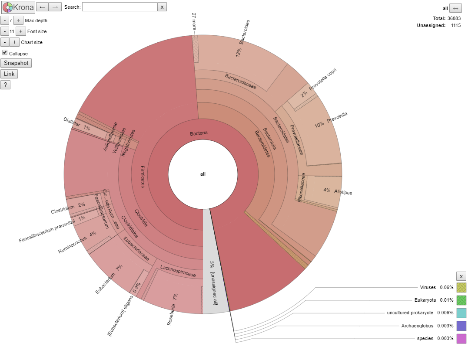
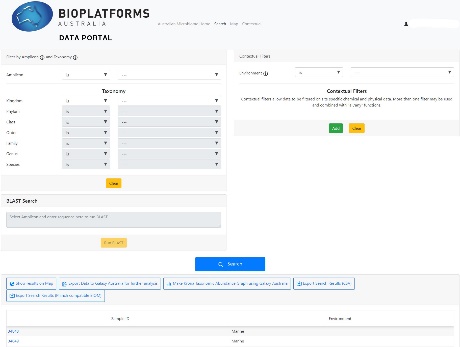
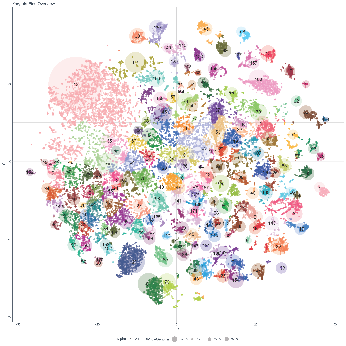
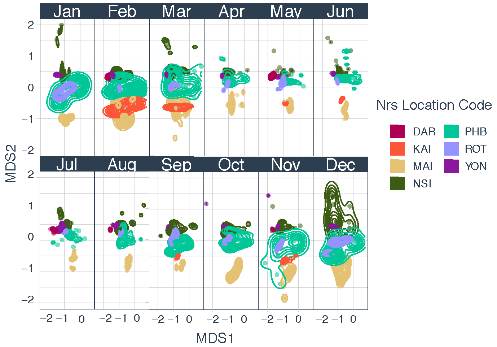
**AUSTRALIAN MICROBIOME INITIATIVE SCIENTIFIC MANUAL**







**Authors**: Mazard S., Bissett A., Koval, J., Tinning, M., Ostrowski M., Smith M., Torsten T., Van de Kamp J., Webster N.

**Date**: October 2019

**Version** 1.2 (draft)

**Table of Contents**

[1 SAMPLE COLLECTION AND PROCESSING 6](#_Toc22208887)

[1.1 SAMPLE: SOIL 6](#_Toc22208888)

[1.1.1 Sampling, processing and nucleic acid extraction equipment: 6](#_Toc22208889)

[1.1.2 Sampling 6](#_Toc22208890)

[1.1.3 DNA extraction 7](#_Toc22208891)

[1.1.4 Chemical analysis and Particle size 7](#_Toc22208892)

[1.2 SAMPLE: MARINE SEDIMENTS 8](#_Toc22208893)

[1.2.1 Sampling, processing and nucleic acid extraction equipment: 8](#_Toc22208894)

[1.2.2 Sampling 8](#_Toc22208895)

[1.2.3 Sediment grain size (% fines) 8](#_Toc22208896)

[1.2.4 Sediment TOC 9](#_Toc22208897)

[1.2.5 RNA extraction 9](#_Toc22208898)

[1.2.6 TurboDNase treatment for 12 RNA samples 10](#_Toc22208899)

[1.2.7 RNA Clean 10](#_Toc22208900)

[1.2.8 DNA extraction 11](#_Toc22208901)

[1.2.9 DNA clean 11](#_Toc22208902)

[1.3 SAMPLE: PELAGIC WATERS 12](#_Toc22208903)

[1.3.1 Sampling, processing and nucleic acid extraction equipment: 12](#_Toc22208904)

[1.3.2 Seawater sampling 12](#_Toc22208905)

[1.3.3 Nucleic acid material extraction 12](#_Toc22208906)

[1.4 SAMPLE: COASTAL SEAWATER 14](#_Toc22208907)

[1.4.1 Sampling, processing and nucleic acid extraction equipment: 14](#_Toc22208908)

[1.4.2 Seawater sampling 14](#_Toc22208909)

[1.4.3 Nucleic acid material extraction 14](#_Toc22208910)

[1.4.4 Preliminary protocol for water sampling for metatranscriptomics 15](#_Toc22208911)

[1.4.5 Sampling for DIC and TA 15](#_Toc22208912)

[1.5 SAMPLE: SEAGRASS 16](#_Toc22208913)

[1.5.1 Sampling, processing and nucleic acid extraction equipment: 16](#_Toc22208914)

[1.5.2 Surficial sediment sampling 16](#_Toc22208915)

[1.5.3 Seagrass sampling 16](#_Toc22208916)

[1.5.4 Sediment nutrients 17](#_Toc22208917)

[1.5.5 Sediment grain size (% fines) 17](#_Toc22208918)

[1.5.6 Sediment TOC 17](#_Toc22208919)

[1.5.7 Sediment DNA and RNA extraction 17](#_Toc22208920)

[1.5.8 Seagrass RNA extraction 17](#_Toc22208921)

[1.5.9 RNA clean 17](#_Toc22208922)

[1.5.10 RNA quality control, transcriptome and metatranscriptome sequencing 18](#_Toc22208923)

[1.6 SAMPLE: SPONGES 19](#_Toc22208924)

[1.6.1 Sampling, processing and nucleic acid extraction equipment: 19](#_Toc22208925)

[1.6.2 General notes on collection: 19](#_Toc22208926)

[1.6.3 Sampling 19](#_Toc22208927)

[1.6.4 DNA extraction for amplicon sequencing 19](#_Toc22208928)

[1.6.5 RNA extraction, host-mRNA removal and rRNA depletion 19](#_Toc22208929)

[1.7 SAMPLE: SEAWEED (*Ecklonia* *radiata*) 21](#_Toc22208930)

[1.7.1 Sampling, processing and nucleic acid extraction equipment: 21](#_Toc22208931)

[1.7.2 General note for the collection: 21](#_Toc22208932)

[1.7.3 Sampling 21](#_Toc22208933)

[1.7.4 DNA extraction 21](#_Toc22208934)

[1.7.5 Additional information: 21](#_Toc22208935)

[1.8 SAMPLE: CORAL 23](#_Toc22208936)

[1.8.1 Sampling, processing and nucleic acid extraction equipment: 23](#_Toc22208937)

[1.8.2 General note for the collection: 23](#_Toc22208938)

[1.8.3 Sampling: 23](#_Toc22208939)

[1.8.4 DNA extraction: 23](#_Toc22208940)

[2 CONTEXTUAL METADATA 24](#_Toc22208941)

[2.1 CORE METADATA 24](#_Toc22208942)

[2.2 ENVIRONMENTAL DESCRIPTORS 24](#_Toc22208943)

[2.2.1 Controlled vocabulary descriptors 25](#_Toc22208944)

[2.3 PHYSICAL DATA 26](#_Toc22208945)

[2.3.1 Controlled vocabulary descriptors 26](#_Toc22208946)

[2.4 CHEMICAL DATA 26](#_Toc22208947)

[2.5 ANALYTICAL METHODS 27](#_Toc22208948)

[2.5.1 SOIL 27](#_Toc22208949)

[2.5.2 SEDIMENT 27](#_Toc22208950)

[2.5.3 SEAWATER 27](#_Toc22208951)

[2.5.4 HOST ASSOCIATED 27](#_Toc22208952)

[3 SEQUENCING PROCEDURE 28](#_Toc22208953)

[3.1 SAMPLE SUBMISSION 28](#_Toc22208954)

[3.2 AMPLICONS 29](#_Toc22208955)

[3.2.1 Sequencing: Bacterial 16S rRNA gene (27F-519R) 29](#_Toc22208956)

[3.2.1.1 Primers for amplification: 27F (Lane 1991) and 519R (Lane et al. 1993) 29](#_Toc22208957)

[3.2.1.2 Preparation of master mix for amplification (for 1 rxn) 29](#_Toc22208958)

[3.2.1.3 Thermocycler conditions for amplification (for 96 well thermocycler) 29](#_Toc22208959)

[3.2.1.4 Amplification method 29](#_Toc22208960)

[3.2.1.5 Sequencing of the 27F and 519R fragment 30](#_Toc22208961)

[3.2.1.6 References 30](#_Toc22208962)

[3.2.2 Sequencing: Archaeal 16S (A2F-519R) 31](#_Toc22208963)

[3.2.2.1 Primers for amplification: A2F and 519R 31](#_Toc22208964)

[3.2.2.2 Preparation of master mix for amplification A2F / 519R (for 1 rxn) 31](#_Toc22208965)

[3.2.2.3 Thermocycler conditions for amplification (for 96 well thermocycler) 31](#_Toc22208966)

[3.2.2.4 Amplification method 31](#_Toc22208967)

[3.2.2.5 Sequencing of the A2F and 519R fragment 32](#_Toc22208968)

[3.2.2.6 References 32](#_Toc22208969)

[3.2.3 Sequencing: Eukaryotic 18S V4 33](#_Toc22208970)

[3.2.3.1 Primers for amplification: V4 region 33](#_Toc22208971)

[3.2.3.2 Preparation of master mix for amplification V4 region (for 1 rxn) 33](#_Toc22208972)

[3.2.3.3 Thermocycler conditions for amplification (for 96 well thermocycler) 33](#_Toc22208973)

[3.2.3.4 Amplification method 33](#_Toc22208974)

[3.2.3.5 Sequencing of the V4 fragment 34](#_Toc22208975)

[3.2.3.6 References 34](#_Toc22208976)

[3.2.3.7 Appendix 1 Primer sequences 34](#_Toc22208977)

[3.2.4 Sequencing: Eukaryotic 18S V9 35](#_Toc22208978)

[3.2.4.1 Primers for amplification: 18S V9 region (1391F and EukBR) 35](#_Toc22208979)

[3.2.4.2 Preparation of master mix for amplification 18S V9 region (for 1 rxn) 35](#_Toc22208980)

[3.2.4.3 Thermocycler conditions for amplification (for 96 well thermocycler) 35](#_Toc22208981)

[3.2.4.4 Amplification method 36](#_Toc22208982)

[3.2.4.5 Sequencing of the 1391F and EukBR 18S V9 fragment 36](#_Toc22208983)

[3.2.4.6 Primer ordering and resuspension: tips and getting started (From EMP protocol) 37](#_Toc22208984)

[3.2.4.7 References 37](#_Toc22208985)

[3.2.4.8 Appendix 1 Primer sequences 38](#_Toc22208986)

[3.2.4.9 Appendix 2 Index sequences for sample sheet 39](#_Toc22208987)

[3.2.5 Sequencing: Fungal ITS (ITS1F and ITS4) 40](#_Toc22208988)

[3.2.5.1 Primers for amplification: ITS1F and ITS4 region 40](#_Toc22208989)

[3.2.5.2 Preparation of master mix for amplification ITS1F and ITS4 region (for 1 rxn) 40](#_Toc22208990)

[3.2.5.3 Thermocycler conditions for amplification (for 96 well thermocycler) 40](#_Toc22208991)

[3.2.5.4 Amplification method 40](#_Toc22208992)

[3.2.5.5 Sequencing of the ITS1F and ITS4 fragment 41](#_Toc22208993)

[3.2.5.6 References 42](#_Toc22208994)

[3.3 METAGENOMICS 43](#_Toc22208995)

[3.3.1 Quality check of the input DNA 43](#_Toc22208996)

[3.3.2 Library preparation 43](#_Toc22208997)

[3.3.3 Quality check of the libraries 43](#_Toc22208998)

[3.3.4 Sequencing 43](#_Toc22208999)

[3.4 METATRANSCRITPTOMICS 44](#_Toc22209000)

[3.4.1 Quality check of the input DNA 44](#_Toc22209001)

[3.4.2 Library preparation 44](#_Toc22209002)

[3.4.3 Quality check of the libraries 44](#_Toc22209003)

[3.4.4 Sequencing 44](#_Toc22209004)

[4 BIOINFORMATIC AND ANALYTICS WORKFLOW 45](#_Toc22209005)

[4.1 AMPLICONS 45](#_Toc22209006)

[4.1.1 Bacterial and Archaeal 16S rRNA gene Amplicon Analysis Workflow 45](#_Toc22209007)

[4.1.2 Eukaryotic 18S rRNA gene Amplicon Analysis Workflow 49](#_Toc22209008)

[4.1.3 Fungal ITS Amplicon Analysis Workflow 53](#_Toc22209009)

[4.2 METAGENOMICS 57](#_Toc22209010)

[4.3 METATRANSCRIPTOMICS 58](#_Toc22209011)

**FOREWORD**

The **Australian Microbiome Initiative** is a continental-scale, collaborative project aspiring to characterise the diversity and ecosystem service provision of the microorganisms inhabiting natural Australian ecosystems.

The initiative aims to develop an Australian microbial genomics resource for management, monitoring, and R&D purposes.

This manual describes the Standard Operating Procedure collection for the Australian Microbiome Initiative. All sampling methods, data generated and analytical workflows for the Australian Microbiome initiative activities are publicly accessible (with free registration) on the Bioplatforms Data portal (<https://data.bioplatforms.com/organization/about/australian-microbiome>). The requirements for access and use of the data, as well as appropriate [acknowledgements](https://www.australianmicrobiome.com/protocols/acknowledgements/) in communications that arise from the initiative consortium’s work and data use are defined in the [Data policy](https://www.australianmicrobiome.com/data-code/data-policy/) and [Communication policy](https://www.australianmicrobiome.com/initiative-activities/comms-policy/)

**Contacts:**

Program Manager – Dr Sophie Mazard, [smazard@bioplatforms.com](mailto:smazard@bioplatforms.com)

Science Lead – Dr Andrew Bissett, [Andrew.Bissett@csiro.au](mailto:Andrew.Bissett@csiro.au)

1. SAMPLE COLLECTION AND PROCESSING

The sample collection SOPs describe the steps involved in:

* sampling microbial communities
* obtaining relevant metadata
* processing sample for subsequent contextual metadata and sequence analysis

The methodologies for definition of the contextual metadata are detailed in section 2 of the manual.

**Please contact the Program Manager for the initiative before starting any collection, so that specific workflow information can be provided.**

* 1. SAMPLE: SOIL
     1. Sampling, processing and nucleic acid extraction equipment:
* Listing to be defined
  + 1. Sampling

1. Select a 25 x25 m plot at the sample site in a reasonably homogenous environment that reflects the characteristics of the site (based on soil, vegetation and land use).
2. Collect soil (comprising between 9-30 samples) from the plot in a manner that adequately samples the whole plot and ensures biological integrity of the sample. These can be sampled regularly on a grid if you think there is not much microgeographic variation, or the sample points could be stratified to take account of anything you think might be important.
3. Collect the samples as two depths (1) top 10 cm; (2) 20 cm and below (define). Homogenise all within plot sub-samples to make a pooled sample for each of the two depths. Samples are best mixed in the field in a large ziplock plastic bags from which aliquots can easily be drawn. Sieve if needed.
4. Ensure collection of adequate soil for nucleic acid extraction and contextual data generation from each sampling unit (e.g., 1 kg from each depth).
   * For DNA extraction: fill 50 mL Falcon tube with soil for DNA analysis (leaving 1-2 cm of space at top of tube) and freeze as soon as possible. (see DNA analysis section)
   * Samples must be gently air dried for all remaining workflows
     + For chemical analysis, sample 250 g dry weight of soil (see chemical analysis section)
     + For particle size, 180 g dry weight of soil. (see particle size analysis section)
5. Assign unique sample identifiers to each sample using the last digits after the “/” i.e. 102.100.100/8202 would be called **8202**. Each depth is to have a separate identifier.
6. Collect and record all other local contextual data listed in the ‘Australian Microbiome contextual data template spreadsheet’.
7. Take photos of plot – soil and surrounding environment (example below).



