TE is needed to add to 1 µL of un-diluted, bead-cleaned PCR product to get a final concentration of 1 ng/µL

- Essentially Concentration – 1

Ex:

sample	measured 1:5 dilution [], ng/uL	calculated full [], ng/uL	vol un-diluted, bead-cleaned PCR product for dilution (uL)	vol TE for dilution (uL)	dilution [] ng/uL
1	4.6	23.0	1	22.0	1
2	2.1	10.5	1	9.5	1
3	7.9	39.5	1	38.5	1
4	5.2	26.0	1	25	1
blank	n/d	0.1	5	0	0.1

73 Accurately pipet calculated volume of TE into PCR plate according to calculations

74 Accurately pipet 1 µL of undiluted, bead-cleaned PCR product into TE

- For blanks or negative controls, concentrations should be below 1 ng/µL, so just add 5 µL of undiluted, bead-cleaned PCR product to dilution plate

75 Cap, vortex, and spin 1 ng/µL dilutions

76 Store cleaned PCR products at 4°C if using within a week. Otherwise, freeze at -20°C.

Pool

77 Pool by adding 1 µL of each sample into one 1.5 mL LoBind tube.

- Don't forget blanks and mock communities
- **Easiest to collect all samples from all users for run into one pool**, instead of individual users each creating own cleaned pool
- Current MiSeq requirements at UC-Davis (10/2015)
- 15 µL
- 5 nM (~1.8 ng/µL)
- If we pool 1 µL of each 1 ng/µL sample, we need 27 samples to meet this requirement, but since we like to pool enough to send half and keep half, we need ≥ 63 samples to meet this requirement (35 µL total vol)
- Store pool at 4°C if using within a week. Otherwise, freeze at -20°C.

Clean and Concentrate

78 Clean pool with SPRIselect (Beckman Coulter B23317) beads.

- SPRIselect beads are just like the Ampure beads, but the SPRIselect beads are QC'd for precise size-selection
- Follow same protocol as described above for Ampure beads, except
- **Elute with 35 µL TE**
- Collect elution into a well-labeled 0.5 mL LoBind tube
- Pipet 15 µL of this into a separate, well-labeled 0.5 mL LoBind tube
- This 15 mL tube is the one that will be sent for sequencing
- Current MiSeq requirements at UC-Davis (10/2015)

- 15 µL
- 5 nM (~1.8 ng/µL)
- The remaining bead-cleaned pool will be used for:
- On-site storage
- Assessment by:
- Quantification (pico-green)
- Size distribution (bioanalyzer trace)
- Store 0.5 mL LoBind tube with 15 µL for sequencing at -20°C until sent for sequencing
- Store remaining cleaned pool (also in 0.5 mL LoBind tube) at 4°C if assessing via pico and bioanalyzer within a week. Otherwise, freeze at -20°C.

Assess Pool: Quantification with Qubit dsDNA HS Assay Kit

79 Sequencing pool needs to be assessed for quantity and size distribution before submitting.

Invitrogen Q32854 Kit Components:
- Qubit dsDNA HS Buffer, RT
- 200X Qubit dsDNA HS Reagent dye, RT
- HS Standard #1 (0 ng/µL), 4°C
- HS Standard #2 (10 ng/µL), 4°C

80 Before beginning
- Get ice
- Locate standards in fridge and place on ice
- Locate RT reagents and assay tubes

81 Procedure
Label assay tube lids, 1 per sample and 2 for standards

82 Prepare assay solution, keep dark, RT

# samples →	1	2	3	4	5	6	7	8	9	10
Buffer (µL)	700	900	2x550	2x650	2x750	2x850	2x950	3x700	3x766.7	3x833.3
200X dye (µL)	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.5
total volume (µL)	703.5	904.5	1105.5	1306.5	1507.5	1708.5	1909.5	2110.5	2311.5	2512.5
tube size	2 mL	2 mL	2 mL	2 mL	2 mL	2 mL	2 mL	5 mL	5 mL	5 mL

83 Add assay solution to assay tubes (keep as dark as possible)
- Standards: 190 µL
- Samples: 199 µL

84 Add DNA to assay tubes
- Standards: 10 µL
- Samples: 1 µL

85 Vortex, spin, keeping dark

86 After ≥ 5 min, run on Qubit:
1. Select correct assay: "Qubit dsDNA HS"
2. Follow directions on prompt
- When measuring samples, change volume of input DNA to 1 µL
-

1. When finished, select "Check Calibration" to record actual fluorescence values for the standards.

87 What to look for:

- Most sequencing centers quantify DNA via Qubit
- Could also quantify with pico-green as outlined above
- No need to dilute sample first
- 63 samples minimum, 1 ng/sample
- 63 ng in 35 µL TE = 1.8 ng/mL
- 256 samples maximum, 1 ng/sample (currently with our 16x16 barcodes)
- 256 ng in 35 uL TE = 7.31 ng/uL
- As long as the pool concentration is $\geq$ 1.8 ng/mL (for this primer set), you can send the pool as is.

If it is < 1.8 ng/uL, may need to either concentrate sample with another bead-cleaning or may need to re-pool samples with higher volumes/sample

Bioanalyzer Chip: High Sensitivity DNA Kit

88 Agilent 5067-4626 Kit Components

- HS Chip, RT
- Spin filters
- DNA Ladder, 4°C
- DNA markers, 4°C
- DNA dye, 4°C
- DNA gel matrix, 4°C

89 Before beginning

Equilibrate gel and dye (or gel-dye mix) to RT in the dark, 30 min.