1. Send each sample for DNA analysis, chemical analysis as described in sections 1.1.3 and 1.1.4 (contact the program manager for initiative specific information).
2. Submit **contextual data** into the excel sheet (<https://www.australianmicrobiome.com/protocols/>) to the Project Manager (Sophie Mazard, [smazard@bioplatforms.com](mailto:smazard@bioplatforms.com))
   * 1. DNA extraction
     2. Fill 50 mL Falcon tube with soil for DNA analysis (leaving 1-2 cm of space at top of tube) and freeze as soon as possible.
     3. Contact Dr Leanne McGrath (AGRF) at the address below to organise soil transfer permit paperwork (Biosecurity SA)
     4. Send frozen samples (on dry ice) to AGRF Adelaide for DNA extraction and sequencing analysis.

|  |  |  |
| --- | --- | --- |
| **Facility** | **Address** | **Contact person** |
| **AGRF** | Australian Genome Research Facility Ltd  Plant Genomics Centre  Hartley Grove, Waite campus  University of Adelaide  Urrbrae SA 5064 | Dr Leanne McGrath  Phone: (08) 8313 7148  Email:  [Leanne.McGrath@agrf.org.au](mailto:Leanne.McGrath@agrf.org.au) |

**Note**: Any samples received that are thawed/room temperature will be deemed inappropriate for extraction.

* The soil samples DNA extraction is carried out in triplicate using the MoBio PowerLyzer PowerSoil DNA Isolation Kit (cat#, <http://www.mobio.com/soil-dna-isolation/powerlyzer-powersoil-dna-isolation-kit.html>).
* The DNA is extracted according to Earth Microbiome protocol <http://www.earthmicrobiome.org/emp-standard-protocols/dna-extraction-protocol/>
* A soil sample input of 250 mg is used for each extraction
* For consistency if added DNA output is required an increased number of extractions will be performed and pooled and noted in the metadata (rather than increasing the soil input into an extraction)
* DNA replicate extractions are pooled for further QC and sequencing analysis
  + 1. Chemical analysis and Particle size
    - For chemical analysis send sample 250 g dry weight preferred [200 g minimum required] to CSBP, Perth WA for the ‘Comprehensive Test’.
    - For particle size include an additional 180 g dry weight of soil [http://www.csbp.com.au/docs/default-source/csbp-lab/csbp-lab-methods-1118.pdf](https://www.csbp.com.au/docs/default-source/csbp-lab/csbp-lab-methods-1118.pdf)

|  |  |  |
| --- | --- | --- |
| **Facility** | **Address** | **Contact person** |
| **CSBP** | CSBP Soil and Plant Laboratory  2 Altona Street  Bibra Lake WA 6163 | Phone: (08) 9434 4600  Email: [analysis@csbp.com.au](mailto:analysis@csbp.com.au) |

* 1. SAMPLE: MARINE SEDIMENTS
     1. Sampling, processing and nucleic acid extraction equipment:
* Liquid nitrogen dewar
* 1.8 ml Nunc Cryotube vials
* Cryolabels
* 250 ml glass jars
* 60 ml specimen jars
* Esky & ice
* Black bag
* MoBio Power Soil RNA Isolation Kit (25 preps)
* MoBio Power Soil DNA Isolation Kit (100 preps)
* Molecular grade ethanol (for RNA/DNA clean)
* DEPC water (for RNA clean)
* Nuclease free water (for DNA clean)
* PCR snap tubes (for RNA/DNA clean)
* SuperMagnet
* Axygen 1.5 mL collection tubes
* Various size pipettes and pipette tips
* Mini centrifuge
* Vortex and vortex adaptor (6 place for RNA and 24 place for DNA)
* Vacuum pump and Sartorius filter holder with sump
* 2 mm and 63 µm mesh
  + 1. Sampling
  1. Sample sediment from a depth of 5-10 m.
  2. If you use a Van Veen grab, empty the sediment into clean tray without disturbing the surface layer. Use sterile gloves to collect 1.8 ml of surface layer sediment (top 1 cm) with a microspatula into a pre-labelled cryovial. Label vial according to the instruction given in the spreadsheet “sampling\_schedule”. Microspatula should be sterilised with ethanol wipes between grabs and gloves changed between samples.
  3. If you collect while diving, use a 50 ml Flacon tubes to scoop up the top 1 cm of sediments. Transfer sediment into pre-labelled cryovial.
  4. Drop cryovial into liquid nitrogen.
  5. Homogenise remaining sediments in the tray thoroughly using gloved hands.
  6. Collect ~60ml sediment in a plastic jar (for grain size analyses). Label jar according to the instruction given in the spreadsheet “sampling\_schedule”.
  7. Collect ~150 ml sediment in a glass jar and store in black bag in esky on ice (for TOC analyses). Label jar according to the instruction given in the spreadsheet “sampling\_schedule”.
  8. Cryovials to be stored in -80 °C freezer and RNA should be extracted within 1 week. Glass jars (TOC) and plastic vials (grain sizes) to be stored in -20 °C freezer and analysed within 2 weeks.
     1. Sediment grain size (% fines)

1. Wash sediment (~5 g) from the top 1 cm through filter holder to retain different fractions; gravel (collected on 2 mm mesh), sand (collected on 63um mesh), and fines (collected in sump).
2. Transfer sediments into pre-weighed and labelled specimen jars.
3. Leave for 24 h to allow sediments to settle and then suction off excess water.
4. Transfer to drying oven at 60 °C for 24-48 h or until sediments are completely dry.
5. Re-weigh jars and calculate the weight of each fraction to determine % fines in the sample.
   * 1. Sediment TOC
6. Acidify sediments with 2 ml of 1 M HCl overnight following Hedges & Stern (1984).
7. Analyse sediments with a LECO CN2000 analyser at a combustion temperature of 1050 °C.
   * 1. RNA extraction

Extract RNA from frozen sediment samples (2 g) with MoBio Power Soil RNA Isolation Kit (http://www.mobio.com) following the modified extraction protocol below.

1. Prep a 15 mL **Bead Tube** with 2.5 ml of Bead Solution, 0.25 ml of Solution SR1 and 0.8 ml of Solution SR2.
2. Add 2 g of sediment

*\*Weigh 6 samples, then continue.*

Move to FUME HOOD

1. Add 3.5 mL of phenol:chloroform:isoamyl alcohol to the bead tube, cap tightly and vortex to mix until the biphasic layer disappears.

*\*All waste needs to be kept in fume cupboard and double bagged in a hazardous waste bag.*

*\*Double glove and change gloves any time working outside of the fume cupboard. All pipettes etc. to be ethanol wiped at the end of use.*

*\*Once phenol:chloroform has been added for 6 samples, start Step 4: vortexing as this takes 15min.*

1. Vortex in vortex adaptor at maximum speed for 15 minutes.

\**While these are vortexing, weigh other 6 samples.*

*\*Start preparing 2 sets of collection tubes with 1) 1.5 ml of Solution SR3, 2) 5mL of Solution SR4.*

1. Remove the Bead Tube from the Vortex Adapter and centrifuge at 2500 x *g* for 10 minutes at room temperature.
2. Remove the Bead Tube from the centrifuge and carefully transfer the upper aqueous phase (avoiding the interphase and lower phenol layer) to the 15 ml Collection Tube with 1.5 mL SR3, vortex to mix.

*\*Avoid tilting the tube.*

*\*MAKE SURE NO PHENOL:CHLOROFORM IS TRANSFERRED before adding to SR3. If*

*interphase is punctured, centrifuge again for 10 min.*

1. Incubate at 4 °C for 10 minutes. Wait until last batch has incubated to catch up.
2. Centrifuge at 2500 x *g* for 10 minutes at room temperature. Transfer the supernatant, without disturbing the pellet (if there is one), to the 15 mL Collection Tube with 5 mL SR4.
3. Invert or vortex to mix, and incubate at room temperature for 30 minutes (choose a cool place in the lab, as heat is not good for RNA – desk cupboard).

*\*Time to start thinking about LUNCH :)*

1. Centrifuge at 2500 x *g* for 30 minutes at room temperature.

*\*In the last 10 minutes of centrifuging, RNAseAway and ethanol treat biological safety cabinet. Get waste beaker and lay out kimwipes to UV sterilise. Any racks, pipettes, open pipette tips that will be used can also be sterilized in the cupboard.*

*FROM NOW ON IN STERILE BSC*

1. Decant the supernatant and invert the 15 ml Collection Tube on a paper towel for 5 minutes. Watch that the pellet does not drop onto kimwipe. This may happen if the sample was left to sit for a while after centrifugation.
2. Shake Solution SR5 to mix. Add 1 ml of Solution SR5 to the 15 ml Collection Tube and resuspend the pellet completely by repeatedly pipetting or vortexing to disperse the pellet.

***\*****Depending on the soil type, the pellet may be difficult to resuspend. Resuspension may be aided by placing the tubes in a heat block or water bath at 45 °C for 10 minutes, followed by vortexing. Repeat until the pellet is resuspended.*

1. With new RNAse free gloves, prepare one RNA Capture Column for each RNA Isolation Sample:
2. Remove the cap of a new 15 ml Collection Tube and place the RNA Capture Column inside the 15 ml Collection Tube. The column will hang in the 15 ml Collection Tube.
3. Add 2 ml of Solution SR5 to the RNA Capture Column and allow it to gravity flow through the column and collect in the 15 ml Collection Tube. Allow Solution SR5 to completely flow through the column.

***\*****DO NOT ALLOW THE COLUMN TO DRY OUT PRIOR TO LOADING THE RNA ISOLATION SAMPLE.*

1. Add the RNA Isolation Sample from Step 12 onto the RNA Capture Column and allow it to gravity flow through the column. Collect the flow through in the 15 ml Collection Tube.

*\*When there is time, with new RNAse free gloves, start carefully capping and labelling 2.2 mL* ***collection tubes from the kit****. Add 1 mL of SR4.*

1. Wash the column with 1 ml of Solution SR5. Allow it to gravity flow and collect the flow through in the 15 ml Collection Tube.
2. Transfer the RNA Capture Column to a new 15 ml Collection Tube. Shake Solution SR6 to mix and then add 1 ml of Solution SR6 to the RNA Capture Column to elute the bound RNA into the 15 ml Collection Tube. Allow Solution SR6 to gravity flow into the 15 ml Collection Tube. Do NOT throw out Capture Column.

*\*PREPARE 15mL DNA COLLECTION TUBES. After the SR6 has completely flowed through the column, transfer the Capture Column to the DNA Collection Tubes.*

1. Transfer the eluted RNA to the 2.2 ml Collection Tubes with 1ml of Solution SR4. Invert at least once to mix and incubate at -20 °C for a minimum of 10 minutes.

*\*Start capping and labelling* ***1.5 mL sterile RNase free tubes*** *(from zip-locked Axygen packets). For each sample, label two tube briefly for DNase treatment and then 4 extra tubes for aliquoting final RNA product, i.e. a total of 6 tubes per sample.*

1. Centrifuge the 2.2 ml Collection Tube at 13,000 x *g* for 15 minutes at room temperature to pellet the RNA.

*\*In the last 5 minutes of centrifugation, UV sterilise KimWipes in the BSC.*

1. Decant the supernatant and invert the 2.2 ml Collection Tube onto a paper towel for 10 minutes to air dry the pellet.

*\*Take TurboDNase buffer out of freezer to defrost. Prepare esky of ice.*

1. Resuspend the RNA pellet in 40 μL of Solution SR7 and place samples on ice.
   * 1. TurboDNase treatment for 12 RNA samples

*\*Remove TurboDNase from freezer and thaw on ice.*

*\*TurboDNase, buffer and DNase inactivation reagent should all be kept on ice while working with them and in the dark (covered esky).*

1. Using one of the 1.5 mL sterile tubes, make a DNase mastermix:
2. FOR 12 SAMPLES: To 52 μL of TurboDNase buffer, gently add 13 μL of TurboDNase and very gently mix with your pipette tip.
3. Transfer 5 µL of this DNase mastermix to each tube (there should be a minimum of 5 µL in each tube).
4. Incubate samples for 20 min at 37 °C

*\*Take DNase inactivation reagent out of freezer to thaw.*

1. Add 4.5 µL of resuspended DNase Inactivation Reagent and mix well.
2. Incubate 5 min at room temperature, mixing occasionally. Flick the tube 2–3 times during the incubation period to redisperse the DNase Inactivation Reagent.

*\*If room temperature cools below 22–26 °C, move the tubes to a heat block or oven to control the temperature. Cold environments can reduce the inactivation of the TurboDNase, leaving residual DNase in the RNA sample.*

1. Centrifuge at 10,000 × *g* for 1.5 min and transfer the RNA to an RNase free PCR plate.

*\*Avoid pipetting/disturbing the DNase Inactivation Reagent, it is difficult to remove, even with the magnet clean and interferes with downstream applications.*

* + 1. RNA Clean

1. In PCR plate, pipette from aliquots 100 µL of RNAclean solution. Gently shake before to resuspend beads.
2. Add ~100 μL sample and pipette to mix. Mix gently as fast pipetting can displace sample from well.
3. Leave to stand for a minimum of 10 min to allow strands to bind to magnet beads.
4. Transfer PCR plate to SuperMagnet. Leave for 10 min for rings to form.
5. Carefully remove solution in wells without disturbing the ring of beads.
6. Add 200 µL 70% ethanol and leave for 30 s.
7. Careful not to disturb the bead ring, pipette mix the ethanol to wash and remove from sample. Repeat for a total of 3 washes.
8. After the last wash, make sure all ethanol is pipetted out from wells.
9. Allow to air dry for 9-10 min. Do NOT allow beads to dry!
10. Remove plate from magnet. Add 40 μL of DEPC RNase free water for RNA samples and pipette to elute beads and strands. Elution is rapid.
11. Place plate back on magnet for 10min.
12. Carefully pipette out the sample into 1.5mL sterile tubes labelled earlier and place on ice.
13. If you suspect carryover of magnet beads, place tube back on magnet plate and begin aliquoting 10 µL RNA to 3 x 1.5 mL tubes. With care, a final drop of sample containing magnet beads can be separated and removed.
14. Place 3 x 10 µL RNA aliquots into -80 °C freezer. Keep 1 x 10 µL “dirty” sample on ice to nanodrop.
15. The final RNA should be labelled according to the instruction given in the spreadsheet “sampling\_schedule” and stored at -80 °C.
    * 1. DNA extraction

Extract DNA from frozen sediment samples (1.5 g) with MoBio Power Soil DNA Isolation Kit (http://www.mobio.com) following manufacturer’s recommendation for extraction protocol.

* + 1. DNA clean

1. In PCR plate, pipette from aliquots 100 μL of AMPure solution. Gently shake before to resuspend beads.
2. Add ~100 μL sample and pipette to mix. Mix gently as fast pipetting can displace sample from well.
3. Leave to stand for a minimum of 10 min to allow strands to bind to magnet beads.
4. Transfer PCR plate to SuperMagnet. Leave for 10 min for rings to form.
5. Carefully remove solution in wells without disturbing the ring of beads.
6. Add 200 μL 70% ethanol and leave for 30 s.
7. Careful not to disturb the bead ring, pipette mix the ethanol to wash and remove from sample. Repeat for a total of 3 washes.
8. After the last wash, make sure all ethanol is pipetted out from wells.
9. Allow to air dry for 9-10 min. Do NOT allow beads to dry!
10. Remove plate from magnet. Add 100 μL nuclease free water for DNA and pipette to elute beads and strands. Elution is rapid.
11. Place plate back on magnet for 10 min.
12. Carefully pipette out the sample into 1.5 mL sterile tubes labelled earlier and place on ice.
13. If you suspect carryover of magnet beads, place tube back on magnet plate and being aliquoting 50 μL DNA to remaining 1.5 mL tube. With care, a final drop of sample containing magnet beads can be separated and removed.
14. Place 1 x 50 μL DNA aliquot at -20°C. Keep 1 x 50 μL DNA aliquot “dirty” sample on ice to nanodrop.
15. The final DNA should be labelled according to the instruction given in the spreadsheet “sampling\_schedule” and stored at -80 °C.
    1. SAMPLE: PELAGIC WATERS

The pelagic seawater sampling carried out as part of the IMOS National Reference Station (NRS) microbial sampling is conducted as detailed in the NRS Biogeochemical operations manual (<https://s3-ap-southeast-2.amazonaws.com/content.aodn.org.au/Documents/IMOS/Facilities/national_mooring/IMOS_NRS_BGCManual_LATEST.pdf>) and according to the AMMBI project SOP.

The protocol for the various sampling aspects has been provided in a series of methodological videos that can be found at to following link: <http://imos.org.au/facilities/nationalmooringnetwork/moorings-documentation/bgcwatersamplingvideos/>

* + 1. Sampling, processing and nucleic acid extraction equipment:
* Lysis buffer
  + 200 mM NaH2PO42H2O (monobasic)
  + 200 mM Na2HPO4 (dibasic) MW 142 142 g/1L =1M 5.68 g/200 mL = 200 mM

To make up 200 mL lysis buffer 39 mL 200 mM Na2HPO4

* + 61 mL 200mM Na2HPO4 17.54 g NaCl
  + 2 g CTAB
  + 4 g PVP K30
  + + ddH20 to make up to 200 ml
  + Adjust to pH 7.0 (using NaOH – try couple of mL of 10 M NaOH)
* Lysozyme
* Proteinase K – 20 mg/ml
* From FastDNA™ Spin Kit for Soil (MP Biomedicals):
  + MT buffer
* From PowerWater® Sterivex™ DNA Isolation Kit (MoBio Laboratories, Inc)
  + Columns and sample recovery tubes, 3 ml and 20 ml syringes, 5 ml tubes Buffer
  + ST4 (warmed to 65 °C before use)
  + Buffer ST5 and ST6
* Inlet and outlet caps for Sterivex filters
* Phenol:Chloroform:Isoamyl (25:24:1) (PCI) Chloroform:Isoamyl (24:1) (CI)
* TE buffer
  + 1. Seawater sampling

Sampling procedures are provided in the NRS Biogeochemical operations manual (<https://s3-ap-southeast-2.amazonaws.com/content.aodn.org.au/Documents/IMOS/Facilities/national_mooring/IMOS_NRS_BGCManual_LATEST.pdf>). Samples (2L) are collected at various depths and filtered each month from the 7 sites, Maria Island (MAI), Port Hacking Bay (PHA), North Stradbroke Island (NSI), Kangaroo Island (KAI) Rottness Island (ROT), Darwin (DAR0 and Yongala (YON). Sterivex filters are used; at the end, filters are dried by pushing some air through them. Filters are then stored at -80 °C from collection and transported to CMAR Hobart on dry ice. Received samples are logged and stored at -80 °C until further processing (bottom shelf, in plastic box) .

* + 1. Nucleic acid material extraction

*(Processing limited by capacity of horizontal vortexer, 2 vortexers x 6 filters = 12 filters per extraction set)*

1. weigh 125 mg lysozyme into 50 mL falcon tube and add 25 mL Lysis Buffer to dissolve (lysozyme final conc. 5 mg/ml).
2. remove filters from -80 °C, remove inlet cap and using a pipette add 1.875 ml Lysis buffer (containing 5mg/ml final concentration of lysozyme) and 0.125 ml MT buffer. *(if filters are covered with RNAlater, remove RNAlater with back pressure from 3 ml syringe)*
3. recap the Sterivex filter and attach filter (with inlet end facing out) to the horizontal vortexer, Speed 5-7 for 60 min (turning the filter a couple of times during the h)
4. using 3 ml syringe, draw back plunger and attach to inlet end of filter until pressure builds up – release plunger and buffer in filter should flow into syringe. Divide approx 2 ml of buffer evenly into 2 × 2.0 ml tubes *(do not use the 2 ml collection tubes that come in the PowerWater kit – they don’t tolerate PCI) (may need to use syringe several times to get all buffer out of filter, should be about 0.800-1.00 ml per tube)*
5. in fume hood, add 900 µl PCI to each tube, invert several times, spin down 13000/10 min/RT
6. combine the aqueous phases from both tubes into one 2.0 ml tube (which will be between 1.2 – 1.5 ml), add 20 µl Proteinase K, onto heat block for 2 h at 60 °C
7. in fume hood, add 500 ul CI, spin down 13000/10 min/RT – put aqueous phase into new tube
8. in fume hood, add a further 500 ul CI, spin down 13000/5 min/RT – put aqueous phase into new tube
9. after 2nd spin, take out 1 ml of aqueous phase, add to 5 ml tube
10. add 3 ml of warmed ST4 buffer (65 °C), mix by inversion
11. attach column to barrel of 20 ml syringe and attach to vacuum manifold
12. pour contents of 5 ml tube into barrel while still warm
13. using vacuum, pull contents through the column
14. while keeping column attached to the manifold, remove barrel and add 800 ul ST5 to column
15. using vacuum, pull contents through the column
16. add 800 µl ST6 to column
17. using vacuum, pull contents through the column, then keep on vacuum for 2 mins
18. turn vacuum off, put column into new 2.0 ml tube and let air dry on bench for 10 mins
19. add 80 µl TE to column, incubate at 37 °C for 45 min
20. spin down column and tube at 13000/2 min/RT to elute DNA
21. Quantify 2 µl on the Nanodrop and record concentration and 260:280 ratio in spreadsheet

**Reference**

Davies, C. and Sommerville, E. (Eds.) (2017), National Reference Stations Biogeochemical Operations Manual Version 3.2.1. Integrated Marine Observing System.

DOI: 10.26198/5c4a56f2a8ae3 (<http://dx.doi.org/10.26198/5c4a56f2a8ae3>)

* 1. SAMPLE: COASTAL SEAWATER
     1. Sampling, processing and nucleic acid extraction equipment:
* Listing to be defined
  + 1. Seawater sampling

1. Wash a ~20 L bucket at least five times with seawater from the site. Drop bucket attached to a rope to ~2 m depth and retrieve. If possible, this is best done from a boat. Alternatively, collect water from a depth of ~2 m using a foldable, acid-washed canister.
2. Suck 50 ml of seawater from bucket or canister into a syringe. Wash syringe a few times. Attach a 0.45 µm syringe filter and push a few millilitres into three 50mL Falcon tubes. Wash tubes with filtrate three times. Then fill tubes with 40 ml of filtrate, close tube and keep on ice. Samples are then frozen once returned to the laboratory. (-20 °C). These three tubes are for nutrient analysis and please label them according to the instruction given in the spreadsheet “sampling\_schedule”.
3. Use the 50 ml syringe from step 2 to 5 times filter 50 ml of seawater (from the container/bucket of step 1) through a syringe filter holder containing a 25 mm Whatman GF/F filter. If the filter does not show colour, then filter more water through it. At the end, press air through it to dry filter. Place GF/F filter into pre-labelled cryovial and snap-freeze in liquid N on site. Repeat three times. These filters will be used for chlorophyll analysis.
4. Pipette 1 ml of seawater (from the container/bucket of step 1) in a pre-labelled cryovial with paraformaldehyde (1% v/v final concentration). Invert tube a few times and then snap-freeze in liquid N on site. Repeat three times. Use gloves and dispose of pipette tip into a sealed container for safe disposal back in the lab. These samples will be used for cell counts.
5. For microbe sampling, attach a prefilter to the intake side of a peristaltic pump tubing. The filter can be simply built by cutting a hole in the cap of a 15 Falcon tube and the placing a 100 micron mesh of the threaded part of the lid. Put the lid back on to jam the mesh between the lid and the tube. Cut the bottom of the tube and attach it to the intake side of the peristaltic pump line. Filter ~2 L of seawater separately three times within 2 hours using a peristaltic pump onto three separate 0.22 micron sterivex filter. Use speeds of up to ~120 rpm and platinum cured silicone pump tubing (8.0 mm dia. Internal diameter 3.2 mm e.g. Masterflex L/S 16) or similar.
6. When the 2 L has been filtered continue to run the pump for 1-2min to remove all of the liquid. If the filter gets blocked before, then note the approximate volume filter, disconnect the inlet and try to pump air through the filter for 1-2 min to remove remaining liquid.
7. Ensure that the sterivex is labelled according to the instruction given in the spreadsheet “sampling\_schedule”. Cap both ends of the sterivex filter [alternatively mould some clean blutak around the ends to seal if you have the version with a luer lok on one side only]. Place the filter in a ziplock bag.
8. Freeze immediately at -80 °C.
   * 1. Nucleic acid material extraction
9. For DNA extraction, above a clean surface (e.g. petri dish) crack open the sterivex filter at the inlet end using long-handled pincers (e.g. Oetiker-14100329-Economy-Tool-Standard-Jaw-Pincers)
10. The PES membrane can be carefully removed by scoring along the edges with a pair of clean watchmakers forceps and peeling off its support and before placing the membrane into an extraction tube.
11. Extract using the Mobio Powersoil kit, which includes a beating step with the vortex adaptor.
12. The final eluted DNA should be labelled according to the instruction given in the spreadsheet “sampling\_schedule” and stored at -80 °C.
    * 1. Preliminary protocol for water sampling for metatranscriptomics
13. Use the same filtering set-up as in step 5 to filter ~2 L of seawater separately three times onto three separate 0.22 micron sterivex filter. The filtering should be completed within 20 min of the water sample being collected.
14. When the 2 L has been filtered continue to run the pump for 1-2 min to remove all of the liquid. If the filter gets blocked before, then note the approximate volume filter, disconnect the inlet and try to pump air through the filter for 1-2 min to remove remaining liquid.
15. Ensure that the sterivex is labelled according to the instruction given in the spreadsheet “sampling\_schedule” and append an “RNA” at the end.
16. Cap the sterivex filter at one end and add 1.6 ml of RNAlater (Sigma). Cap the other end and shake well to distribute solution. Place the filter in a ziplock bag.
17. Freeze immediately at -80 °C.
    * 1. Sampling for DIC and TA
18. Use (250 mL)- borosilicate glass bottles with stopper or with screw cap to collect samples. All bottles should be acid washed and well rinsed with seawater.
19. Collect at depth of 2 m. Either fill bottles underwater and seal underwater to avoid excess gas exchange. Alternatively, lower bottle in bucket (see above) and then close under water. If sampling using a Niskin bottles, then attach tubing (make sure there are no gas bubbles in tubing) to tap of Niskin bottle, insert the tubing into the bottom of the sample bottle and begin water flow. Allow the bottle to fill completely and to overflow 3 full volumes.
20. Collect n=3 per site. Labelled bottles according to the instruction given in the spreadsheet “sampling\_schedule”.
21. Leave 1-2cm head-space in sample bottles.
22. Add 100 µl for 250 mL bottle (adjust for volume of container) of HgCl2 using a pipette. Make sure to use a pipette dedicated for toxics. Do not submerge pipette tip into sample.
23. Note: Prepare a saturated solution of mercuric chloride: 1:10 ratio for a saturated HgCl2 solution; e.g. 10 g mercuric chloride salt per 100 mL DIW. Standard volumes used for saturated HgCl2 solutions are 0.02-0.05% of the total sample volume.
24. Invert the sample several times to mix the mercuric chloride thoroughly.
25. Seal the sample bottle with silicon grease around the stopper and a rubber band. If using screw cap seal with parafilm.
26. Store bottles at cold/ambient temperature in the dark.
    1. SAMPLE: SEAGRASS
       1. Sampling, processing and nucleic acid extraction equipment:

* Van Veen grab
* Liquid nitrogen dewar (to be used both, on site and in the lab)
* 1.8 mL Nunc Cryotube vials
* Cryolabels
* Aluminium foil
* Gloves
* Scissors
* 250 mL glass jar
* 100 mL specimen jars
* Esky and ice
* Black bag
* Plastic resealable bags
* MoBio Power Soil DNA Isolation Kit and MoBio Power Soil RNA Isolation Kit
* PureLink RNA Mini Kit and PureLink DNase
* Instrumentation for nutrients analysis of porewater and TOC analysis of sediment (AutoAnalyzer and standard spectrophotometric analysis)
* Vacuum pump and Sartorius filter holder with sump
* 2 mm and 63 mm sieves

First time the site is visited, abundance of *Zostera muelleri* will be measured over 10 randomly thrown 50cmx50cm quadrats.

* + 1. Surficial sediment sampling

NOTE: samples to be collected *immediately before* seagrass samples (for each plant, both surficial sediment and plant tissue will be collected).

1. Pull out a seagrass plant and, out of water, collect surficial sediment, using sterile gloves onto a clean tray.
2. Homogenize sediments with gloved hands and avoid collecting roots/rhizome.
3. Fill a 1.8 mL Nunc cryotube vial up to the top (collect up to 3 cryotube vials/sample). Label tubes according to the scheme outlined in the spreadsheet “sample\_schedule”.
4. Flash freeze cryotube vials in liquid nitrogen, using the dewar brought to the field, and keep the samples on ice for transport to the laboratory.
5. Store samples in -80 °C freezer for further analysis.
6. RNA and DNA should be extracted from sediment within 2 weeks.
7. Collect ~200 mL sediment in a standard glass jar and store in black bag in esky on ice (for nutrients and TOC analyses). Label according to the scheme outlined in the spreadsheet “sample\_schedule”.
8. Store glass jars in -20 °C freezer to be analysed within 2 weeks.
9. Collect ~60 mL sediment in a standard plastic jar (for grain size analyses). Label according to the scheme outlined in the spreadsheet “sample\_schedule”.
10. Store plastic jars at room temperature for further analysis.
    * 1. Seagrass sampling

NOTE: samples to be collected *immediately after surficial* sediment samples (for each plant, both surficial sediment and plant tissue will be collected).

1. Collect seagrass healthy plants, using sterile gloves. Select *only* green, no-damaged plants.
2. Separate *only* the leaves (above ground biomass) from the rest of plant tissues by using sterile scissors (section at the interface between leaves and roots).
3. Discard the seagrass roots/rhizome tissue (below ground biomass), making sure that necrotic/dark plant tissue is *not* collected.
4. Wash seagrass leaves with seawater *directly in the ocean* until all sediment and littler is detached. Be careful not to damage the leaves.
5. Pack the seagrass leaves without sediment into aluminium foil envelopes previously labelled according to the scheme outlined in the spreadsheet “sample\_schedule”, and flash freeze these envelopes in liquid nitrogen (or place in dewar).
6. After return to the laboratory, store aluminium envelops in a labelled box in -80 °C freezer for further analysis.
7. RNA should be extracted from leaves within 2 weeks.
   * 1. Sediment nutrients
8. Collect a 20 mL pore water sample by centrifuging sediment for nutrients analysis (phosphorus, ammonia, nitrate + nitrite).
9. Analyse phosphorus, ammonia, nitrate + nitrite in 20 mL of pore water using AutoAnalyzer and standard spectrophotometric methods (Strickland & Parsons, 1972; Grasshoff, 1976; Oudot & Montel, 1988).
   * 1. Sediment grain size (% fines)
10. Wash sediment (~60 mL) of top 1 cm through sieves to collect three size fractions; gravel (collected on 2 mm mesh), sand (collected on 63 µm mesh) and fines (collected in sump).
11. Transfer sediments into pre-weighed and labelled specimen jars. Label jars according to the scheme outlined in the spreadsheet “sample\_schedule”.
12. Leave for 24 h to allow sediments to settle and then suction off excess water.
13. Transfer to drying oven at 60 °C for 24-48 h or until sediments are completely dry.
14. Re-weigh jars and calculate the weight of each fraction to determine % fines in the sample.
    * 1. Sediment TOC
15. Acidify sediments left (~140 mL) with 60 mL of 1 M HCl overnight following Hedges & Stern (1984).
16. Analyse sediments with a LECO TOC analyzer at a combustion temperature of 1050°C.
    * 1. Sediment DNA and RNA extraction
17. Extract frozen sediment samples (0.25g) with MoBio Power Soil DNA Isolation Kit (http://www.mobio.com) following extraction protocol.
18. Extract frozen sediment samples (up to 2g) with MoBio Power Soil RNA Isolation Kit (http://www.mobio.com) following extraction protocol.
    * 1. Seagrass RNA extraction

Extract RNA from seagrass leaves using PureLink RNA Mini Kit as per the manufacturer’s protocol (Life Technologies) with the addition of On-column PureLink DNase (Life Technologies) to remove contaminating gDNA.

* + 1. RNA clean

1. In PCR plate, pipette from aliquots 100uL of RNAclean solution. *(Gently shake before to resuspend beads)*
2. Add ~50 µl sample and pipette to mix \*Mix gently as fast pipetting can displace sample from well.
3. Leave to stand for a minimum of 10 min to allow strands to bind to magnet beads.
4. Transfer PCR plate to SuperMagnet. Leave for 10 min for rings to form.
5. Carefully remove solution in wells without disturbing the ring of beads.
6. Add 200uL 70% ethanol and leave for 30 s.
7. Careful not to disturb the bead ring, pipette mix the ethanol to wash and remove from sample. Repeat for a total of 3 washes.
8. After the last wash, make sure all ethanol is pipetted out from wells.
9. Allow to air dry for 9-10 min. Do NOT allow beads to dry!
10. Remove plate from magnet. Add 40 μL of DEPC RNase free water for RNA samples and pipette to elute beads and strands. Elution is rapid.
11. Place plate back on magnet for 10min.
12. Carefully pipette out the sample into 1.5 µl sterile tubes prepared earlier and place on ice.
13. If you suspect carryover of magnet beads, place tube back on magnet plate and being aliquoting 10uL RNA to remaining 1.5 µl tubes. With care, a final drop of sample containing magnet beads can be separated and removed.
14. Place 3x 10 µl RNA aliquots into -80 °C freezer. Keep one “dirty” sample on ice to nanodrop. Labelled according to the scheme outlined in the spreadsheet “sample\_schedule”.
    * 1. RNA quality control, transcriptome and metatranscriptome sequencing
15. Check RNA quantity and quality using Nanochip technology (Agilent 2100 Bioanalyzer) according to manufacturer’s instructions (Agilent).
16. Store high-quality RNA samples (integrity number >7) in -80 °C freezer for further sequencing.
17. Send RNA samples by sequencing (5GB per sample).
    1. SAMPLE: SPONGES
       1. Sampling, processing and nucleic acid extraction equipment:

* Listing to be defined
  + 1. General notes on collection:

1. First time the site is visited, sponge species abundance will be measured over 10 random quadrants. This will be repeated as required during the 14 months observation period.
2. Select only sponge individuals that appear healthy and document specimen with a photo.
3. Three different individuals will be collected at each sampling event.
4. Record the following parameter for each individual: size (approximate in cm); covered with detritus (yes/no); any contact with other benthic, sessile organisms (yes, no; if yes, specify); morphological anomalies (discolouration; outgrowth etc.); epiphytic growth (type and coverage)
   * 1. Sampling
5. Cut sponge samples underwater into small pieces (~5 g) using sterile blades. For each sponge specimen (n=3) collect six pieces each covering the entire body of the sponge
6. Transfer pieces into 50 ml Falcon tube. Label containers according to the instruction given in the spreadsheet “sampling\_schedule.
7. On the surface after the dive, use sterile gloves to squeeze water out of three sponge, pieces per sponge, transfer to a fresh 15 ml Falcon tube and plunge into liquid nitrogen (for RNA extraction). Label tubes according to the instruction given in the spreadsheet “sampling\_schedule” and add “R” at the end (for RNA extraction).
8. On the surface, wash three pieces three times in sterile seawater (5 min), transfer to a fresh 15 ml Falcon tube and then plunge into liquid nitrogen (for amplicon sequencing). Label tubes according to the instruction given in the spreadsheet “sampling\_schedule and add “A” at the end (for amplicon analysis).
   * 1. DNA extraction for amplicon sequencing
9. Cut all three piece of each specimen (n=3) into small pieces under liquid nitrogen. Mix the small pieces for each specimen, but don’t mix pieces of the sponge replicates.
10. Extract frozen samples (~0.5 g) for each replicate (n=3) separately with MoBio Power Soil I[solation kit (http://www.mobio.com/soil](http://www.mobio.com/soil-dna-isolation/powersoil-dna-isolation-kit.html))-[dna-isolation/powersoil-dna-isolation-kit.html).](http://www.mobio.com/soil-dna-isolation/powersoil-dna-isolation-kit.html))
11. The final eluted DNA should be labelled according to the instruction given in the spreadsheet “sampling\_schedule” and stored at -80 °C.
    * 1. RNA extraction, host-mRNA removal and rRNA depletion
12. Cut all three pieces of each specimen (n=3) into small pieces under liquid nitrogen. Mix the small pieces for each specimen, but don’t mix pieces of the sponge replicates.
13. Add approximately 0.5 g of material to a tube separately for each sponge replicate.
14. Add 1ml of Trizol directly in the tube (still frozen) and let it defrost while cutting the tissue with small scissors
15. Transfer the liquid (with the crushed tissue) into a tube with beads and beat for 30 s at 5.5 speed
16. Follow the protocol for Trizol extraction (i.e. PureLink RNA Mini Kit) with the DNase I treatment step, elute in 100 µl. (This usually results in 700 – 1000 ng/µl of total RNA by Qubit)
17. Check the samples by agarose gel and Bioanalyzer before continuing
18. Deplete the polyA + sequences with “Poly(A) purist Kit” following the protocol, but discarding the beads and recover the RNA from the supernatant with the PureLink RNA Mini Kit doing another DNase I treatment step
19. Check the samples with universal bacterial 16S rRNA gene primers (to confirm that there is no contaminant DNA)
20. Use this RNA with another round of with “Poly(A) purist Kit” as before (DNA I treatment can be omitted, if previous PCR was negative)
21. Use 5 µg of the previous RNA (PolyA-) to deplete in ribosomal RNA using “RiboZero Bacteria Kit” following the protocol.
22. Remove the beads and precipitate the supernatant with the ethanol method (adding 2 µl Glycogen, 1/10 volume 3M NaOAc, 2.5 volumes of ethanol and incubate overnight at -20 °C).
23. Resuspend the final pellet in water. The final RNA should be labelled according to the instruction given in the spreadsheet “sampling\_schedule” and stored at -80 °C.
    1. SAMPLE: SEAWEED (*Ecklonia* *radiata*)
       1. Sampling, processing and nucleic acid extraction equipment:

* Listing to be defined
  + 1. General note for the collection:

1. First time the site is visited, abundance of *E. radiata* will be measured over 10 randomly thrown 50x50 cm quadrats. This will be repeated as required during the 14 months observation period.
2. Select randomly from a kelp bed only those seaweed individuals that appear healthy and document specimen with a photo.
3. Three different healthy specimens are analysed at each sampling event.
4. Record the following parameter for each specimen: length (approximate in cm); any contact with other benthic, sessile organisms (yes, no; if yes, specify); morphological condition (discolouration; outgrowth, spots etc.); epiphytic growth (type and coverage via categories <10, 10-25, 25-50, 50-75, >75%), photosynthetic efficiency *in situ* (quantum yield using a Diving-PAM fluorometer; if available (see Additional Information below), evidence of grazing (presence/ absence and estimated % of thallus affected via categories <10, 10-25, 25-50, 50-75, >75%)
   * 1. Sampling
5. Collect for each of the three replicate specimens the middle section of a secondary lamina located at approximately the same distance from the meristem (see Additional Information below). Make sure the collected tissue has an area >=30 cm2 (enough for swabbing and PAMing; see below).
6. Enclose samples individually inside press-seal bags *in situ* and bring to the surface. Label the bag according to the instruction given in the spreadsheet “sampling\_schedule”.
7. On the surface, rinse the algal samples with filtered-sterilised seawater. Use a 50 ml syringe with a 0.22 µm filter to run sterilised seawater for about 10-20 second over thallus to remove parts (e.g. detritus) that are not biofilm.
8. Use a sterile cotton tip (head size about 1 cm) and swab an area of approximately 20 cm2 of the lamina gently (i.e. without breaking the algal tissue) for 30 seconds
9. Transfer the tips of the swabs into individual sterile cryo-tubes, label the tubes according to the instruction given in the spreadsheet “sampling\_schedule” and immediately store in a dry-shipper with liquid nitrogen onsite (or -80 °C for long-term).
   * 1. DNA extraction
10. Extract each replicate swab (n=3) separately with the MoBio Power Soil Isolation kit (<http://www.mobio.com/soil-dna-isolation/powersoil-dna-isolation-kit.html>).
11. For this, aseptically transfer swab from cryotube to extraction tube by tipping from one to the other (or where necessary use clean, sterile forceps). Then follow the protocol as outlined in the manual, with the following modifications:
    * Step #5. Instead of using a vortex etc., replace this step by bead beating at full speed (30 Hz) for 10 min.
    * Step #18. Make sure filter is completely dry after this step. If not, repeat.
    * Steps #20-21. Split into 2 steps using 50 μL (rather than 100 μL once). Heat up solution C6 to 50 °C (when you start the whole lot), add 50 μL and leave it for 5 min, centrifuge at room temperature for 30 sec at 10,000 x *g*, and repeat (final volume = 100 μL).
12. The final eluted DNA should be labelled according to the instruction given in the spreadsheet “sampling\_schedule” and stored at -20 °C.
    * 1. Additional information:

*PAM fluorometry measurements in situ*: After bringing the algal tissue to the surface and swabbing

~20 cm2 to sample surface-associated microorganisms, use a ‘leaf clip’ provided with a diving Pulse Amplitude Modulated (diving-PAM) fluorometer (Walz, Germany) to first ‘dark-adapt’ for 15 minutes an area of algal thallus adjacent to where the swabbing was done (i.e. not swabbed) that is not covered by visible epiphytes (note: the leaf clip must be in the ‘closed’ position). A 0.8 s ‘saturating pulse’ of light (> 4500 µmol photons m-2s-1) is delivered to the algal surface *via* a 1.5 mm diameter optical fibre. The same fibre also delivers the ‘measuring light’ (which should be set

< 0.15 µmol photons m-2s-1, below the level required to initiate photosynthesis. Chlorophyll fluorescence (wavelength > 710 nm) will be measured on the Diving-PAM (gain = 4). Minimum dark-adapted fluorescence (F) is determined prior to the saturating flash, while the maximum dark-

adapted fluorescence (FMʹ) will be determined as the fluorescence value during the saturating flash. The difference between FMʹ and F is the variable fluorescence, ΔF. Effective quantum yield (EQY, ΔF/FMʹ) is a measure of photosystem II photochemical efficiency in the dark-adapted algae.

Area of thallus appropriate for sampling and PAM measurements: Because *E. radiata* has a basal meristem, the tissues at the top of the thallus (primary lamina) and the ends of each lateral ‘branch’ are typically older, may be more poorly chemically defended and are thus more likely to show evidence of epibiosis or natural senescence. We wish to sample young, healthy tissue, so avoid those areas. Select a lateral ‘branch’ that is in the lower half of the alga and sample that lateral ‘branch’ on the half closest to the primary lamina.



* 1. SAMPLE: CORAL
     1. Sampling, processing and nucleic acid extraction equipment:
* Listing to be defined
  + 1. General note for the collection:

1. Relative abundance of chosen coral species at each site should be contextualised on each visit by 10 randomly thrown 50x50 cm quadrats.
2. Select several colonies that are of dominant and consistent appearance (colour morph, growth pattern etc.) and document each colony with a photo.
3. Three different healthy specimens are analysed at each sampling event.
4. Record the following parameters for each specimen: size (approximate in cm); any contact with other benthic, sessile organisms (yes, no; if yes, specify); morphological condition (discolouration; outgrowth, spots etc.); epiphytic growth (type and coverage (%)), evidence of grazing (presence/ absence and estimated % affected)
   * 1. Sampling:
5. Collect coral mucus samples of 3 colonies with autoclaved cotton swabs placed in 2 ml Eppendorf tubes (see picture). Invert those tubes and open underwater next to the coral surface to minimize contamination with microbes from the surrounding seawater. Swabs should be gently rolled over the coral surface and then placed back in the tube. Labelled tubes according to the instruction given in the spreadsheet “sampling\_schedule”.
6. Once on the surface, remove excess seawater from the tube without touching the swab and snap freeze sample in liquid Nitrogen.
7. Also remove 3 random samples per coral colony of the same 3 colonies (avoid sampling of the swabbed location). Enclose the 3 random samples separated for each colony inside press-seal bags (or Falcon tubes) with surrounding seawater *in situ* and bring to the surface. Each sample should comprise of a fragment 5-10mm in length or a small branch (species dependent). Labelled tubes according to the instruction given in the spreadsheet “sampling\_schedule”
8. Water is emptied before sample is snap frozen in liquid nitrogen to preserve the mucus layer onsite (liquid N is transported in dry shipper).
9. In the laboratory, remove mucus and tissue with an air gun and 1X PBSE pH 7.4 (137 mM NaCl (8.07 g/L), 2.7 mM KCl (0.201 g/L), 4.3 mM Na2HPO4 (0.611 g/L), 1.4 mM KH2PO4 (0.191 g/L), 10 mM EDTA (20 ml 0.5 M EDTA). Liquefy coral tissue by repeatedly drawing into a 5 ml syringe or using a hand-held homogenizer. Divide samples into 1.5 ml cryo-tubes and centrifuge at 13,000 rpm for 10 minutes. Remove top liquid phase and discard keeping the cell pellet. Freeze samples at -80 °C (or -20 °C).
   * 1. DNA extraction:
10. Extracting DNA with the MoBio PowerPlant Pro DNA Isolation kit using Standard Protocols [(http:// http://www.mobio.com/images/custom/file/protocol/13400.pdf).](http://www.mobio.com/images/custom/file/protocol/13400.pdf)) This kit has the ability to remove many of the inhibitors present in a variety of samples.
11. Store DNA at -80 °C until required.
12. The final eluted DNA should be labelled according to the instruction given in the spreadsheet “sampling\_schedule” and stored at -80 °C.
13. CONTEXTUAL METADATA

Environmental samples that are part of the Australian Microbiome initiative are accompanied by a rich suite of contextual metadata describing the environmental features in which the sample(s) was taken, as well as the physical and chemical characteristics of the environment. For each sample, we require a set of mandatory contextual metadata and strongly encourage the procurement of additional recommended metadata (indicated in italics in the section below). The metadata captured is aligned with international efforts of international community-driven standards (e.g. EnvO environmental ontologies) as recommended by international groups such as the [Genomic Standards Consortium](https://press3.mcs.anl.gov/gensc/).

For each metadata type, there are shared features between the different environment investigated within the initiative, these are presented as overview at the start of each section. Metadata that are specific for an environment are then detailed. The environment type codes are as follow:

* SO: soils
* SE: sediments
* SW: coastal seawater, pelagic seawater and NRS
* HA: host associated
  1. CORE METADATA

The core metadata are absolute mandatory sets of data that must be provided providing the core identifying features of each sample.

|  |  |
| --- | --- |
| **Core metadata for all sample type** | |
| Sample\_ID | Sample\_attribution |
| Date\_sampled [YYYY-MM-DD] | Sample\_descriptor |
| Time\_sampled [hh:mm] | *Citation* |
| Latitude [decimal degrees] | Funding\_source |
| Longitude [decimal degrees] | *Notes (incl. phys/chem measurement method variation)* |
| Depth [m] | Sample\_collection\_device\_method |
| Geo\_loc [country:subregion] | Sample\_material\_processing |
| Sample\_site (location\_description) | Sample\_storage\_method |
|  | DNA\_extraction\_method |

* In addition, for sediments, and coastal water sampling, (host-associated sampling when required):

Coastal\_ID

* In addition, for NRS water sampling:

NRS\_TRIP\_CODE; NRS\_SAMPLE\_CODE; IMOS\_SITE\_CODE

* In addition, for research voyage (pelagic water sampling:

Voyage\_code; Voyage\_survey\_link; Operation/Cast\_ID; Bottle\_number

* 1. ENVIRONMENTAL DESCRIPTORS

The environmental descriptors provide a refined description of the sampled site. The required set of mandatory environmental descriptors (standard font) is listed for each sample type and a list of strongly recommended metadata (in italics). Several descriptors have controlled vocabulary for input as detailed on the contextual metadata pro forma (<https://www.australianmicrobiome.com/protocols/>).

|  |  |
| --- | --- |
| **Environmental descriptor metadata for all sample type** | |
| Environment | Detailed\_land\_use (control vocab 2) |
| Env\_material (control vocab 0) | General\_Env\_feature (control vocab 3) |
| Broad\_land\_use (control vocab 2) | Env\_biome (control vocab 4) |

|  |  |
| --- | --- |
| **SO** | **HA** |
| Horizon (control vocab 1) | Host\_type (see vocab) |
| Elevation [m] | Host\_species |
| *Vegetation\_Total\_cover [%]* | Host\_state |
| *Vegetation\_Dom\_Trees [%]* | Host\_associated\_microbiome\_zone (see vocab) |
| *Vegetation\_Dom\_Shrubs [%]* | *Average\_host\_abundance [% of individuals per m2]* |
| *Vegetation\_Dom\_Grasses [%]* | *Host\_abundance [mean number ind per m2 +/- SE]* |
| *Slope [%]* | *Length [cm]* |
| *Slope\_Aspect [Direction or degrees; NW or 315°]* | *Fouling [%]* |
| *Profile\_Position (control vocab 5)* | *Fouling\_organisms* |
| *Australian\_Soil\_Classification (control vocab 6)* | *Grazing\_number* |
| *FAO\_soil\_classification (control vocab 7)* | *Grazing [%]* |
| *Immediate\_Previous\_Land\_Use (control vocab 2)* | *Bleaching [%]* |
| *Date\_since\_change\_in\_Land\_Use* | *Touching\_organisms* |
| *Crop\_rotation\_1yr\_since\_present* |  |
| *Crop\_rotation\_2yrs\_since\_present* |  |
| *Crop\_rotation\_3yrs\_since\_present* |  |
| *Crop\_rotation\_4yrs\_since\_present* |  |
| *Crop\_rotation\_5yrs\_since\_present* |  |
| *Agrochemical\_Additions* |  |
| *Tillage (control vocab 9)* |  |
| *Fire* |  |
| *Fire\_intensity (if known)* |  |
| *Flooding* |  |
| *Extreme\_Events* |  |

* + 1. Controlled vocabulary descriptors

Env\_material (control vocab 0)

Horizon (control vocab 1) [soil specific]

Broad\_land\_use (control vocab 2)

Detailed\_land\_use (control vocab 2)

Immediate\_Previous\_Land\_Use (control vocab 2) [soil specific]

General\_Env\_feature (control vocab 3)

Env\_biome (control vocab 4)

Profile\_Position (control vocab 5) [soil specific]

Australian\_Soil\_Classification (control vocab 6) [soil specific]

FAO\_soil\_classification (control vocab 7) [soil specific]

Tillage (control vocab 9) [soil specific]

Host\_type (see vocab) [host-associated specific]

Host\_associated\_microbiome\_zone (see vocab) [host-associated specific]

* 1. PHYSICAL DATA

For each sample type, we require a set of mandatory physical metadata (standard font) and strongly encourage the procurement of additional recommended metadata (in italics)

|  |  |  |  |
| --- | --- | --- | --- |
| **SO** | **SED** | **HA** | **SW** |
| Temperature [deg C] | Temperature [deg C] | Temperature [deg C] | Temperature [deg C] |
| *Soil\_moisture [%]* | Water\_depth [m] | *PAM\_fluorometer* | *Secchi\_Depth [m]* |
| *Water\_Holding\_capacity* | %\_fine\_sediment [grain size] | *Light\_intensity\_surface [µmol/m²/s¯¹]* | *Bottom\_depth* |
| *Bulk\_density [g/cm3 or kg/m3]* | *Sedimentation\_rate [g /[cm2 x y]r]* | *Light\_intensity\_bottom [µmol/m²/s¯¹]* | *Pressure\_bottle* |
| *Microbial\_Biomass* | *Bulk\_density [g/cm3 or kg/m3]* |  | *Fluorescence [AU]* |
| Color (control vocab 10) |  |  | *Transmittance [%]* |
| Gravel >2.0mm [%] |  |  | *Conductivity [S/m]* |
| Texture |  |  | *Turbidity* |
| Corse\_Sand 200-2000µm [%] |  |  | *Density [Kg/m^3]* |
| Fine\_Sand 20-200µm [%] |  |  |  |
| Sand [%] |  |  |  |
| Silt 2-20µm [%] |  |  |  |
| Clay <2µm [%] |  |  |  |

* + 1. Controlled vocabulary descriptors

Color (control vocab 10) [soil specific]

* 1. CHEMICAL DATA

For each sample type, we require a set of mandatory physical metadata (standard font) and strongly encourage the procurement of additional recommended metadata (in italics)

|  |  |  |
| --- | --- | --- |
| **SO** | **SE** | **SW** |
| Ammonium\_Nitrogen [mg/Kg] | %\_total\_nitrogen [TN] | Salinity [PSU] |
| Nitrate\_Nitrogen [mg/Kg] | %\_total\_phosphorous [TP] | Nitrate\_Nitrite [μmol/L] |
| Phosphorus\_Colwell [mg/Kg] | %\_total\_carbon [TC] | Phosphate [μmol/L] |
| Potassium\_Colwell [mg/Kg] | *Porewater\_pH* | Ammonium [μmol/L] |
| Sulphur [mg/Kg] | *TOC* | Silicate [μmol/L] |
| Organic\_Carbon [%] | *Metals* | *Total\_CO2 [μmol/kg]* |
| Conductivity [dS/m] | *Acid\_volatile\_sulphides* | *Total\_alkalinity [μmol/kg]* |
| pH\_Level [CaCl2] [pH] | *Organic\_matter\_content [LOI]* | *Chlorophyll\_a [μg/L]* |
| pH\_Level [H2O] [pH] | *Sed\_porewater\_PO43- [μmol/L]* | *Organic\_Fraction [mg/L]* |
| DTPA\_Copper [mg/Kg] | *Sed\_porewater\_H4SiO4 [μmol/L]* | *Inorganic\_Fraction [mg/L]* |
| DTPA\_Iron [mg/Kg] | *Sed\_porewater\_NH4+ [μmol/L]* | *TSS [mg/L]* |
| DTPA\_Manganese [mg/Kg] | *Sed\_porewater\_NO3- [μmol/L]* | *Nitrite [μmol/L]* |
| DTPA\_Zinc [mg/Kg] | *Sed\_porewater\_NO2- [μmol/L]* | *pH\_Level [H2O] [pH]* |
| Exc\_Aluminium [meq/100g] |  | *Oxygen [μmol/L] Lab* |
| Exc\_Calcium [meq/100g] |  | *Oxygen [ml/L] CTD* |
| Exc\_Magnesium [meq/100g] |  | *Microbial\_Biomass* |
| Exc\_Potassium [meq/100g] |  | *Microbial\_abundance [cells/ml]* |
| Exc\_Sodium [meq/100g] |  |  |
| Boron\_Hot\_CaCl2 [mg/Kg] |  |  |

* In addition for soils

*Cation Exchange Capacity; Arsenic; Scandium; Vanadium; Chromium; Cobalt; Nickel; Gallium; Germanium; Arsenic; Selenium; Rubidium; Strontium; Yttrium; Zirconium; Niobium [Columbium]; Molybdenum; Ruthenium; Rhodium; Palladium; Silver; Cadmium; Tin; Antimony; Cesium; Barium; Lanthanum; Cerium; Praseodymium; Neodymium; Samarium; Europium; Gadolinium; Terbium; Dysprosium; Holmium; Erbium; Thulium; Ytterbium; Lutetium; Hafnium; Tantalum; Tungsten or Wolfram; Osmium; Iridium; Platinum; Gold; Lead; Thorium; Uranium*

* In addition for seawater

*ALLO [mg/m3]; ALPHA\_BETA\_CAR [mg/m3]; ANTH [mg/m3]; ASTA [mg/m3]; BETA\_BETA\_CAR [mg/m3]; BETA\_EPI\_CAR [mg/m3]; BUT\_FUCO [mg/m3]; CANTHA [mg/m3]; CPHL\_A [mg/m3]; CPHL\_B [mg/m3]; CPHL\_C1C2 [mg/m3]; CPHL\_C1 [mg/m3]; CPHL\_C2 [mg/m3]; CPHL\_C3 [mg/m3]; CPHLIDE\_A [mg/m3]; DIADCHR [mg/m3]; DIADINO [mg/m3]; DIATO [mg/m3]; DINO [mg/m3]; DV\_CPHL\_A\_and\_CPHL\_A [mg/m3]; DV\_CPHL\_A [mg/m3]; DV\_CPHL\_B\_and\_CPHL\_B [mg/m3]; DV\_CPHL\_B [mg/m3]; ECHIN [mg/m3]; FUCO [mg/m3]; GYRO [mg/m3]; HEX\_FUCO [mg/m3]; KETO\_HEX\_FUCO [mg/m3]; LUT [mg/m3]; LYCO [mg/m3]; MG\_DVP [mg/m3]; NEO [mg/m3]; PERID [mg/m3]; PHIDE\_A [mg/m3]; PHYTIN\_A [mg/m3]; PHYTIN\_B [mg/m3]; PRAS [mg/m3]; PYROPHIDE\_A [mg/m3]; PYROPHYTIN\_A [mg/m3]; VIOLA [mg/m3]; ZEA [mg/m3]*

* 1. ANALYTICAL METHODS
     1. SOIL

The analytical methods used for the soils should be carried out to the same standards and methodology as conducted through the CSBP laboratories <https://www.csbp.com.au/CSBP-Lab>

The methodologies used are described in their available manual <https://www.csbp.com.au/docs/default-source/csbp-lab/csbp-lab-methods-1118.pdf> (April 2019 version) .