90 Make sure gel-dye mix is prepared and < 6 weeks old, otherwise, make new

- Vortex dye 10 sec, spin
- Pipet 15 uL dye into gel matrix vial.
- Vortex 10 sec
- Pipet into spin filter (provided)
- Spin 10 min, RT, 2240 g
- Label tube cap with date of preparation
- Good for 5 chips

91 Test syringe for gel priming station for pressure by holding gloved finger against tip while dispensing 1 mL air

92 Add 400 uL new, clean PCR water to bioanalyzer's electrode washing chip

- Place into bioanalyzer and close lid. Let electrodes sit in new water until samples are loaded. At that time, lift lid of bioanalyzer to let electrodes dry

93 Procedure

1. Prime the plate
2. Set chip priming station to bottom position ("C"), move plunger to 1 mL volume, and insert new HS chip
3. Pipet 9 uL gel-dye mix into well marked with black "G"
4. Set timer for 60 seconds, then lock the latch of priming station
5. Press the plunger until it is locked into the clip
6. Wait exactly 60 seconds, then release the plunger with the clip release mechanism
7. Visually inspect that the plunger moves back at least to the 0.3 mL mark
8. Wait for 5 s, then slowly pull back the plunger to the 1 mL position
9. Open the chip priming station
10. Open bioanalyzer lid to let electrodes dry

94 Load remaining reagents

1. Pipet 9 µL of gel-dye mix to each of the wells marked with a grey "G"
2. Pipet DNA marker into all remaining wells (ladder and sample wells)

- 5 µL for ladder well
- 5 µL for sample wells
- 6 µL for sample wells where no sample is to be loaded

1. Pipet 1 µL DNA ladder into well marked with ladder symbol
2. Pipet 1 µL DNA into appropriate sample well

95 Mix reagents

- Vortex plate for 60 sec at 2400 rpm

96 Load into Agilent 2100 Bioanalyzer

1. Select Assay: dsDNA, High Sensitivity DNA.xsy
2. Enter sample information
3. If running fewer than 11 samples, indicate how many wells to run (Run samples 1 to 9, for example)

- All samples indicated need to be run if user would like to toggle between time and size of fragment assessment.
- Could manually stop run after sample 9, for example, but then option to alter x-axis to fragment length would not be available

Press Start

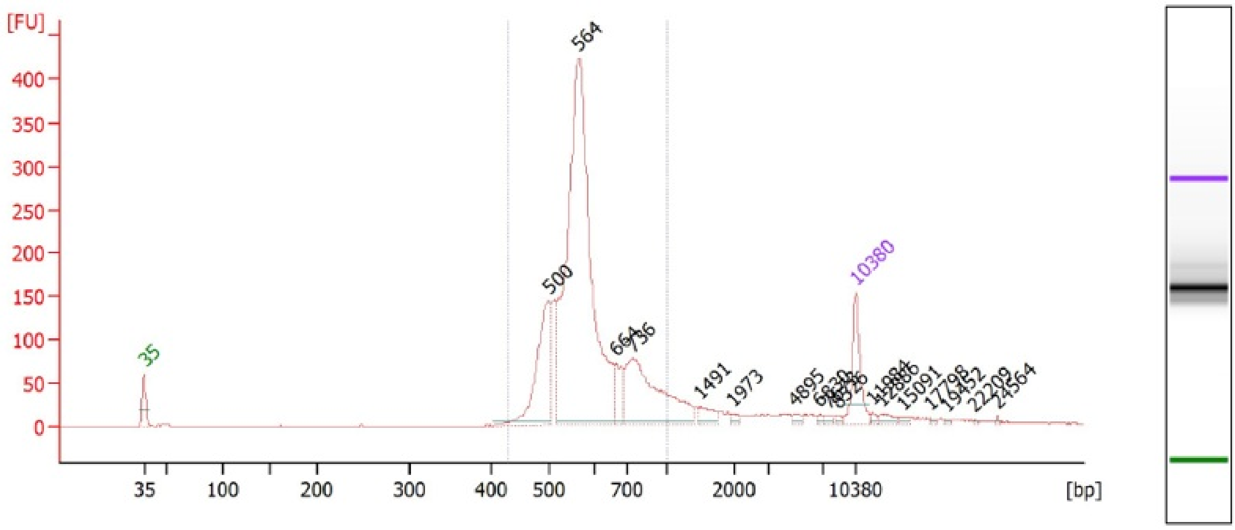
97 After chip is run (~30 min)

1. Remove HS chip and throw away (can clean and re-use…)
2. Add 400 uL new, clean PCR water to bioanalyzer's electrode washing chip

- Add washing chip to bioanalyzer and close lid. Keep electrodes stored this way.

98 What to look for:

- Can have sequencing facility run chip for us instead
- UC Davis currently charges $184/chip or $53/sample for 1-3 samples, 11/2015
- We currently pay $49.14/chip (7/2015)
- Trace can identify primer dimers
- Greatly interfere with sequencing since these are preferentially sequenced
- If present, do another clean-up before sending pool for sequencing
- Trace can show if multiple fragment sizes are present
- May indicate separate 16S and 18S pools as this primer set can detect both
- May indicate a greater issue with amplicons
- Trace indicates amplicon size
- Need amplicon size for sequencing form
- Can better estimate molarity of sample



Overall Results for sample 1 : Alma_Erin_Ella_pool_dil

Number of peaks found:	17	Corr. Area 1:	2,085.8
Noise:	0.2		

Peak table for sample 1 : Alma_Erin_Ella_pool_dil

Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	500	237.98	721.5	
3	564	1,068.00	2,867.3	
4	664	53.56	122.2	
5	736	272.66	561.2	
6	1,491	23.47	23.8	

Submit

99
- Each sequencing run should include:
- ≥ 1 PCR blank (1 per person)
- ≥ 1 even mock community
- ≥ 1 staggered mock community
- Dilute multiple pools to an equal concentration
- If submitting one pool:
- Mix multiple pools in the ratio in which you desire sequencing depth
- Ex: Pool 1 has 50 samples and is 1.8 ng/µL; Pool 2 has 100 samples and is 1.8 ng/µL. If you desire each sample to have equal sequencing depth, mix 11.7 µL Pool 1 and 23.3 uL Pool 2.
- If submitting multiple pools,
- Typically charged for pooling: Currently $96 at UCDavis (11/2015)
- Need to send bioanalyzer trace and quantification information for each pool sent
- Need to send:
- Submission form (quantification information)
- Bioanalyzer trace for each pool sent
- Indices used for demultiplexing

Appendix 1: Primer Strategy

100 Primer Strategy from Hilary Morrison (et al. at MBL). Also basically used in http://aem.asm.org/content/suppl/2011/05/13/77.11.3846.DC1/Bartram_ms_AEM_S-corrected.pdf[1], though I think they only did reverse indexing with no inline barcode on the forward.

The primers are quite 'universal'; the details:

515F (modified from the EMP primer, here's a reference[4])

NOTE: Currently used 515F primer is modified to contain a Y in 4th position of the 5' end[2], instead of the published C, this increases the

Thaumarchaea hits to from 0.4% to 70%, and the total Archaea from 50% to 80%

Matches From the Silva v106 (using 515F with C in 4th position)
Arch (0mm): 54.4%
Notably, does not hit most of the Thaumarchaea with zero mismatch, the mismatch is 4 bases from the 5' end.
Bact (0mm): 95.4%
Euk (0mm): 92.2
Arch (1mm): 96.2%
Bact (1mm): 98.4%
Euk (1mm): 97%

926R (perhaps this is a paper where this primer appears, i.e., [3]: http://www.biomedcentral.com/1471-2105/12/38/ and referenced in http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3721114/pdf/ismej201344a.pdf, Headwaters are critical reservoirs of microbial diversity for fluvial networks,
and Development of a standardized approach for environmental microbiota investigations related to asthma development in children
for what that's worth.

Matches from the Silva 115
Arch (0mm): 89%
Bact (0mm): 92.7%
Euk (0mm): 92.2%
Arch (1mm): 94.1%
Bact (1mm): 97.4%
Euk (1mm): 97.5%

Indices. This is one aspect that, if ordering new primers, could be changed. Since the indexing read is a dedicated read that does not reduce the data from other reads, having the commonly used 12 base index would only benefit. But practically, if might create a big problem if one was to mix indeces with 6 bases and others with 12 bases in the same run, please contact sequencing facility if this would even work. Generally taken from https://wikis.utexas.edu/display/GSAF/Illumina+-+all+flavors

Appendix 2: Mock Community and Blanks

101 **Background:**
We generated a mock community that approximates a surface microbial community at the San Pedro Ocean Time-series (SPOT) for several reasons: 1.) examine the extent that PCR bias influences the assessment of a community with a known input, 2.) assist in examining run-to-run variability of sequencing runs, 3.) serve as a control for clustering, classification, or other analytical procedures (e.g., is the data quality high enough for oligotyping? Yes: https://www.nature.com/articles/ismej201729).

The mock community is made from pooled 16S rRNA gene clones generated from SPOT, and 3 clones from the Giovannoni lab's clone collection (taxa that we did not have: Chloroflexi and Planctomycetes). The clones were re-grown overnight from their -80C freezer stocks, and plasmids purified via Miniprep. We then amplified the plasmids with vector primers (remove any extra *E. coli* contamination), followed a PCR with 16S rRNA gene primers that amplified nearly the full 16S primer (generally a 8-28F position and 1392-1492R). We then quantified the products and pooled them in equal proportions (even community) or from 0.1%-30% (staggered community), taking into consideration each taxon's amplicon length and sequence.

Originally the even community consisted of 10 clones (even version 1) and the staggered community consisted of 25 clones (staggered version 1). However, we later added 1 Thaumarchaea clone to the even mock community (even version 2) and 2 Thaumarchaea clones to the staggered community (staggered version 2).

102 **Usage:**
We amplify an 'even' and 'staggered' mock communities as 'positive controls' with each set of PCRs that we run. For example for every 48 PCR samples, we will include 2 mock communities (1 even and 1 staggered) and at least 1 no template control.

The lyophilized mock community should be diluted with 125 µL of TE (the final concentration will be about 0.001ng/µL). Add the TE, and then vortex, wait 5 minutes and vortex again (like you would do for a primer). Aliquot this into about 10 aliquots. I have observed some degradation from multiple freeze thaw cycles. We use 1 µL of 0.001ng/µL of mock community for each amplification because this is approximately the concentration that we expect 16S rRNA gene abundances to be found in our environmental samples. This is because we usually load 1ng environmental DNA to each PCR reaction, and 16S is roughly 1/1000 of each genome. Do not take the mock communities into clean, UV/PCR hood.

It would be interesting to add the mock communities to an environmental sample and leave the same environmental sample blank in a separate reaction. If you do this, you could see how much the mock community results are skewed by the presence of other templates, we have not done this would be interested in your experiences if you were to experiment with this.