Any deviation to the methods should be discussed before submission and noted in the contextual metadata provided.

* + 1. SEDIMENT

To be added

* + 1. SEAWATER

The analytical methods used for the Coatal/Pelagic Seawater should be carried out as per NRS microbial genomics sampling detailed in the IMOS Biogeochemical Operations manual <https://s3-ap-southeast-2.amazonaws.com/content.aodn.org.au/Documents/IMOS/Facilities/national_mooring/IMOS_NRS_BGCManual_LATEST.pdf>

* + 1. HOST ASSOCIATED

To be added

1. SEQUENCING PROCEDURE
   1. SAMPLE SUBMISSION

Submission to the facility: procedures and conditions (e.g. contextual data, sample ID…) will be detailed here shortly.

* 1. AMPLICONS

Add attributions for the protocols and review the currency

* + 1. Sequencing: Bacterial 16S rRNA gene (27F-519R)

The bacterial communities in environmental samples are investigated by amplification of the 16S rRNA gene using the primer set 27F and 519R (single and dual indexed) and paired-end sequencing on the Illumina MiSeq platform.

* + - 1. Primers for amplification: 27F (Lane 1991) and 519R (Lane et al. 1993)

**ILM\_*27F*\_Uv3 –forward primer**

Both un-indexed and indexed versions are used, depending on the need for dual indexing

1. 5' Illumina adapter
2. Barcode (indexed primer only)
3. Forward primer pad
4. Forward primer linker
5. Forward primer (1391f)

Non-indexed: AATGATACGGCGACCACCGAGATCTACAC TATGGCGAGT GA ***AGAGTTTGATCMTGGCTCAG***

Indexed: AATGATACGGCGACCACCGAGATCTACAC XXXXXXXX TATGGCGAGT GA ***AGAGTTTGATCMTGGCTCAG***

**ILM\_*519R*\_NNNN – reverse primer**

1. Reverse complement of 3' Illumina adapter
2. Golay barcode\*
3. Reverse primer pad
4. Reverse primer linker
5. Reverse primer (EukBr)

CAAGCAGAAGACGGCATACGAGAT XXXXXXXXXXXX *AGTCAGTCAG GG* ***GWATTACCGCGGCKGCTG***

\* This primer includes a 12 base Golay barcode as described by Caporaso et al.

* + - 1. Preparation of master mix for amplification (for 1 rxn)

|  |  |  |
| --- | --- | --- |
| **Component** | **Volume (µl)** | **Final concentration** |
| 10x ImmoBuffer (a) | 2.50 | 1 **X** |
| 10 mM dNTP | 0.50 | 200 nM |
| 50mM MgCl2 | 1.25 | 2.5 mM |
| ILM\_27F\_Uv3 (forward) (5 µM) | 2.50 | 500 nM |
| ILM\_519R\_XXXX (5µM) | 2.50 | 500 nM |
| Immolase DNA Polymerase (5U/µL)(a) | 0.20 | 1 Unit |
| H2O | 14.55 | -­‐ |
| Template | 1.0 | -­‐ |
| **Total Volume** | **25** | **-­‐** |

a IMMOLASE™ DNA Polymerase (Bioline, #BIO-­‐21047)

* + - 1. Thermocycler conditions for amplification (for 96 well thermocycler)

|  |  |  |
| --- | --- | --- |
| **Reaction Step** | **Temperature** | **Time (mm:ss)** |
| Activation | 95°C | 10:00 |
| Amplification (35 cycles) | 94°C | 00:30 |
| 55°C | 00:10 |
| 72°C | 00:45 |
| Final Extension | 72°C | 10:00 |

* + - 1. Amplification method

1. Use neat DNA for initial attempt, 1:10 dilution for failed samples (2nd attempt)
2. Amplify samples with conditions outlined above
3. Run amplicons on an agarose gel. Expected band size for 27F/519R is approx. 530 bp.
4. Clean and normalise samples in a one-step process using the SequelPrep Normalisation Plate Kit according to manufacturer instructions (Invitrogen Cat No. A10510-01
5. Combine equivalent volumes of normalised amplification into a single maximum-recovery tube.
6. Perform a double cleanup of the pool using 0.8x beads
7. Perform library QC on the pool using Qubit (concentration) and Tapestation (size). Calculate final molarity of the pool.
   * + 1. Sequencing of the 27F and 519R fragment

* Sequencing Primers Read 1 Primer: ACACTATGGCGAGTGA**AGAGTTTGATCMTGGCTCAG**
* Read 2 Primer: AGTCAGTCAGGG**GWATTACCGCGGCKGCTG**
* Index Primer: CAGCMGCCGCGGTAATWCCCCTGACTGACT

**Sequencing Setup**

1. If required, dilute pool prepared in **step 7** above to **4nM**.

Denature and dilute down to 20 pM according to Ilumina protocol. See *Preparing Libraries for Sequencing on the MiSeq* (part #15039740).

1. Prepare MiSeq Reagent Cartridge. *See MiSeq Reagent Preparation Guide* (part # 15044983).
2. Add custom sequencing primers into reservoirs 12-14. See *Using Custom Primers on the Miseq* (part # 15041638).
3. Load 600 μl of library pool into the MiSeq reagent cartridge in designated reservoir
4. Modify sample sheet to include custom primer’s sequence/indexes (see index sequences in appendix 2)
5. Start sequencing run following *MiSeq System User Guide* (part # 15027617).
   * + 1. References

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J

Lane DJ. (1991) 16S/23S rRNA sequencing, p 115–175. In Stackebrandt E, Goodfellow M (ed), Nucleic acid techniques in bacterial systematics. Wiley, New York, NY.

Lane DJ, et al. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U. S. A. 82:6955–6959.

* + 1. Sequencing: Archaeal 16S (A2F-519R)

The archaeal communities in environmental samples are investigated by amplification of the archaeal 16S rRNA gene (A16S) and paired-end sequencing on the Illumina MiSeq platform. The preparation of archaeal 16S enriched libraries uses a two-stage PCR strategy. The first round of PCR uses locus specific primers with overhang adapters (A2F\_Nex and 519R\_Nex). The locus specific region of the forward primer was based on the A2F primer from Reysenbach et al (1995), which is specific to archaeal targets. The reverse primer has a locus specific region that is universal for prokaryotes. The second round PCR and subsequent steps of library preparation and sequencing follow the “16S Metagenomic Sequencing Library Preparation” guidelines from Illumina.

* + - 1. Primers for amplification: A2F and 519R

**A2F\_Nex PCR Primer Sequence – Forward primer**

1. Forward overhang

2. Locus specific sequence (A2F primer)

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG TTCCGGTTGATCCYGCCGGA

**519R\_Nex – Reverse primer**

1. Reverse overhang

2. Locus specific sequence (16S universal primer)

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GWATTACCGCGGCKGCT

* + - 1. Preparation of master mix for amplification A2F / 519R (for 1 rxn)

|  |  |
| --- | --- |
| **Component** | **Volume (µl)** |
| PCR Grade H2O(a) | 17.55 |
| 10x immoBuffer(b) | 2.50 |
| 50 mM MgCl2(b) | 0.75 |
| 10 mM dNTPs | 0.50 |
| A2F Nex primer (10 µM)(c) | 1.25 |
| 519R Nex primer (10 µM) (c) | 1.25 |
| Immolase DNA polymerase(b) | 0.20 |
| Template DNA | 1.0 |
| **Total Volume** | 25 |

(a) PCR grade water was purchased from MoBio Laboratories (MoBio Labs: Item#17000l 11)

(b) IMMOLASE DNA Polymerase Kit (Cat number: BIO-2146)

(c) Final primer concentration in reaction: 0.5 µM

* + - 1. Thermocycler conditions for amplification (for 96 well thermocycler)

|  |  |  |
| --- | --- | --- |
| **Reaction Step** | **Temperature** | **Time (mm:ss)** |
| Activation | 95°C | 10:00 |
| Amplification (35 cycles) | 95°C | 00:30 |
| 60°C | 00:15 |
| 72°C | 00:50 |
| Final Extension | 72°C | 5:00 |

* + - 1. Amplification method

1. Use undiluted DNA for first attempt, and 1:10 dilution for second attempt (if failed samples).
2. Amplify samples with conditions outlined above.
3. Run amplicons on an agarose gel. The expected band size is roughly 520 bp
4. If there is no band present, repeat PCR using a 1:10 dilution of the sample. Use the concentration of the DNA extract to determine if the DNA should be further diluted or used at higher volumes.
5. Clean amplicons with Agencourt AMPure XP beads, according to manufacturer’s instructions.
6. Perform a second round PCR (Index PCR) following Illumina’s 16S Metagenomic Sequencing Library Preparation, section Index PCR (part #15044223). A half reaction (5µl total rxn volume) can be used.
7. Clean and normalise the PCR products equalPrep Normalisation plates (Invitrogen, A1051001) according to manufacturer instructions.
8. Pool equal volumes of each normalised amplicon.
9. Perform QC on pool using Qubit (concentration) and Tapestation (size) and calculate molarity
   * + 1. Sequencing of the A2F and 519R fragment

**Sequencing Setup**

1. Dilute pool prepared in **step 8** to 4nM.
2. Denature according to IIlumina protocol, with increased PhiX control spike as recommended for low diversity libraries. see *Preparing Libraries for Sequencing on the MiSeq* (part #15039740).
3. Prepare MiSeq Reagent Cartridge (v3 600 cycles). see *MiSeq Reagent Kit v3 Reagent Preparation Guide* (part # 15044983).
4. Load 600 μL of library pool into the MiSeq reagent cartridge in designated sample well.
5. Start sequencing run following *MiSeq System User Guide* (part # 15027617).
   * + 1. References

Reysenbach AL, Pace NR. In: Robb, F.T., Place, A.R. (Eds.), Archaea: A Laboratory Manual Thermophiles. CSHLP, 101l 107 (1995).

* + 1. Sequencing: Eukaryotic 18S V4

The protocol detailed here is designed to amplify the V4 region of the 18S rRNA gene for paired-end 18S community sequencing on the Illumina MiSeq platform. This protocol is based on Illumina’s 16S Metagenomic Sequencing Library Preparation guide and the protocol used by Ocean Sampling Day, modified to amplify the target and add indexed adapter sequences in a single PCR step.

* + - 1. Primers for amplification: V4 region

See section 3.2.3.7 Appendix 1 for full list of primer sequences.

**Forward primer**

1. Field number (space-delimited), description:
2. 5' Illumina adapter
3. Nextera XT i5 index sequence Illumina forward overhang sequence 18S V4 forward

AATGATACGGCGACCACCGAGATCTACAC XXXXXXXX TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CCAGCASCYGCGGTAATTCC

**Reverse primer**

1. Field number (space-delimited), description:
2. Reverse complement of 3' Illumina adapter Nextera XT i7 index sequence
3. Illumina reverse overhang sequence 18S V4 reverse

CAAGCAGAAGACGGCATACGAGAT XXXXXXXX GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG ACTTTCGTTCTTGATYRATGA

* + - 1. Preparation of master mix for amplification V4 region (for 1 rxn)

|  |  |
| --- | --- |
| **Component** | **Volume (µl)** |
| KAPA HiFi Hot Start Readymix (2x) (a) | 12.5 |
| H2O | 9.0 |
| Forward primer (10 µM) | 1.25 |
| Reverse primer (10 µM) | 1.25 |
| Template | 1.0 |
| **Total Volume** | 25 |

(a) Kit code KK2601 or KK2602

* + - 1. Thermocycler conditions for amplification (for 96 well thermocycler)

|  |  |  |
| --- | --- | --- |
| **Reaction Step** | Temperature | Time (mm:ss) |
| Activation | 98°C | 0:30 |
| Amplification (10 cycles) | 98°C | 0:10 |
| 44°C | 0:30 |
| 72°C | 0:15 |
| Amplification (20 cycles) | 98°C | 0:10 |
| 62°C | 0:30 |
| 72°C | 0:15 |
| Final Extension | 72°C | 7:00 |

* + - 1. Amplification method

1. Use undiluted DNA as a first attempt, and 1:10 diluted for repeats/failed reactions
2. Amplify samples with conditions outlined above.
3. Run amplicons on an agarose gel. Expected band size for 18S-V4 is approximately 536bp.
4. Clean and normalize the PCR products using SequalPrep Normalization plates according to manufacturer’s instructions (Invitrogen cat no. A10510-01)
5. Pool equal volumes of each normalised amplicon.
6. Perform QC on pool using Qubit (concentration) and Tapestation (size) and calculate molarity of pool.
   * + 1. Sequencing of the V4 fragment

**Sequencing Setup**

1. Dilute pool prepared in **step 8** to **4nM**.
2. Denature according to IIlumina protocol. See *Preparing Libraries for Sequencing on the MiSeq* (part #15039740).
3. Prepare MiSeq Reagent Cartridge (v2 500-cycles). See *MiSeq Reagent Kit v2 - Reagent Preparation Guide* (part # 15034097)*.*
4. Load 600 μl of library pool into the MiSeq reagent cartridge in designated reservoir
5. Prepare sample sheet to include the appropriate index sequences.
6. Start sequencing run following *MiSeq System User Guide* (part # 15027617).
   * + 1. References

16S Metagenomic Sequencing Library Preparation (Illumina Part # 15044223 Rev. B) available here: http://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry\_documentati on/16s/16s-metagenomic-library-prep-guide-15044223-[b.pdf](http://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf)

LifeWatch Italy Ocean Sampling Day 2014 Protocol – available here: <http://mb3is.megx.net/osd-> files/download?path=/2014/protocols&files=OSD2014\_protocol\_B\_18S\_V4andV9\_Sequencing\_LifeWath\_MoBiLab\_BARI.pdf

* + - 1. Appendix 1 Primer sequences

|  |  |
| --- | --- |
| Primer Name | Primer Sequence |
| 18S-V4f\_S502 | AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S503 | AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S505 | AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S506 | AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S507 | AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S508 | AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S510 | AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S511 | AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S513 | AATGATACGGCGACCACCGAGATCTACACTCGACTAGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S515 | AATGATACGGCGACCACCGAGATCTACACTTCTAGCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S516 | AATGATACGGCGACCACCGAGATCTACACCCTAGAGTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S517 | AATGATACGGCGACCACCGAGATCTACACGCGTAAGATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S518 | AATGATACGGCGACCACCGAGATCTACACCTATTAAGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S520 | AATGATACGGCGACCACCGAGATCTACACAAGGCTATTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S521 | AATGATACGGCGACCACCGAGATCTACACGAGCCTTATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4f\_S522 | AATGATACGGCGACCACCGAGATCTACACTTATGCGATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCASCYGCGGTAATTCC |
| 18S-V4Lr\_N701 | CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N702 | CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N703 | CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N704 | CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N705 | CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N706 | CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N707 | CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N710 | CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N711 | CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N712 | CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N714 | CAAGCAGAAGACGGCATACGAGATTCATGAGCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N715 | CAAGCAGAAGACGGCATACGAGATCCTGAGATGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N716 | CAAGCAGAAGACGGCATACGAGATTAGCGAGTGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N718 | CAAGCAGAAGACGGCATACGAGATGTAGCTCCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N719 | CAAGCAGAAGACGGCATACGAGATTACTACGCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N720 | CAAGCAGAAGACGGCATACGAGATAGGCTCCGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N721 | CAAGCAGAAGACGGCATACGAGATGCAGCGTAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N722 | CAAGCAGAAGACGGCATACGAGATCTGCGCATGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N723 | CAAGCAGAAGACGGCATACGAGATGAGCGCTAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N724 | CAAGCAGAAGACGGCATACGAGATCGCTCAGTGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N726 | CAAGCAGAAGACGGCATACGAGATGTCTTAGGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N727 | CAAGCAGAAGACGGCATACGAGATACTGATCGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N728 | CAAGCAGAAGACGGCATACGAGATTAGCTGCAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |
| 18S-V4Lr\_N729 | CAAGCAGAAGACGGCATACGAGATGACGTCGAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTCGTTCTTGATYRATGA |

* + 1. Sequencing: Eukaryotic 18S V9

The 18S protocol detailed here is designed to amplify eukaryotes broadly with a focus on microbial eukaryotic lineages. The primers are based on those of Amaral-­‐Zettler et al 2009 and designed to be used with the Illumina platform. The protocol is based on that used by the Earth Microbiome Project (EMP), found here: http://www.earthmicrobiome.org/emp-­‐standard-­‐protocols/18s/

* + - 1. Primers for amplification: 18S V9 region (1391F and EukBR)

The primer sequences in this protocol are always listed in the 5’ -­‐> 3’ orientation. See Section 4 for more information on ordering, concentration, and resuspension. Primer constructs designed by Laura Wegener Parfrey.

**ILM\_Euk\_1391f PCR Primer Sequence – Forward primer**

Field number (space-­‐delimited), description:

1. 5' Illumina adapter

2. Forward primer pad

3. Forward primer linker

4. Forward primer (1391f)

AATGATACGGCGACCACCGAGATCTACAC TATCGCCGTT CG GTACACACCGCCCGTC

**ILM\_EukBr PCR primer sequence – Reverse primer, barcoded**

Each sequence contains different 12 base Golay barcode as described by Caporaso et al.