The following table includes the clone names, accession number of the clones, and the SILVA taxonomy for the sequences in the mock community, and the proportion of each of the clones in version 2 of the mock community. For version 1, remove the Thaumarchaea sequences (red text) and re-calculate proportions.

clone_names	ARISA	Accession	Silva taxonomy v. 119	% ID to Silva119 rep_seq/99 match (full/ 515-926/ no primers)	proportion in staggered.v2	proportion ineven.v2
SAR11	ARISA_667.6	DQ009197.1	Bacteria; Proteobacteria; Alphaproteobacteria; SAR11 clade; Surface 1	99.22/99.76/99.73	31.5	9.1
OCS155_a	ARISA_435.5	D1009124.1	Bacteria; Actinobacteria; Acidimicrobiia; Acidimicrobiales; OM1 clade; Candidatus Actinomarina	98.36/100/100	15.8	9.1
OCS155_b	ARISA_419.5	DQ009123.1	Bacteria; Actinobacteria; Acidimicrobiia; Acidimicrobiales; OM1 clade; Candidatus Actinomarina; uncultured bacterium	99.17/99.76/99.73	9.0	
Tharumarchaea_MGI_a*	thaum_890_2_0.0997	KT036446	Archaea; Thaumarchaeota; Marine Group I; Unknown Order	99.1/99.76/99.73	9.0	9.1
Prochlorococcus	ARISA_828.8	DQ009356.1	Bacteria; Cyanobacteria; Cyanobacteria; SubsectionI; FamilyI; Prochlorococcus; uncultured bacterium	99.51/99.76/100	6.8	9.1
SAR86_a	ARISA_402.4	DQ009149.1	Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; SAR86 clade; uncultured bacterium	99.39/99.75/99.73	4.5	9.1
AEGEAN-169	ARISA_676.9	DQ009262.1	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales; Rhodospirillaceae; AEGEAN-169 marine group	99.09/100/100	2.3	

SAR116_a	ARISA_653.1	DQ009264.1	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR116 clade; uncultured bacterium	100/100/100	2.3	9.1
Euryarchaea_MGII*	eury25	KT036445	Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales; Marine Group II; uncultured archaeon	99.17/99.76/99.73	1.8	9.1
Flavobacteria	ARISA_726.4	DQ009108.1	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; NS2b marine group	99.03/100/100	1.8	9.1
Planctomyces**	6013	AF029077	Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales; Planctomycetaceae; Blastopirellula	100/100/100	1.8	9.1
SAR116_b	ARISA_703.7	DQ009270.1	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR116 clade	100/100/100	1.8	
SAR202_a	bats256	AY534095	Bacteria; Chloroflexi; SAR202 clade; uncultured Chloroflexi bacterium	100/100/100	1.8	9.1
SAR406	ARISA_627.8	DQ009157.1	Bacteria; Deferribacteres; Deferribacteres; Deferribacterales; SAR406 clade(Marine group A); uncultured bacterium	98.47/99.76/99.73	1.4	9.1
Flavobacteria_Formosa	ARISA_779.2	DQ009099.1	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Formosa; uncultured bacterium	98.83/99.76/99.73	0.9	

Flavobacteria_NS9	ARISA_540.1	EF572761.1	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; NS9 marine group; uncultured bacterium	99.17/99.01/98.91	0.9	
Pseudospirillum	ARISA_933.7	DQ009153.1	Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; Oceanospirillaceae; Pseudospirillum	100/100/100	0.9	
SAR86_b	ARISA_634.7	DQ009142.1	Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; SAR86 clade; uncultured bacterium	100/99.76/99.73	0.9	
SAR92	ARISA_762.8	DQ009136.1	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Alteromonadaceae; SAR92 clade; uncultured marine bacterium	100/100/100	0.9	
Thaumarchaea_MGI_b*	thaum_890_3_0.0100	KT036447	Archaea; Thaumarchaeota; Marine Group I; uncultured marine archaeon	98.03/100/100	0.9	
Verrucomicrobia	ARISA_738.8	DQ009368.1	Bacteria; Verrucomicrobia; Opitutae; MB11C04 marine group	99.46/100/100	0.9	
Rhodobacteraceae	ARISA_840	EU804911.1	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; uncultured	100/100/100	0.7	
SAR86_d	ARISA_657.6	**DQ009141.1**	Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; SAR86 clade; uncultured bacterium	100/100/100	0.7	

Flavobacteria_NS5	ARISA_749.6	DQ009088.1	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; NS5 marine group; uncultured bacterium	98.37/100/100	0.5	
SAR86_c	ARISA_584	**DQ009125.1**	Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; SAR86 clade	100/100/100	0.2	
SAR116_c	ARISA_765.7	DQ009276.1	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR116 clade; uncultured bacterium	97.82/99.76/99.73	0.1	
SAR202_b	bats259	AY534094	Bacteria; Chloroflexi; SAR202 clade; uncultured Chloroflexi bacterium	100/100/100	0.1	

full sequence (27f-1492r)/ primed amplicon(515f-926r)
*full sequence (20f-1392r)/ primed amplicon(515f-926r)
**full sequence (27f-1392r)/ primed amplicon(515f-926r)

Appendix 2: Mock Community, blanks and simple data analyses and WARNING on 18S.

103 **Data analysis:**
We add perfect sequences in with the quality filtered reads from the sequencer to easily assess the accuracy of the PCR-to-OTU table procedures. If you just analyze the mock communities by themselves, with the perfect sequences, you might over or underestimate the ability of clustering algorithms ability to correctly split your environmental sequences. For example, we found that pairing the amplified mock community sequences with those of environmental sequences split some of the taxa with certain settings or pipelines, which did not occur when we analyzed them separately.

Find 2 files on the Fuhrman lab website that will help with assessment and analysis of the mock communities. These files happen to be formatted for the QIIME pipeline, which you may use, but you may like to instead format them for mothur, or another 16S analysis tool of choice. The two 4 of files are correspond to the version 1 or 2 of the staggered and even communities. These files have the full length expected 16S sequences for the mock communities, in fasta format, in the proportions that would be if the mock community performed perfectly. You will probably want to trim the sequence files to only include the 16S fragment that your assay amplified. If you are using the 515F/926R (our preferred primers), the following primer removal will work, if you download the cutadapt software:
http://cutadapt.readthedocs.io/en/stable/installation.html

```
#remove reverse primer (reverse complement of primer)
cutadapt -a AAACTYAAAKRAATTGRCGG in.silico.full.even.fa -o no.rev.in.silico.even.fa -e 0.2 --discard-untrimmed
#remove forward primer (5'-3' orientation of primer)
cutadapt -g GTGYCAGCMGCCGCGGTAA no.rev.in.silico.even.fa -o no.prime.no.rev.in.silico.even.fa -e 0.2 --discard-untrimmed

#remove reverse primer (reverse complement of primer)
cutadapt -a AAACTYAAAKRAATTGRCGG in.silico.full.stag.fa -o no.rev.in.silico.stag.fa -e 0.2 --discard-untrimmed
#remove forward primer (5'-3' orientation of primer)
cutadapt -g GTGYCAGCMGCCGCGGTAA no.rev.in.silico.stag.fa -o no.prime.no.rev.in.silico.stag.fa -e 0.2 --discard-untrimmed
```

Now that you have a file with the trimmed version of your mock community sequences, formatted for QIIME, add them to your sequences read files for clustering (probably at the bottom so they are least like to be seeds for cluster generation). If you trimmed and removed the primers from the files above, one good option is to add them to your sequences after they have been demulitplexed, merged, quality filtered, and primers trimmed. Like this:

#simple unix concatenate command to concatenate all of your sample sequence reads, including the amplified mock community reads and the *in silico* (perfect) mock community reads
cat seqs.to.cluster.fa no.prime.no.rev.in.silico.even.fa no.prime.no.rev.in.silico.stag.fa > seqs.to.cluster.plus.mock.fa

Then, follow through your chimera checking, clustering, classification, and OTU table generation pipeline just as you normally would. If you are using QIIME, with the file formats above, the mock communities will be called “in.silico.even” and “in.silico.stag” in your OTU table.

WARNING: 18S forward and reverse reads DO NOT overlap, therefore in merging steps common to most rRNA analyses all 18S sequences would be lost if the step requires overlap. Therefore we recommend first trimming for quality control then merging forward and reverse reads separated by a single N and using k-mer base classifier.

Here is a sample set of unix commands that our lab uses for this:

```
#!/bin/bash
# Program:
# This pipeline is used for 16s and 18s sequences analysis
# Pipelines and applications used in this script:
# Trimmomatic, Usearch, QIIME,seqtk, cutadapt, BLAST
# Required files:
# R1 and R2 fastq files for each sample separately
# Final outputs:
# A 16S OTU table, a 18S OTU table, and two chloroplast table showing  how the chloroplast OTUs are reassigned in PhytoRef and NCBI databases
# History:
# Liv, 8/21/2017: First release
# Liv, 8/23/2017: add steps of classifying chloroplast reads using blastn against NCBI and PhytoRef

# Note that the way to create a list depends on how you name your sample
ls *R1.fastq| cut -d '.' -f 1-13>sample.name
for i in `cat sample.name`;
do

# Do qualify filtering on fwd and rev reads using Trimmomatic
java -jar /korriban/liv/Trimmomatic-0.36/trimmomatic-0.36.jar PE -phred33 "$i"_R1.fastq "$i"_R2.fastq R1_pe R1_se R2_pe R2_se SLIDINGWINDOW:4:20 MINLEN:200;
# Rename files that pass QC and remove files that don't pass QC
mv R1_pe "$i"_R1_trimSW4_20.fastq;
mv R2_pe "$i"_R2_trimSW4_20.fastq;
rm R1_se;
rm R2_se;

# Merge paired ends using Usearch (alternative: Vsearch or FLASH) (Note that merged reads are 16s, and non-merged reads are 18s)
~/bin/usearch9.2.64_i86linux32 -fastq_mergepairs "$i"_R1_trimSW4_20.fastq -reverse "$i"_R2_trimSW4_20.fastq -fastqout "$i".merged.fastq -fastq_maxdiffs 3 -fastq_minmergelen 300 -fastqout_notmerged_fwd notmerged_"$i".R1_trimSW4_20.fastq -fastqout_notmerged_rev notmerged_"$i".R2_trimSW4_20.fastq

########## for 16s reads ##########
# Convert to fasta file
convert_fastaqual_fastq.py -f "$i".merged.fastq -o 16s.fna/ -c fastq_to_fastaqual

# Remove primers using cutadapt
cutadapt -g ^GTGYCAGCMGCCGCGGTAA -o 16s.fna/test.fwdprimer.removed.fna 16s.fna/"$i".merged.fna --discard-untrimmed;

cutadapt -a AAACTYAAAKRAATTGRCGG$ -o 16s.fna/primers_removed_"$i".merged.fna 16s.fna/test.fwdprimer.removed.fna --discard-untrimmed
# Do chimera checking by sample and remove chimeric reads
identify_chimeric_seqs.py -m usearch61 -i 16s.fna/primers_removed_"$i".merged.fna -o 16s.fna/usearch61_chimera_checking/ --suppress_usearch61_ref
filter_fasta.py -f 16s.fna/primers_removed_"$i".merged.fna -o 16s.fna/primers_removed_nochimera_"$i".merged.fna -s 16s.fna/usearch61_chimera_checking/chimeras.txt -n

########## for 18s reads ##########
```

 # Trim all the non-merged fwd reads to the same length
java -jar /korriban/liv/Trimmomatic-0.36/trimmomatic-0.36.jar SE -phred33 notmerged_"$i".R1_trimSW4_20.fastq notmerged_"$i".R1_trimSW4_20_len190.fastq CROP:190 MINLEN:190;

 # Trim all the non-merged rev reads to the same length
java -jar /korriban/liv/Trimmomatic-0.36/trimmomatic-0.36.jar SE -phred33 notmerged_"$i".R2_trimSW4_20.fastq notmerged_"$i".R2_trimSW4_20_len190.fastq CROP:190 MINLEN:190