1. Reverse complement of 3' Illumina adapter

2. Golay barcode

3. Reverse primer pad

4. Reverse primer linker

5. Reverse primer (EukBr)

CAAGCAGAAGACGGCATACGAGAT XXXXXXXXXXXX AGTCAGTCAG CA TGATCCTTCTGCAGGTTCACCTAC

Full list of primer sequences is provided in Section 9.1.2.7 Appendix 1.

* + - 1. Preparation of master mix for amplification 18S V9 region (for 1 rxn)

|  |  |
| --- | --- |
| **Component** | **Volume (µl)** |
| PCR Grade H2O (a) | 13.0 |
| 5 Primer Hot MM (b) | 10.0 |
| Forward primer (10µM) (c) | 0.5 |
| Reverse primer (10µM) (c) | 0.5 |
| Template DNA | 1.0 |
| **Total reaction volume** | 25.0 |

(a) PCR grade water was purchased from MoBio Laboratories (MoBio Labs: Item#17000-­‐11)

(b) 5 PRIME HotMasterMix (5 PRIME: Item# 2200410)

(c) Final primer concentration of mastermix: 0.2 µM

* + - 1. Thermocycler conditions for amplification (for 96 well thermocycler)

|  |  |  |
| --- | --- | --- |
| **Reaction step** | **Temperature** | **Time (mm:ss)** |
| Activation | 94°C | 3:00 |
| Amplification (35 cycles) | 94°C | 00:45 |
| 57°C | 01:00 |
| 72°C | 01:30 |
| Final Extension | 72°C | 10:00 |

* + - 1. Amplification method

1. Dilute DNA 1:10.
2. Amplify samples with conditions outlined above.
3. Run amplicons on an agarose gel. Expected band size for 1391F/EukBR is roughly 200 bp.
4. If there is no band present, repeat PCR using either the undiluted DNA or a 1:100 dilution. Use the concentration of the DNA extract to determine if the DNA should be further diluted or used neat.
5. Clean and normalise the PCR products. For this step, both Agencourt AMPure XP bead clean-up and SequalPrep Normalisation plates (Invitrogen, A1051001) are acceptable. If using Agencourt AMPure XP beads:
6. Perform a bead clean-­‐up following manufacturer’s instructions.
7. Quantify amplicon yields. Acceptable methods include Picogreen (see manufacturers protocol; Invitrogen Item #P11496) or the high sensitivity Quant-­‐iT™ DNA Assay Kit (Life Technologies, Q-­‐33120).
8. Run clean amplicons on Agilent Bioanalyzer, LabChip GX, or 2200 TapeStation to get accurate sizing information.
9. Normalise amplicons by diluting to 10nM with variable volume of buffer (EB buffer, or 10mM Tris pH 8.5) and a set volume of the amplicons.
10. Pool equal volumes of each normalised amplicon.
11. If using SequalPrep Normalisation plates:
12. Perform a plate normalisation following manufacturer’s instructions.
13. Pool equal volumes of each normalised amplicon.
14. Run clean amplicon pool on Agilent Bioanalyzer, LabChip GX, or 2200 TapeStation to get accurate sizing information.
15. Optional: If spurious bands were present on gel (in step 3), or in the clean amplicons (step 5c) or amplicon pool (step 5h), a portion of the final amplicon pool can go through a gel extraction or a second round of bead clean up. A lower bead to template ratio will get rid of small bands.
16. Measure concentration of the final clean pool using Qubit or picogreen.
    * + 1. Sequencing of the 1391F and EukBR 18S V9 fragment

**Sequencing Primers**

**ILM\_Euk\_R1: Read 1 Sequencing Primer**

Field description (space - delimited):

1. Forward primer pad

2. Forward primer linker

3. Forward primer

TATCGCCGTT CG GTACACACCGCCCGTC

**ILM\_Euk\_R2: Read 2 Sequencing Primer**

Field description (space - delimited):

1. Reverse primer pad

2. Reverse primer linker

3. Reverse primer

AGTCAGTCAG CA TGATCCTTCTGCAGGTTCACCTAC

**ILM\_Euk\_INDEX: Index Read Sequencing Primer**

Field description (space-­‐delimited):

1. Reverse complement of reverse primer

2. Reverse complement of reverse primer linker

3. Reverse complement of reverse primer pad

GTAGGTGAACCTGCAGAAGGATCA TG CTGACTGACT

**Sequencing Setup**

1. Dilute pool prepared in step 2.4.5 to 4nM.
2. Denature according to Illumina protocol, with increased PhiX control spike-in as recommended for low diversity libraries. See Preparing Libraries for Sequencing on the MiSeq (part #15039740).
3. Prepare MiSeq Reagent Cartridge (v2 300 - cycles). See MiSeq Reagent Kit v2 - Reagent Preparation Guide (part # 15034097).
4. Using an extra-long pipette tip set to 75 μL, add 3.4 µL of Read1 sequencing primer (100 μM) into well 12 of the MiSeq Reagent Cartridge and mix 10 times. Repeat adding the Index Primer into well 13 and the Read2 sequencing primer into well 14.
5. Load 600 μL of library pool into the MiSeq reagent cartridge in designated sample well.
6. Modify sample sheet to include the custom index sequences (see index sequences in Appendix 2).
7. Start sequencing run following MiSeq System User Guide (part # 15027617).
   * + 1. Primer ordering and resuspension: tips and getting started (From EMP protocol)

**Resuspension of primers** is a critical step and must be done with Ultra pure water under sterile conditions. It is recommended that resuspension be done in a hood.

• Stock plates are generally 100 µM and are aliquoted and diluted to 10 µM for use.

• When making stock plates it is a good idea to make multiple replicate plates so that there are back up plates of primers.

• The primer sequences in this protocol are always listed in the 5’ -­‐> 3’ orientation. This is the orientation that should be used for ordering.

• We use the standard desalting cleanup option.

• If you are using a robot the volume in the plates should be 30 µL per well or greater. 15µL per well is the absolute minimum.

**Ordering barcoded primers**

• Contact IDT or your primer provider to get a quote prior to ordering.

• Barcoded primers are ordered in 96 well plates at the 100 nMole scale.

• Make sure to select the option of receive the same amount per well of each primer (do not get full yield). This is generally the 8 nMole / well option.

• Select the rows option for load scheme: [Rows [A01,A02,A03]]

• If you will be archiving plates of primers for later use consider selecting the option for replicate plates.

* + - 1. References

Amaral-­‐Zettler, LA, EA McCliment, HW Ducklow, SM Huse. 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-­‐subunit ribosomal RNA genes. PLoS ONE 4:e6372.

Caporaso JG, Lauber CL, Walters WA, Berg-­‐Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-­‐high-­‐throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012 Aug; 6(8): 1621–1624.

* + - 1. Appendix 1 Primer sequences

Primer Name Primer Sequence

ILM\_Euk\_R1 TATCGCCGTTCGGTACACACCGCCCGTC

ILM\_Euk\_R2 AGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC

ILM\_Euk\_INDEX GTAGGTGAACCTGCAGAAGGATCATGCTGACTGACT

ILM\_Euk\_1391F\_U AATGATACGGCGACCACCGAGATCTACACTATCGCCGTTCGGTACACACCGCCCGTC

ILM\_EukBR\_0097 CAAGCAGAAGACGGCATACGAGATTACCGCTTCTTCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0098 CAAGCAGAAGACGGCATACGAGATTGTGCGATAACAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0099 CAAGCAGAAGACGGCATACGAGATGATTATCGACGAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0100 CAAGCAGAAGACGGCATACGAGATGCCTAGCCCAATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0101 CAAGCAGAAGACGGCATACGAGATGATGTATGTGGTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0102 CAAGCAGAAGACGGCATACGAGATACTCCTTGTGTTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0103 CAAGCAGAAGACGGCATACGAGATGTCACGGACATTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0104 CAAGCAGAAGACGGCATACGAGATGCGAGCGAAGTAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0105 CAAGCAGAAGACGGCATACGAGATATCTACCGAAGCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0106 CAAGCAGAAGACGGCATACGAGATACTTGGTGTAAGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0107 CAAGCAGAAGACGGCATACGAGATTCTTGGAGGTCAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0108 CAAGCAGAAGACGGCATACGAGATTCACCTCCTTGTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0109 CAAGCAGAAGACGGCATACGAGATGCACACCTGATAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0110 CAAGCAGAAGACGGCATACGAGATGCGACAATTACAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0111 CAAGCAGAAGACGGCATACGAGATTCATGCTCCATTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0112 CAAGCAGAAGACGGCATACGAGATAGCTGTCAAGCTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0113 CAAGCAGAAGACGGCATACGAGATGAGAGCAACAGAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0114 CAAGCAGAAGACGGCATACGAGATTACTCGGGAACTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0115 CAAGCAGAAGACGGCATACGAGATCGTGCTTAGGCTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0116 CAAGCAGAAGACGGCATACGAGATTACCGAAGGTATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0117 CAAGCAGAAGACGGCATACGAGATCACTCATCATTCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0118 CAAGCAGAAGACGGCATACGAGATGTATTTCGGACGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0119 CAAGCAGAAGACGGCATACGAGATTATCTATCCTGCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0120 CAAGCAGAAGACGGCATACGAGATTTGCCAAGAGTCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0121 CAAGCAGAAGACGGCATACGAGATAGTAGCGGAAGAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0122 CAAGCAGAAGACGGCATACGAGATGCAATTAGGTACAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0123 CAAGCAGAAGACGGCATACGAGATCATACCGTGAGTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0124 CAAGCAGAAGACGGCATACGAGATATGTGTGTAGACAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0125 CAAGCAGAAGACGGCATACGAGATCCTGCGAAGTATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0126 CAAGCAGAAGACGGCATACGAGATTTCTCTCGACATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0127 CAAGCAGAAGACGGCATACGAGATGCTCTCCGTAGAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0128 CAAGCAGAAGACGGCATACGAGATGTTAAGCTGACCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0129 CAAGCAGAAGACGGCATACGAGATATGCCATGCCGTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0130 CAAGCAGAAGACGGCATACGAGATGACATTGTCACGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0131 CAAGCAGAAGACGGCATACGAGATGCCAACAACCATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0132 CAAGCAGAAGACGGCATACGAGATATCAGTACTAGGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0133 CAAGCAGAAGACGGCATACGAGATTCCTCGAGCGATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0134 CAAGCAGAAGACGGCATACGAGATACCCAAGCGTTAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0135 CAAGCAGAAGACGGCATACGAGATTGCAGCAAGATTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0136 CAAGCAGAAGACGGCATACGAGATAGCAACATTGCAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0137 CAAGCAGAAGACGGCATACGAGATGATGTGGTGTTAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0138 CAAGCAGAAGACGGCATACGAGATCAGAAATGTGTCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0139 CAAGCAGAAGACGGCATACGAGATGTAGAGGTAGAGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0140 CAAGCAGAAGACGGCATACGAGATCGTGATCCGCTAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0141 CAAGCAGAAGACGGCATACGAGATGGTTATTTGGCGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0142 CAAGCAGAAGACGGCATACGAGATGGATCGTAATACAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0143 CAAGCAGAAGACGGCATACGAGATGCATAGCATCAAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0144 CAAGCAGAAGACGGCATACGAGATGTGTTAGATGTGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0145 CAAGCAGAAGACGGCATACGAGATTTAGAGCCATGCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0146 CAAGCAGAAGACGGCATACGAGATTGAACCCTATGGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0147 CAAGCAGAAGACGGCATACGAGATAGAGTCTTGCCAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0148 CAAGCAGAAGACGGCATACGAGATACAACACTCCGAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0149 CAAGCAGAAGACGGCATACGAGATCGATGCTGTTGAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0150 CAAGCAGAAGACGGCATACGAGATACGACTGCATAAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0151 CAAGCAGAAGACGGCATACGAGATACGCGAACTAATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0152 CAAGCAGAAGACGGCATACGAGATAGCTATGTATGGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0153 CAAGCAGAAGACGGCATACGAGATACGGGTCATCATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0154 CAAGCAGAAGACGGCATACGAGATGAAACATCCCACAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0155 CAAGCAGAAGACGGCATACGAGATCGTACTCTCGAGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0156 CAAGCAGAAGACGGCATACGAGATTCAGTTCTCGTTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0157 CAAGCAGAAGACGGCATACGAGATTCGTGCGTGTTGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0158 CAAGCAGAAGACGGCATACGAGATGTTATCGCATGGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0159 CAAGCAGAAGACGGCATACGAGATGATCACGAGAGGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0160 CAAGCAGAAGACGGCATACGAGATGTAAATTCAGGCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0161 CAAGCAGAAGACGGCATACGAGATAGTGTTTCGGACAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0162 CAAGCAGAAGACGGCATACGAGATACACGCGGTTTAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0163 CAAGCAGAAGACGGCATACGAGATTGGCAAATCTAGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0164 CAAGCAGAAGACGGCATACGAGATCACCTTACCTTAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0165 CAAGCAGAAGACGGCATACGAGATTTAACCTTCCTGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0166 CAAGCAGAAGACGGCATACGAGATTGCCGTATGCCAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0167 CAAGCAGAAGACGGCATACGAGATCGTGACAATAGTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0168 CAAGCAGAAGACGGCATACGAGATCGCTACAACTCGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0169 CAAGCAGAAGACGGCATACGAGATTTAAGACAGTCGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0170 CAAGCAGAAGACGGCATACGAGATTCTGCACTGAGCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0171 CAAGCAGAAGACGGCATACGAGATCGCAGATTAGTAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0172 CAAGCAGAAGACGGCATACGAGATTGGGTCCCACATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0173 CAAGCAGAAGACGGCATACGAGATCACTGGTGCATAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0174 CAAGCAGAAGACGGCATACGAGATAACGTAGGCTCTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0175 CAAGCAGAAGACGGCATACGAGATAGTTGTAGTCCGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0176 CAAGCAGAAGACGGCATACGAGATTCGTCAAACCCGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0177 CAAGCAGAAGACGGCATACGAGATTAATCGGTGCCAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0178 CAAGCAGAAGACGGCATACGAGATTTGATCCGGTAGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0179 CAAGCAGAAGACGGCATACGAGATCGGGTGTTTGCTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0180 CAAGCAGAAGACGGCATACGAGATTTGACCGCGGTTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0181 CAAGCAGAAGACGGCATACGAGATGTGCAACCAATCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0182 CAAGCAGAAGACGGCATACGAGATGCTTGAGCTTGAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0183 CAAGCAGAAGACGGCATACGAGATCGCTGTGGATTAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0184 CAAGCAGAAGACGGCATACGAGATCTGTCAGTGACCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0185 CAAGCAGAAGACGGCATACGAGATACGATTCGAGTCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0186 CAAGCAGAAGACGGCATACGAGATGGTTCGGTCCATAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0187 CAAGCAGAAGACGGCATACGAGATCTGATCCATCTTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0188 CAAGCAGAAGACGGCATACGAGATTATGTGCCGGCTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0190 CAAGCAGAAGACGGCATACGAGATTGTAAGACTTGGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0191 CAAGCAGAAGACGGCATACGAGATCGGATCTAGTGTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0192 CAAGCAGAAGACGGCATACGAGATCGATCTTCGAGCAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC ILM\_EukBR\_0193 CAAGCAGAAGACGGCATACGAGATGTCGAATTTGCGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC

* + - 1. Appendix 2 Index sequences for sample sheet

I7\_Index\_ID index I7\_Index\_ID index I7\_Index\_ID index

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ILM\_EukBR\_0097 | GAAGAAGCGGTA | ILM\_EukBR\_0098 | TGTTATCGCACA | ILM\_EukBR\_0099 | TCGTCGATAATC |
| ILM\_EukBR\_0100 | ATTGGGCTAGGC | ILM\_EukBR\_0101 | ACCACATACATC | ILM\_EukBR\_0102 | AACACAAGGAGT |
| ILM\_EukBR\_0103 | AATGTCCGTGAC | ILM\_EukBR\_0104 | TACTTCGCTCGC | ILM\_EukBR\_0105 | GCTTCGGTAGAT |
| ILM\_EukBR\_0106 | CTTACACCAAGT | ILM\_EukBR\_0107 | TGACCTCCAAGA | ILM\_EukBR\_0108 | ACAAGGAGGTGA |
| ILM\_EukBR\_0109 | TATCAGGTGTGC | ILM\_EukBR\_0110 | TGTAATTGTCGC | ILM\_EukBR\_0111 | AATGGAGCATGA |
| ILM\_EukBR\_0112 | AGCTTGACAGCT | ILM\_EukBR\_0113 | TCTGTTGCTCTC | ILM\_EukBR\_0114 | AGTTCCCGAGTA |
| ILM\_EukBR\_0115 | AGCCTAAGCACG | ILM\_EukBR\_0116 | ATACCTTCGGTA | ILM\_EukBR\_0117 | GAATGATGAGTG |
| ILM\_EukBR\_0118 | CGTCCGAAATAC | ILM\_EukBR\_0119 | GCAGGATAGATA | ILM\_EukBR\_0120 | GACTCTTGGCAA |
| ILM\_EukBR\_0121 | TCTTCCGCTACT | ILM\_EukBR\_0122 | GTACCTAATTGC | ILM\_EukBR\_0123 | ACTCACGGTATG |
| ILM\_EukBR\_0124 | GTCTACACACAT | ILM\_EukBR\_0125 | ATACTTCGCAGG | ILM\_EukBR\_0126 | ATGTCGAGAGAA |
| ILM\_EukBR\_0127 | TCTACGGAGAGC | ILM\_EukBR\_0128 | GGTCAGCTTAAC | ILM\_EukBR\_0129 | ACGGCATGGCAT |
| ILM\_EukBR\_0130 | CGTGACAATGTC | ILM\_EukBR\_0131 | ATGGTTGTTGGC | ILM\_EukBR\_0132 | CCTAGTACTGAT |
| ILM\_EukBR\_0133 | ATCGCTCGAGGA | ILM\_EukBR\_0134 | TAACGCTTGGGT | ILM\_EukBR\_0135 | AATCTTGCTGCA |
| ILM\_EukBR\_0136 | TGCAATGTTGCT | ILM\_EukBR\_0137 | TAACACCACATC | ILM\_EukBR\_0138 | GACACATTTCTG |
| ILM\_EukBR\_0139 | CTCTACCTCTAC | ILM\_EukBR\_0140 | TAGCGGATCACG | ILM\_EukBR\_0141 | CGCCAAATAACC |
| ILM\_EukBR\_0142 | GTATTACGATCC | ILM\_EukBR\_0143 | TTGATGCTATGC | ILM\_EukBR\_0144 | CACATCTAACAC |
| ILM\_EukBR\_0145 | GCATGGCTCTAA | ILM\_EukBR\_0146 | CCATAGGGTTCA | ILM\_EukBR\_0147 | TGGCAAGACTCT |
| ILM\_EukBR\_0148 | TCGGAGTGTTGT | ILM\_EukBR\_0149 | TCAACAGCATCG | ILM\_EukBR\_0150 | TTATGCAGTCGT |
| ILM\_EukBR\_0151 | ATTAGTTCGCGT | ILM\_EukBR\_0152 | CCATACATAGCT | ILM\_EukBR\_0153 | ATGATGACCCGT |
| ILM\_EukBR\_0154 | GTGGGATGTTTC | ILM\_EukBR\_0155 | CTCGAGAGTACG | ILM\_EukBR\_0156 | AACGAGAACTGA |
| ILM\_EukBR\_0157 | CAACACGCACGA | ILM\_EukBR\_0158 | CCATGCGATAAC | ILM\_EukBR\_0159 | CCTCTCGTGATC |
| ILM\_EukBR\_0160 | GCCTGAATTTAC | ILM\_EukBR\_0161 | GTCCGAAACACT | ILM\_EukBR\_0162 | TAAACCGCGTGT |
| ILM\_EukBR\_0163 | CTAGATTTGCCA | ILM\_EukBR\_0164 | TAAGGTAAGGTG | ILM\_EukBR\_0165 | CAGGAAGGTTAA |
| ILM\_EukBR\_0166 | TGGCATACGGCA | ILM\_EukBR\_0167 | ACTATTGTCACG | ILM\_EukBR\_0168 | CGAGTTGTAGCG |
| ILM\_EukBR\_0169 | CGACTGTCTTAA | ILM\_EukBR\_0170 | GCTCAGTGCAGA | ILM\_EukBR\_0171 | TACTAATCTGCG |
| ILM\_EukBR\_0172 | ATGTGGGACCCA | ILM\_EukBR\_0173 | TATGCACCAGTG | ILM\_EukBR\_0174 | AGAGCCTACGTT |
| ILM\_EukBR\_0175 | CGGACTACAACT | ILM\_EukBR\_0176 | CGGGTTTGACGA | ILM\_EukBR\_0177 | TGGCACCGATTA |
| ILM\_EukBR\_0178 | CTACCGGATCAA | ILM\_EukBR\_0179 | AGCAAACACCCG | ILM\_EukBR\_0180 | AACCGCGGTCAA |
| ILM\_EukBR\_0181 | GATTGGTTGCAC | ILM\_EukBR\_0182 | TCAAGCTCAAGC | ILM\_EukBR\_0183 | TAATCCACAGCG |
| ILM\_EukBR\_0184 | GGTCACTGACAG | ILM\_EukBR\_0185 | GACTCGAATCGT | ILM\_EukBR\_0186 | ATGGACCGAACC |
| ILM\_EukBR\_0187 | AAGATGGATCAG | ILM\_EukBR\_0188 | AGCCGGCACATA | ILM\_EukBR\_0190 | CCAAGTCTTACA |
| ILM\_EukBR\_0191 | ACACTAGATCCG | ILM\_EukBR\_0192 | GCTCGAAGATCG | ILM\_EukBR\_0193 | CGCAAATTCGAC |

* + 1. Sequencing: Fungal ITS (ITS1F and ITS4)

The protocol detailed here is designed to investigate the fungal diversity in environmental samples through amplification of the ITS1F and ITS4 region of the fungal ITS for paired-end sequencing on the Illumina MiSeq platform.

* + - 1. Primers for amplification: ITS1F and ITS4 region

**ILM\_ITS1F\_Uv2 – Forward primer**

Field number (space-­‐delimited), description:

1. 5' Illumina adapter

2. Forward primer pad

3. Forward primer linker

4. Forward primer (ITS1F)

AATGATACGGCGACCACCGAGATCTACAC TGTCCGGCTT CG CTTGGTCATTTAGAGGAAGTAA

**ILM\_ITS4Rv2\_00NN – Reverse primer**

Each sequence contains different 12 base Golay barcode as described by Caporaso et al.

1. Reverse complement of 3' Illumina adapter

2. Golay barcode

3. Reverse primer pad

4. Reverse primer linker

5. Reverse primer (ITS4)

CAAGCAGAAGACGGCATACGAGAT XXXXXXXXXXXX AGTCCGTCCG GA TCCTCCGCTTATTGATATGC

* + - 1. Preparation of master mix for amplification ITS1F and ITS4 region (for 1 rxn)

|  |  |  |
| --- | --- | --- |
| **Component** | **Volume (µL)** | **Final Conc.** |
| 10x ImmoBuffer (a) | 2.50 | 1 X |
| 10 mM dNTP | 0.50 | 200 nM |
| 50mM MgCl2 | 1.25 | 2.5 mM |
| ILM\_ITS1F\_Uv2 (5µM) | 2.50 | 500 nM |
| ILM\_ITS4Rv2 \_XXXX (5µM) | 2.50 | 500 nM |
| Immolase DNA Polymerase (5U/µL) (a) | 0.20 | 1 Unit |
| H2O | 14.55 | -­‐ |
| Template | 1.0 | -­‐ |
| **Total Volume** | 25 | -­‐ |

(a) Immolase DNA Polymerase (Bioline, #BIO-­‐21047)

* + - 1. Thermocycler conditions for amplification (for 96 well thermocycler)

|  |  |  |
| --- | --- | --- |
|  | Temperature | Time (mm:ss) |
| Activation | 94°C | 10:00 |
| Amplification (35 cycles) | 94°C | 00:30 |
| 55°C | 01:00 |
| 72°C | 01:00 |
| Final Extension | 72°C | 10:00 |

* + - 1. Amplification method

1. Optional: Dilute DNA 1:10.
2. Amplify samples with conditions outlined above.
3. Run amplicons on an agarose gel. Expected band size for ITS1F / ITS4 fragment is approximately 850 bp.
4. If there is no band present, repeat PCR using either the undiluted DNA or a 1:100 dilution. Use the concentration of the DNA extract to determine if the DNA should be further diluted or used neat.
5. Clean and normalise the PCR products. For this step, both Agencourt AMPure XP bead clean-up and SequalPrep Normalisation plates (Invitrogen, A1051001) are acceptable. If using Agencourt AMPure XP beads:
6. Perform a bead clean-­‐up following manufacturer’s instructions.
7. Quantify amplicon yields. Acceptable methods include Picogreen (see manufacturers protocol; Invitrogen Item #P11496) or the high sensitivity Quant‐iT™ DNA Assay Kit (Life Technologies, Q-33120).
8. Run clean amplicons on Agilent Bioanalyzer, LabChip GX, or 2200 TapeStation to get accurate sizing information.
9. Normalise amplicons by diluting to 10 nM with variable volume of buffer (EB buffer, or 10 mM Tris pH 8.5) and a set volume of the amplicons.
10. Pool equal volumes of each normalised amplicon.
11. If using SequalPrep Normalisation plates:
12. Perform a plate normalisation following manufacturer’s instructions.
13. Pool equal volumes of each normalized amplicon.
14. Run clean amplicon pool on Agilent Bioanalyzer, LabChip GX, or 2200 TapeStation to get accurate sizing information.
15. Optional: If spurious bands were present on gel (in step 3), or in the clean amplicons (step 5c) or amplicon pool (step 5h), a portion of the final amplicon pool can go through a gel extraction or a second round of bead clean up. A lower bead to template ratio will get rid of small bands.
16. Measure concentration of the final clean pool using Qubit or picogreen.
    * + 1. Sequencing of the ITS1F and ITS4 fragment

**Sequencing Primers**

**Read 1 Primer (ILM\_ITS\_R1v2)**

Field description (space-­‐delimited):

1. Forward primer pad

2. Forward primer linker

3. Forward primer

ACACTGTCCGGCTT CG CTTGGTCATTTAGAGGAAGTAA

**Read 2 Primer (ILM\_ITS\_R2v2)**

Field description (space-­‐delimited):

1. Reverse primer pad

2. Reverse primer linker

3. Reverse primer

AGTCCGTCCG GA TCCTCCGCTTATTGATATGC

**Index Read Primer (ILM\_ITS\_INDEXv3)**

Field description (space-­‐delimited):

1. Reverse complement of reverse primer

2. Reverse complement of reverse primer linker

3. Reverse complement of reverse primer pad

GCATATCAATAAGCGGAGGA TC CGGACGGACT

**Sequencing Setup**

1. Dilute pool prepared in step 2.4.5 to 4nM.
2. Denature according to Ilumina protocol, with increased PhiX control spike-­‐in as recommended for low diversity libraries. See Preparing Libraries for Sequencing on the MiSeq (part #15039740).
3. Prepare MiSeq Reagent Cartridge (v3 600-­‐cycles). See MiSeq Reagent Kit v3 -­‐ Reagent Preparation Guide (part # 15044983).
4. Using an extra-long pipette tip set to 75 μL, add 3.4 µL of Read1 sequencing primer (100 μM) into well 12 of the MiSeq Reagent Cartridge and mix 10 times. Repeat adding the Index Primer into well 13 and the Read2 sequencing primer into well 14.
5. Load 600 μL of library pool into the MiSeq reagent cartridge in designated sample well.
6. Modify sample sheet to include the custom index sequences.
7. Start sequencing run following MiSeq System User Guide (part # 15027617).
   * + 1. References

Caporaso JG, Lauber CL, Walters WA, Berg-­‐Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-­‐high-­‐throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J

Gardes, M., and T. D. Bruns. 1993. ITS primers with enhanced specificity for basidiomycetes -­‐ application to the identification of mycorrhizae and rusts. Mol. Ecol. 2: 113-­‐118

White, T. J., T. Bruns, S. Lee, and J. W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-­‐322 In: PCR Protocols: A Guide to Methods and Applications, eds. Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White. Academic Press, Inc., New York.

* 1. METAGENOMICS

**NOTE: this protocol was established for the Marine Microbes project in 2016, an updated protocol is being prepared)**

This protocol describes the procedure for preparing metagenomic libraries from marine samples to be sequenced on the Illumina HiSeq 2500 platform. The library preparation follows Illumina’s Nextera XT DNA Library Prep Reference Guide (Document # 15031942 v01, January 2016) except for the bead-based normalisation and pooling steps.

* + 1. Quality check of the input DNA

Verify the quality of the extracted genomic DNA using the NanoDrop (Thermo Scientific) or DropSense 16 (Trinean, formerly known as Xpose). The concentration should be verified with the Qubit or PicoGreen assay (Invitrogen).

* + 1. Library preparation

1. Dilute 1ng of DNA to a concentration of 0.2ng/ul.
2. DNA tagmentation follows Illumina’s Nextera XT DNA Library Prep Reference Guide (Document # 15031942 v01, January 2016).
3. Library amplification follows the protocol and includes 12 cycles of PCR to enrich for fragments with correctly added indexed adapters.
4. Clean up the amplified libraries using AMPure XP or AxyPrep Mag PCR clean-up beads according to the protocol.
   * 1. Quality check of the libraries
5. Verify the size of the libraries on the LabChip GXII (Perkin Elmer) or the TapeStation (Agilent).
6. Measure the concentration of the libraries using Qubit, PicoGreen, or qPCR.
7. Normalise libraries to 3 nM and pool the libraries manually, do not do follow bead-based normalisation and pooling protocol detailed in the sections of the Prep Guide entitled “Normalize Libraries” and “Pool Libraries”.
   * 1. Sequencing
8. Following Illumina’s NovaSeq 6000 System Guide (Document # 1000000019358 v11, February 2019), NovaSeq 6000 SP Reagent Kit (500 cycles) reagent cartridge .
9. Dilute the library pool to 1.75 nM and denature according to the NovaSeq 6000 System Guide.
10. Start the NovaSeq 6000 sequencing run according to the system guide referenced above.
11. After sequencing is complete, perform the base-calling and demultiplexing using bcl2fastq Conversion Software.
    1. METATRANSCRITPTOMICS

**NOTE: this protocol was established for the Marine Microbes project in 2016, an updated protocol is being prepared)**

**NOTE: Illumina discontinued the Ribo-Zero rRNA removal kits for Bacteria in early 2019. Illumina do plan to relaunch this product in the latter half of 2019**

This protocol describes the procedure for preparing metatranscriptomic libraries form samples to be sequenced on the Illumina HiSeq 2500 platform. The library preparation follows the Ribo-Zero rRNA Removal Kit Reference Guide (Document # 15066012 v02, August 2016) and TruSeq Stranded mRNA Sample Preparation Guide (Document # 15031047 Rev. E, October 2013) with some protocol variations.

* + 1. Quality check of the input DNA

Verify the quality of the extracted RNA using the Bioanalyzer RNA 6000 Nano or RNA 6000 Pico Kit.

* + - 1. Library preparation

Note: The Ribo-Zero rRNA Removal kits support rRNA depletion from 1-5 µg of total RNA. Probes from the Bacteria and Pant Ribo-Zero kits can be combined (50:50 ratio) to simultaneously deplete both bacterial and plant rRNA. In this instance the inputs of RNA should be halved (i.e. 500 ng -2.5 ng of total RNA). The inputs for the sediment RNA sample maybe much lower than these recommended inputs.

1. Bring ≤1 µg of total RNA to a final volume of 28µl with RNase-free water.
2. Perform rRNA depletion following the Ribo-Zero Kit Reference guide
3. Perform clean-up of rRNA depleted supernatant using Agencourt RNAClean XP Kit according to the Ribo-Zero Kit Reference Guide. Elute samples in 8.5 µl of RNase-free water
4. Perform TruSeq Stranded mRNA-seq library preparation according the TruSeq Stranded mRNA Sample Preparation Guide, skipping the purification of poly(A) RNAs.

Add 13 µl of Fragment, Prime and Finish Mix to 5 µl of rRNA depleted rRNA sample and follow the library prep protocol from the Incubate RFP step (page 20)

1. At the Enrich DNA Fragments step (page 38-41) perform 13-15 cycles of PCR

For inputs of 1 µg use 13 cycles, for inputs < 200 ng use 15 cycles

* + - 1. Quality check of the libraries

1. Asses the size of the RNA-seq libraries via electrophoresis using the Agilent TapeStation TapeScreen DNA 1000 Assay or similar (Agilent Bioanalyzer, Perkin Elmer LabChip GXII)
2. Quantify libraries using qPCR (KAPA Library Quantification Kits for Illumina or similar)
3. Normalise the libraries to 3 nM and pool libraries for sequence following the library prep protocol (Normalise and Pool Libraries, page 44).
   * + 1. Sequencing
4. Following Illumina’s NovaSeq 6000 System Guide (Document # 1000000019358 v11, February 2019), thaw the appropriate NovaSeq 6000 SP Reagent Kit reagent cartridge .
5. Dilute the library pool to 1.75 nM and denature according to the NovaSeq 6000 System Guide.
6. Start the NovaSeq 6000 sequencing run according to the system guide referenced above.
7. After sequencing is complete, perform the base-calling and demultiplexing using bcl2fastq Conversion Software.
8. BIOINFORMATIC AND ANALYTICS WORKFLOW

Data generated as part of the Australian Microbiome initiative activities are publicly accessible (with free registration) on the Bioplatforms Data portal (<https://data.bioplatforms.com/organization/about/australian-microbiome>). The Bioinformatic and analytic workflow are accessible through the GitHub page for the initiative: <https://github.com/AusMicrobiome>

The data portal give access to the “raw” sequence data as well as processed data for the microbial diversity surveys through the portal service itself and the BioRDC analytics linked to the Galaxy Australia platform (currently only available for zOTU, to be expanded to [gen]omics).

* 1. AMPLICONS

Note: All amplicons have the primers removed in the end product through various methods:

* Bacterial and archaeal 16S do not have primer sequence in the raw sequence. Removed by the sequencing facility before provision of the data
* Eukaryotic 18Sv4: sequences are hard trimmed 20 nt from the 5’ of R1, 21 nt from the 5’ end of R2
* Eukaryotic 18sV9: primers are removed as part of the R1/R2 merging process
* Fungal ITS: primers are removed when ITS regions are extracted from the flanking mRNA
  + - 1. Bacterial and Archaeal 16S rRNA gene Amplicon Analysis Workflow

This section outlines the workflow required to analyse 16S rRNA amplicon sequences for Bacteria (27f - 519r) and Archaea (A2f-519r), to produce Amplicon Sequence Variant (ASV) information for the Australian Microbiome database

Analysis is completed on a per sequencing run (sequencing plate) basis. The workflow consists of the following stages:

**A] Sequence preparation and merging**

1. Merge paired end reads (non-merged reads are discarded)
2. Convert fastq file format to fasta file format
3. Add sampleID, runID and "sample=" information to the sequence headers
4. Concatenate all sequences per sequencing run into a single file

**B] Sequence analysis**

1. Quality screening, zotu/ASV calling and sequence mapping
2. Classify and remove flipped sequences
3. Replace arbitrary ASV ID's with the sequence itself in the table index

**C] Following analysis, all data is combined to give a single dataset, by the following steps:**

1. Merge tables into a single table
2. Remove controls from the abundance tables, to create separate samples and control datasets
3. Make a fasta from the abundance table
4. Classify Sequences

**Software Required**

The following software is used in the steps below:

1. FLASH2 (Magoc and Salzberg, 2011)
2. Mothur (Schloss, et al., 2009)
3. USEARCH (Edgar 2010)
4. Seqtk (<https://github.com/lh3/seqtk>)
5. Fastx (<http://hannonlab.cshl.edu/fastx_toolkit/>)
6. Python3

**A] Sequence preparation and merging**

**Merge the paired end reads**

Paired end reads are merged using FLASH2 (Magoc and Salzberg, 2011). Flash is run with the following arguments,

--min-overlap=30 --max-overlap=250

Following merging, the merge quality is manually checked by examining the FLASH log file for the percentage of reads that were merged. Plates with low merge rates (< 70%) are manually checked to see if the alignments can be improved.