 # Convert the non-merged rev reads to its reverse complement with seqkt
seqtk seq -r notmerged_"$i".R2_trimSW4_20_len190.fastq > notmerged_"$i".R2_trimSW4_20_len190_rc.fastq

 # Split fwd and rev fastq files into fasta and quality files
convert_fastaqual_fastq.py -f notmerged_"$i".R1_trimSW4_20_len190.fastq -o 18s.fna/ -c fastq_to_fastaqual

convert_fastaqual_fastq.py -f notmerged_"$i".R2_trimSW4_20_len190_rc.fastq -o 18s.fna/ -c fastq_to_fastaqual

 # Add N at the beginning of reverse complement
sed -e 's/^/N/' 18s.fna/notmerged_"$i".R2_trimSW4_20_len190_rc.fna | sed -e 's/N>/>/' > 18s.fna/N.added.notmerged_"$i".R2_trimSW4_20_len190_rc.fna

 # Extract headers from fwd reads
sed -n '1~2p' 18s.fna/notmerged_"$i".R1_trimSW4_20_len190.fna >18s.fna/R1;

 # Extract headers from rev reads
sed -n '1~2p' 18s.fna/notmerged_"$i".R2_trimSW4_20_len190_rc.fna >18s.fna/R2;

 # Get shared headers
cat 18s.fna/R1 18s.fna/R2 | sort | uniq -d | cut -d '>' -f 2 >18s.fna/filter.txt;

 # filter fasta file
filter_fasta.py -f 18s.fna/notmerged_"$i".R1_trimSW4_20_len190.fna -o 18s.fna/temp_R1.fna -s 18s.fna/filter.txt;

filter_fasta.py -f 18s.fna/notmerged_"$i".R2_trimSW4_20_len190_rc.fna -o 18s.fna/temp_R2.fna -s 18s.fna/filter.txt;

 # Remove headers from reverse complement
cut -d '>' -f 1 18s.fna/temp_R2.fna > 18s.fna/r2.no.header;

 # Paste r1 and reverse complement of r2 together
paste 18s.fna/temp_R1.fna 18s.fna/r2.no.header -d '' > 18s.fna/"$i".mergedN.18s.fna;

 # Remove primers using cutadapt
cutadapt -g ^GTGYCAGCMGCCGCGGTAA -o 18s.fna/temp.fwdprimer.removed.fna 18s.fna/"$i".mergedN.18s.fna -e 0.2 --discard-untrimmed

cutadapt -a AAACTYAAAKRAATTGRCGG$ -o 18s.fna/primers_removed_"$i".mergedN.18s.fna 18s.fna/temp.fwdprimer.removed.fna -e 0.2 --discard-untrimmed

 # Do chimera checking by sample and remove chimeric reads
identify_chimeric_seqs.py -m usearch61 -i 18s.fna/"$i".mergedN.18s.fna -o 18s.fna/usearch61_chimera_checking/ --suppress_usearch61_ref

filter_fasta.py -f 18s.fna/"$i".mergedN.18s.fna -o 18s.fna/nochimera_"$i".mergedN.18s.fna -s 18s.fna/usearch61_chimera_checking/chimeras.txt -n

done

mkdir 16s.fna/no.chimera
mv 16s.fna/primers_removed_nochimera_* 16s.fna/no.chimera/
mkdir 18s.fna/no.chimera
mv 18s.fna/primers_removed_nochimera_* 18s.fna/no.chimera/

 # Combine all the 16S fasta files and in silico 16S mock fasta files into a single fasta file
cat 16s.fna/no.chimera/*.fna ../insilico_16s/*fasta>16s.insilico_mock_combined_chimera_removed.fna

 # Combine all the 18S fasta files and in silico 18S mock fasta files into a single fasta file
cat 18s.fna/no.chimera/*.fna ../insilico_18s/*fasta>18s.insilico_mock_combined_chimera_removed.fna

 # OTU picking using uclust
pick_otus.py -i 16s.insilico_mock_combined_chimera_removed.fna -o otu_picking_16s_uclust99/ -s 0.99

pick_otus.py -i 18s.insilico_mock_combined_chimera_removed.fna -o otu_picking_18s_uclust99/ -s 0.99

 # Pick the representative fna
pick_rep_set.py -i otu_picking_16s_uclust99/16s.insilico_mock_combined_chimera_removed_otus.txt -f 16s.insilico_mock_combined_chimera_removed.fna -m most_abundant -o otu_picking_16s_uclust99/rep_otus.fasta

pick_rep_set.py -i otu_picking_18s_uclust99/18s.insilico_mock_combined_chimera_removed_otus.txt -f 18s.insilico_mock_combined_chimera_removed.fna -m most_abundant -o otu_picking_18s_uclust99/rep_otus.fasta

 # Assign taxonomy (reference database: SILVA (for 16S) and PR2 (for 18S))
assign_taxonomy.py -i otu_picking_16s_uclust99/rep_otus.fasta –t /Silva119_release/taxonomy/99/taxonomy_99_7_levels.txt -r /Silva119_release/rep_set/99/Silva_119_rep_set99.fna --similarity 0.97 -o otu_picking_16s_uclust99/silva119_taxonomy/

assign_taxonomy.py -i otu_picking_18s_uclust99/rep_otus.fasta –r /pr2_gb203_version_4.5/pr2_gb203_version_4.5.fasta -t /pr2_gb203_version_4.5/pr2_gb203_version_4.5.taxo -o otu_picking_18s_uclust99/pr2_taxonomy/ -m rdp --rdp_max_memory 40000

 # make otu table
make_otu_table.py -i otu_picking_16s_uclust99/16s.insilico_mock_combined_chimera_removed_otus.txt -t otu_picking_16s_uclust99/silva119_taxonomy/rep_otus_tax_assignments.txt -o otu_picking_16s_uclust99/16s.insilico_mock.biom

make_otu_table.py -i otu_picking_18s_uclust99/18s.insilico_mock_combined_chimera_removed_otus.txt -t otu_picking_18s_uclust99/pr2_taxonomy/rep_otus_tax_assignments.txt -o otu_picking_18s_uclust99/18s.insilico_mock.biom

 # Convert biom to txt
biom convert -i otu_picking_16s_uclust99/16s.insilico_mock.biom -o otu_picking_16s_uclust99/16s.insilico_mock.txt --header-key taxonomy --to-tsv
biom convert -i otu_picking_18s_uclust99/18s.insilico_mock.biom -o otu_picking_18s_uclust99/18s.insilico_mock.txt --header-key taxonomy --to-tsv

#####for 16S reads assigned as chloroplast in SILVA database ##########
 # Get a Chloroplast only OTU table from 16s OTU table
filter_taxa_from_otu_table.py -i otu_picking_16s_uclust99/16s.insilico_mock.biom -o otu_picking_16s_uclust99/chloroplast_only.biom -p Chloroplast

 # Extract a Chloroplast only fasta file based the Chloroplast only OTU table
filter_fasta.py -f otu_picking_16s_uclust99/rep_otus.fasta -o otu_picking_16s_uclust99/chloroplast_rep_otus.fasta -b otu_picking_16s_uclust99/chloroplast_only.biom

 # A BLAST search against NCBI nucleotide
blastn -db /home/korriban/db/blast_nt_genomic_refseq/nt –query chloroplast_rep_otus.fasta -out result3.out -num_threads 120 -max_target_seqs 1 -outfmt "6 qseqid sseqid pident length mismatch gapopen qstart qend sstart send evalue bitscore sseqid sallseqid sgi sacc staxids sscinames scomnames stitle"

 # Create a custom PhytoRef database
makeblastdb -in PhytoRef_with_taxonomy.fasta -out phytoref.db.taxo -dbtype nucl

 # A BLAST search against PhytoRef database
blastn -db /korriban/liv/PhytoRef/phytoref.db.taxo -query otu_picking_16s_uclust99/chloroplast_rep_otus.fasta -out 16s_chl_blast_phytoref.out -num_threads 8 -max_target_seqs 1 -outfmt 6

Appendix 3: General PCR Considerations

104 Generally PCR strips should be performed in the UV, Biosafety hood (we typically use the Stratlinker for plastics and PCR water, and other reagents that can be UV'd). In general we have not been able to trace contamination issues to lab practices, but it is probably a good idea to take some steps to reduce contamination such as cross contamination or from the lab.

Some steps to perhaps reduce contamination
Wipe down the inside of the Biosafety hood with 10% Bleach. Wipe down the pipettes with 10% Bleach.
Leave the UV light on the hood for about 15 minutes.

Aliquot all PCR reagents and ideally only use them one time, especially the buffer, MgSO4, and water into thin walled PCR strip tubes.
You can also treat all plastic ware in the Stratalinker and/or soak in a 10% bath of bleach.
Consider opening tubes with DNA individually, being careful where your thumb goes as to avoid cross contamination.
Add master mix to PCR tubes before DNA.
Treat reagents, that are not harmed by UV, with UV: the water, buffer and magnesium. This can be performed in 300uL strip tubes in the Stratalinker by setting the timer to 9999. This is about 10 minutes. I typically perform this step at the beginning of a days PCR setup.
Load all samples + blanks into thermocycler before moving on to the mock communities which are relatively concentrated with 16S PCR product.

We used to typically run reactions in triplicate by adding, say 72uL of MM to a tube and then 3uL of DNA and then split into 3 tubes of 25uL each, though we have not confirmed that this makes a difference, if nothing else it gives more PCR product with fewer cycles.

After PCR combine triplicate reactions on lab bench. At this point if there is slight cross contamination it is not as big deal of a deal as it was before since now the samples are all barcoded. The worst it could do is mess up your quantifications, which is not ideal (potentially, for example, making your blanks quantify higher than they should).

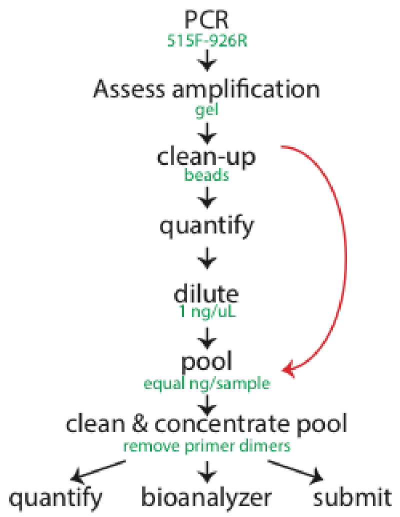
The polymerase master mix indicated below was used for some Fuhrman Lab amplicons, and Alma/David found that the mock communities generated with this polymerase were equivalent to those generated with the 5-Prime Taq master mix indicated in this protocol.

Platinum Hifi PCR Master Mix---------------------------1x
Invitrogen Hifi Buffer (11304-029) (10x)--------2.5 µL----1x [final]
dNTPs (10mM) (promega cat #: U1515) ------------0.5 µL--200uM [final]
Forward Primer (10µM)--------------------------1 µL-----0.4uM [final]
Reverse Primer (10µM)--------------------------1 µL-----0.4uM [final]
Invitrogen MgSO4 (11304-029)(50mM)-------------1 µL------2mM [final]
Invitrogen Platinum Hifi Taq (11304-029)--------0.2 µL--x Unit [final]
DNA Template (generally, 1ng/µl)---------------1 µL------1ng [final]
Molecular Grade Water (VWR cat#:95043-414)-----17.8 µL
Total Reaction Volume ------------------------25µL

Appendix 4: PCR clean-up notes

105 When more high-throughput sequencing is desired, can clean-up PCR reactions with normalization plates. We have successfully used SequalPrep Normalization Plate (Invitrogen A10510-01), but tested when we were doing triplicate PCR reactions/sample.

- According to protocol, most PCR reactions give 25-100 ng/uL in 20 µL reactions.
- Add at least 250 ng of PCR product suggested per well
- Standard elutions using 20 µL elution volume yields concentrations of 1-2 ng/µL
- aka, 25 ng kept
-
- Theoretically negates the need to quantify and dilute cleaned PCR reactions before pooling.



Appendix 5: Sequencing Submission

106 Current requirements/pricing (11/2015)

		UCDavis	USC
HiSeq 2500 Rapid Run, PE250	min vol	15 mL	
	min nM	5 nM, ~1.8 ng/uL for our amplicons	
	$	$5380/lane	$3700/lane
MiSeq 600 cycles v3 PE300	min vol	15 mL	
	min nM	5 nM, ~1.8 ng/uL for our amplicons	
	$	$2669	
MiSeq 500 cycles v2 PE250	min vol	15 mL	
	min nM	5 nM, ~1.8 ng/uL for our amplicons	
	$	$1967	
Library Pooling	$	$96, up to 6 samples	
DNA Quant (Qubit)	$	$96, up to 12 samples	
Library Quant (qPCR)	$	$36/sample	
Bioanalyzer HS	$	$53, 1-3 samples $184, 4-11 samples	
QC: Qubit, bioanalyzer, bead-clean	$		$120/sample, 1-11

107 Current UCDavis form (11/2015)

Customer Prepared Library- Sequencing Submission Form

Indicate the Illumina sequencing platform:
___MiSeq __HiSeq2500 __ HiSeq3000

PI on Genome Center Account: ________ Date: ________
PI email: ________ Your Name: ________
Institute: ________ Email: ________
DaFIS / PO to be billed (required): ________ Phone No. ________

Please email this form to dnatech@ucdavis.edu and submit a printed copy with your sample to 1410 GBSF.

IMPORTANT: If libraries require custom sequencing primers please indicate under 'special instructions'

	Sample #1
Library Name	Fuhrman lab pool
Organism (Scientific and Common Name)	Natural Micorbial Community
What library type (WGS, RNA-seq, RAD/GBS, Amplicon, 16S Amplicon, PCR-free) and **kit** (TruSeq, Kapa, custom) were used?	Amplicon
Sample **concentrations** - 5 nM minimum **From:**__ Nanodrop | _x Qubit | _x BioAnalyzer	x ng/uL qubit
Sample **volume** (ul) - 15 ul minimum	15 uL
Library size with adapters (e.g. 250 bp, 450 bp)	560 bp
Bioanalyzer traces are required. **Do you need us to run a Bioanalyzer trace** (at extra cost)?	no, see attached
Specify **indexing read** (none [in-line], 6bp, 8bp, dual)	6 base index to be read during a dedicated indexing read
Do we need to **demultiplex**? If yes, write vendor, strategy, and fill in BarcodeInfo worksheet	Yes. Demultiplex on 6-bp indices to be read during a dedicated indexing read, see attached
Pooling requested (at extra cost) (i.e. none, all into 1 pool, or describe custom)	Please pool x% amplicon pool, x% metagenomes. Metagenomes should replace PhiX
No. of seq. **lanes** requested (for pool if applicable)	n/a
Type of sequencing run (SR50, PE100, PE150, PE250, Rapid PE250, etc.)	PE300
Does the library contain **low complexity** regions? (i.e. enzyme recognition sites, GC or AT rich areas, in-line barcodes, amplicons)	5-base in-line barcodes (to be read in forward read)
Special Instructions (i.e. custom seq. primers)	3 reads: Forward Read, Index Read, Reverse Read required with standard illumina sequencing primers. We are sending x tubes (labels) to be run together (x% amplicon pool, x% metagenomes). Metagenomes should replace PhiX.

UCDAVIS DNA Technologies Core - Revised 9/21/2015

sample	subsample name	Index	Index (RC)-demultiplex with this
Fuhrman Lab Pool	Euk-V4-YRR1-B	AAACAC	GTGTTT
Fuhrman Lab Pool	Euk-V4-YRR2-B	TGAAGG	CCTTCA
Fuhrman Lab Pool	Euk-V4-YRR3-B	AACATA	TATGTT
Fuhrman Lab Pool	Euk-V4-YRR4	CGCGTC	GACGCG
Fuhrman Lab Pool	Euk-V4-YR1-B	TCGGCA	TGCCGA
Fuhrman Lab Pool	Euk-V4-YR2-B	CTCAGA	TCTGAG
Fuhrman Lab Pool	Euk-V4-YR3-B	ACTGAT	ATCAGT
Fuhrman Lab Pool	Euk-V4-YR4	ATGAGC	GCTCAT
Fuhrman Lab Pool	cDNAquart_11	GTCCGC	GCGGAC
Fuhrman Lab Pool	SIPspot1014_12	GTGAAA	TTTCAC
Fuhrman Lab Pool	SIPspot1014_13	CACCGG	CCGGTG

Primers

108

Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads. *Appl. Environ. Microbiol.* **77,** 3846–3852 (2011).

Parada, A., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples. *Environ. Microbiol.* (2015). doi:10.1111/1462-2920.13023

Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. Removing Noise From Pyrosequenced Amplicons. *BMC Bioinformatics* **12,** 38 (2011).

Walters, W. A. *et al.* PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* **27,** 1159–1161 (2011).