Unmerged reads are discarded

**File naming**

File names are formatted to the following format:

sampleID\_plateID.fasta

The sampleID of the mock communities and negative controls on each plate are standardised to the following naming convention (all listed for completeness):

* 16S Bacteria mock communities: **BACMOCK**
* A16 Archaeal mock communities: **ARCMOCK**
* 18Sv4 Eukaryote mock communities: **EUKMOCKv4**
* 18Sv9 Eukaryote mock communities: **EUKMOCKv9**
* ITS Fungal mock communities: **FUNMOCK**
* Negative control: **NEG**
* STAN: **STAN**

fastq files are then converted to fasta using *seqtk*

**Add sample and file name to the sequence headers**

Sample identifiers are added to the header of each sequence for downstream processing. As each fasta is now named SampleID\_plateID.fasta we simply add the file name with the extension to the sequence header. At this stage we also add any other information and delimiters that downstream programs will likely require. For Usearch we add "sample=" and ";", for Qiime we add "\_".

#!/bin/bash

#this initiates the loop for all files with extension .fasta in the list generated by ls -l

for file in `ls -1 \*.fasta`

do

#tell me which file you're working on

echo working on $file

#cut the appropriate string from the name of the file.

VAR1=`echo $file | cut -d "." -f 1`

#tell me what you've cut and called VAR1

echo VAR1 is $VAR1

#use VAR1 to add the sample name to the sequence id. Substitute > for >VAR1\_ in all files. The variable (VAR1) needs to be in quotation marks to be passed from bash to perl, the perl script needs to be embedded in the bash script with the ` ` marks

`perl -pi -e 's/\>/\>sample='$VAR1';\_16S\_/' $file;`

done

**Identify unique sequences per sample**

An abundance table of all unique sequences in each sample on the plate is then prepared. Unique sequences are first identified using fastx using the following Usearch command:

usearch64 -fastx\_uniques SampleID\_plateID.fasta -fastaout SampleID\_plateID\_uniques.fasta -sizeout

***Unique sequences per sample (non-Denoised or quality filtered) are provided as a data output and are available from the Australian Microbiome Website.***

**Concatenate all sequences per sequencing run into a single file**

All of the files for each plate are concatenated into a single file for zotu/ASV calling. The resulting file name is standardized to the format: plateID\_all\_16S.fasta

**B] Sequence analysis**

**Quality screening, zotu/ASV calling and sequence mapping**

* The first step removes sequences that are too short or too long, have ambiguous bases, or have more than 12 homopolymers.

mothur "#set.dir(modifynames=F);summary.seqs(fasta=plateID\_all.fasta, processors=10); screen.seqs(fasta=current, minlength=432, maxlength=520, maxambig=0, maxhomop=12, processors=10); summary.seqs()"

* Reads are dereplicated

usearch64 -fastx\_uniques plateID\_all\_16S.good.fasta -fastaout plateID\_all\_16S.good\_uniques.fasta -sizeout

* Sort unique reads by abundance

usearch64 -sortbysize plateID\_all\_16S.good\_uniques.fasta -fastaout plateID\_all\_16S.good\_sorted\_uniques.fasta -sizeout

* ASV/zotus are called by UNOISE3, from sequences that have => 4 representatives

usearch64 -unoise3 plateID\_all\_16S.good\_sorted\_uniques.fasta -zotus plateID\_all\_16S.good\_sorted\_uniques\_zotus.fasta -ampout plateID\_all\_16S.good\_sorted\_uniques\_ampout.fasta -tabbedout plateID\_all\_16S.good\_sorted\_uniques\_unoise3.txt -minsize 4

* Map reads to zotus to generate abundances

Reads are mapped against the zotus using USEARCH. Note the termination conditions on the mapping run

(-maxaccepts 0), this seems to be required to ensure the best match is found and to produce consistent results when adding multiple plates together, as we do later.

usearch64 -otutab plateID\_all\_16S.fasta -zotus plateID\_all\_16S.good\_sorted\_uniques\_zotus\_renamed.fasta -otutabout plateID\_all\_16S.good\_sorted\_uniques\_zotutab\_MA0.txt -mapout plateID\_all\_16S.good\_sorted\_uniques\_zmap\_MA0.txt -maxaccepts 0 -threads X

**Classify and remove flipped sequences**

A final QC step is performed to remove likely erroneous sequences. The ASVs are classified, with those that do not align to the 16S database in the correct orientation being removed. Those that need to be "flipped" to a new orientation are likely errors, since we know the reads should be in 27f - 519r orientation.

**1. classify the seqs against 16S database**

"#set.dir(modifynames=F); classify.seqs(fasta=plateID\_all\_16S.good\_sorted\_uniques\_zotutab\_relabelled\_MA0.fasta, reference=gg\_13\_8\_99.fasta, taxonomy=gg\_13\_8\_99.gg.tax, cutoff=60, probs=FALSE)"

**2. use the \*acnos.flip list to remove sequences from the table**

The above produces the final abundance table, with ASV sequences as index for each sequencing run. These tables are then combined as below to produce a single dataset for the Australian Microbiome.

**Replace table indexes with ASV sequence**

After denoising and mapping the ASV tables have an arbitrary zOTU number as the index, we replace this with the sequence that the arbitrary number represents. These sequences are unique strings and allow tables to be merged etc. downstream easily. They also negate the need to maintain a dictionary of ASVs and the sequences they represent.

**C] Prepare the single dataset**

Now we have an ASV abundance table for each plate, with ASV's as row and SampleID\_runID as column headers. To prepare this data for ingest into the AMD database the following steps are carried out:

1. Each table is converted from short to long format (from rectangular to 3 column, with columns ASV, sampleID, Abundance)
2. All of these 3 column tables are concatenated into a single table
3. Controls and samples are split into separate tables
4. sequencing run ID's are removed from the column headers and any sample sequenced more than once has the abundances from these runs summed to give a single abundance per sample
5. A fasta file of unique ASV's is created from all ASV's in this final table
6. Sequences are classified.

**Classify the sequences**

Sequences are classified to provide taxonomies relative to the Silva database (Quast et al., 2013; Yilmaz et al 2014; Glöckner 2017) as below:

mothur "#classify.seqs(fasta=seqs.fasta, reference=SILVA132.ng.fasta, taxonomy=SILVA132.tax, cutoff=60, probs=FALSE, processors=X)"

**References**

Magoc, T. and Salzberg, S. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27(21): 2957-2963.

Schloss, P.D., et al., (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75(23):7537-7541.

Edgar, R.C., (2010), Search and clustering orders of magnitude faster than BLAST, Bioinformatics 26(19): 2460-2461.

Seqtk available at: <https://github.com/lh3/seqtk> (last accessed 23 Jan 2019).

Fastx available at: (<http://hannonlab.cshl.edu/fastx_toolkit/>) (last accessed 23 Jan 2019).

DeSantis T.Z., Hugenholtz P., Larsen N., Rojas M., Brodie E.L., Keller K., Huber T., Dalevi D., Hu P., Andersen G.L.(2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl. Environ. Microbiol. 72(7): 5069-5072; DOI: 10.1128/AEM.03006-05

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl. Acids Res. 41 (D1): D590-D596.

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucl. Acids Res. 42:D643-D648

Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, Bruns G, Yarza P, Peplies J, Westram R, Ludwig W (2017) 25 years of serving the community with ribosomal RNA gene reference databases and tools. J. Biotechnol.

* + 1. Eukaryotic 18S rRNA gene Amplicon Analysis Workflow

This section outlines the workflow required to analyse 18S rRNA gene amplicon sequences to produce Amplicon Sequence Variant information for the Australian microbiome database.

This workflow covers both 18S variable region 4 (18Sv4) amplified by the 18S\_V4f/18S\_V4r primer set and 18S variable region 9 (18Sv9) amplified by the ILM\_Euk\_1391f/ILM\_EukBr primer set

Analysis is completed on a per sequencing run basis. The workflow consists of the following stages:

**A] Sequence preparation and merging**

1. Trim and merge paired end reads (non-merged reads are discarded)
2. Convert fastq file format to fasta file format
3. Add sampleID, runID and "sample=" information to the sequence headers
4. Concatenate all sequences per sequencing run into a single file

**B] Sequence analysis**

1. Quality screening, zotu/ASV calling and sequence mapping
2. Classify and remove flipped sequences
3. Replace arbitrary ASV ID's with the sequence itself in the table index

**C] Following analysis, all data is combined to give a single dataset, by the following steps:**

1. Merge tables into a single table
2. Remove controls from the abundance tables, to create separate samples and control datasets
3. Make a fasta from the abundance table
4. Classify Sequences

**Software Required**

The following software is used in the steps below:

1. FLASH2 (Magoc and Salzberg, 2011) (for 18Sv9) or FLASH2 (Magoc and Salzberg, 2011)(for 18Sv4)
2. Mothur (Schloss, et al., 2009)
3. USEARCH (Edgar 2010)
4. Seqtk (<https://github.com/lh3/seqtk>)
5. Fastx (<http://hannonlab.cshl.edu/fastx_toolkit/>)
6. Python3

**A] Sequence preparation and merging**

**Merge the paired end reads**

For 18Sv4 amplicons, primer removal is performed using seqTk by hard trimming 20 nucleotides from the 5' end of R1 sequences and 21 nucleotides from the 5' end of its respective R2 paired end read. Sequences are merged using FLASH2 (Magoc and Salzberg, 2011). FLASH2 is run with the following arguments,

--min-overlap=50 --max-overlap=160 --allow-outies

18Sv9 paired end reads are merged using FLASH2 with the following arguments

--min-overlap=50 --max-overlap=120 --allow-outies

Following merging, the merge quality is manually checked by examining the FLASH log file for the percentage of reads that were merged. Plates with low merge rates (< 70%) are manually checked to see if the alignments can be improved.

Unmerged reads are discarded

**Rename files**

File names are formatted to the following format:

sampleID\_plateID.fasta

The sampleID of the mock communities and negative controls on each plate are standardised to the following naming convention (all listed for completeness):

* 16S Bacteria mock communities: **BACMOCK**
* A16 Archaeal mock communities: **ARCMOCK**
* 18Sv4 Eukaryote mock communities: **EUKMOCKv4**
* 18Sv9 Eukaryote mock communities: **EUKMOCKv9**
* ITS Fungal mock communities: **FUNMOCK**
* Negative control: **NEG**
* STAN: **STAN**

fastq files are then converted to fasta using *seqtk*

**Add sample and file name to the sequence headers**

Sample identifiers are added to the header of each sequence for downstream processing. As each fasta is now named SampleID\_plateID.fasta we simply add the file name with the extension to the sequence header. At this stage we also add any other information and delimiters that downstream programs will likely require. For Usearch we add "sample=" and ";", for Qiime we add "\_".

#!/bin/bash

#this initiates the loop for all files with extension .fasta in the list generated by ls -l

for file in `ls -1 \*.fasta`

do

#tell me which file you're working on

echo working on $file

#cut the appropriate string from the name of the file.

VAR1=`echo $file | cut -d "." -f 1`

#tell me what you've cut and called VAR1

echo VAR1 is $VAR1

#use VAR1 to add the sample name to the sequence id. Substitute > for >VAR1\_ in all files. The variable (VAR1) needs to be in quotation marks to be passed from bash to perl, the perl script needs to be embedded in the bash script with the ` ` marks

`perl -pi -e 's/\>/\>sample='$VAR1';\_18SV\_/' $file;`

done

**Identify unique sequences per sample**

An abundance table of all unique sequences in each sample on the plate is then prepared. Unique sequences are first identified using using fastx using the following Usearch command:

usearch64 -fastx\_uniques SampleID\_plateID.fasta -fastaout SampleID\_plateID\_uniques.fasta -sizeout

***Unique sequences per sample (non-Denoised or quality filtered) are provided as a data output and are available from the Australian Microbiome Website.***

**Concatenate all sequences per sequencing run into a single file**

All of the files for each plate are concatenated into a single file zotu/ASV calling. The resulting file name is standardised to the format: plateID\_all\_18SVn.fasta.

Where 18SVn represents the 18S variable region being analysed (18SV4 or 18SV9)

**B] Sequence analysis**

**Quality screening, zotu/ASV calling and sequence mapping**

* The first step performs some quality control on the sequences, with each amplicon having different parameters:

For 18SV4, the first step removes sequences that are too short, have ambiguous bases, or have more than 12 homopolymers.

mothur "#set.dir(modifynames=F);summary.seqs(fasta=plateID\_all\_18SVn.fasta, processors=10); screen.seqs(fasta=current, minlength=300, maxambig=0, maxhomop=12, processors=10); summary.seqs()"

For 18SV9 the first step removes sequences have ambiguous bases, or have more than 12 homopolymers.

mothur "#set.dir(modifynames=F);summary.seqs(fasta=plateID\_all\_18SVn.fasta, processors=10); screen.seqs(fasta=current, maxambig=0, maxhomop=12, processors=10); summary.seqs()"

* Next reads are dereplicated

usearch64 -fastx\_uniques plateID\_all\_16S.good.fasta -fastaout plateID\_all\_18SVn.good\_uniques.fasta -sizeout

* Sort unique reads by abundance

usearch64 -sortbysize plateID\_all\_18SVn.good\_uniques.fasta -fastaout plateID\_all\_16S.good\_sorted\_uniques.fasta -sizeout

* ASV/Zotus are called by UNOISE3, from sequences that have => 4 representatives

usearch64 -unoise3 plateID\_all\_18SVn.good\_sorted\_uniques.fasta -zotus plateID\_all\_18SVn.good\_sorted\_uniques\_zotus.fasta -ampout plateID\_all\_18SVn.good\_sorted\_uniques\_ampout.fasta -tabbedout plateID\_all\_18SVn.good\_sorted\_uniques\_unoise3.txt -minsize 4

* Map reads to zotus to generate abundances

Reads are mapped against the zotus using USEARCH. Note the termination conditions on the mapping run

(-maxaccepts 0), this seems to be required to ensure the best match is found and to produce consistent results when adding multiple plates together, as we do later.

usearch64 -otutab plateID\_all\_18SVn.fasta -zotus plateID\_all\_18SVn.good\_sorted\_uniques\_zotus\_renamed.fasta -otutabout plateID\_all\_18SVn.good\_sorted\_uniques\_zotutab\_MA0.txt -mapout plateID\_all\_18SVn.good\_sorted\_uniques\_zmap\_MA0.txt -maxaccepts 0 -threads X

**Classify and remove flipped sequences**

A final QC step is performed to remove likely erroneous sequences. The ASVs are classified, with those that do not align to the 18S Silva database (Quast et al., 2013; Yilmaz et al 2014; Glöckner 2017) in the correct orientation being removed. Those that need to be "flipped" to a new orientation are likely errors, since we know the reads should be in correct orientation for their respective primer sets 18S\_V4f/18S\_V4r or ILM\_Euk\_1391f/ILM\_EukBr.

**1. classify the seqs against 18S database**

"#set.dir(modifynames=F); classify.seqs(fasta=plateID\_all\_18SVn.good\_sorted\_uniques\_zotutab\_relabelled\_MA0.fasta, reference=//OSM/CBR/OA\_AMD/amd-work/DBs/GG\_mothur/gg\_13\_8\_99.fasta, taxonomy=//OSM/CBR/OA\_AMD/amd-work/DBs/GG\_mothur/gg\_13\_8\_99.gg.tax, cutoff=60, probs=FALSE, processors=5)"

**2. use the \*acnos.flip list to remove sequences from the table**

The above produces the final abundance table, with sequences as index for each sequencing run. These tables are then combined as below to produce a single dataset

**Replace table indexes with ASV sequence**

Currently the tables have an arbitrary OTU number as the index, replace this with the sequence that the arbitrary number represents. These sequences are unique strings and allow tables to be merged etc. downstream easily. They also negate the need to maintain a dictionary of ASVs and the sequences they represent.

**C] Prepare the single dataset**

Now we have an ASV abundance table for each plate, with ASV's as row and SampleID\_runID as column headers. To prepare this data for ingest into the AMD database the following steps are carried out:

1. Each table is converted from short to long format (from rectangular to 3 column, with columns ASV, sampleID, Abundance)
2. All of these 3 column tables are concatenated into a single table
3. Controls and samples are split into separate tables
4. sequencing run ID's are removed from the column headers and any sample sequenced more than once has the abundances from these runs summed to give a single abundance per sample
5. A fasta file of unique ASV's is created from all ASV's in this final table
6. Sequences are classified.

**Classify the sequences**

Sequences are classified to provide taxonomies relative to the Silva database (Quast et al., 2013; Yilmaz et al 2014; Glöckner 2017) as below:

mothur "#classify.seqs(fasta=seqs.fasta, reference=SILVA132.ng.fasta, taxonomy=SILVA132.tax, cutoff=60, probs=FALSE, processors=X)"

**References**

Magoc, T. and Salzberg, S. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27(21): 2957-2963.

Schloss, P.D., et al., (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75(23):7537-7541.

Edgar, R.C., (2010), Search and clustering orders of magnitude faster than BLAST, Bioinformatics 26(19): 2460-2461.

Seqtk available at: <https://github.com/lh3/seqtk> (last accessed 23 Jan 2019).

Fastx available at: (<http://hannonlab.cshl.edu/fastx_toolkit/>) (last accessed 23 Jan 2019).

DeSantis T.Z., Hugenholtz P., Larsen N., Rojas M., Brodie E.L., Keller K., Huber T., Dalevi D., Hu P., Andersen G.L.(2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl. Environ. Microbiol. 72(7): 5069-5072; DOI: 10.1128/AEM.03006-05

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl. Acids Res. 41 (D1): D590-D596.

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucl. Acids Res. 42:D643-D648

Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, Bruns G, Yarza P, Peplies J, Westram R, Ludwig W (2017) 25 years of serving the community with ribosomal RNA gene reference databases and tools. J. Biotechnol.

* + 1. Fungal ITS Amplicon Analysis Workflow

This section outlines the workflow required to analyse amplicon sequences of the internal transcribed spacer (ITS) region located between the small and large rRNA subunits to produce Amplicon Sequence Variant (ASV) information for the Australian microbiome database. Amplicons are derived from primers targeting the fungal ITS1 and ITS4 regions (ITS1F and ITS4).

Analysis is completed on a per sequencing run (sequencing plate) basis. The workflow consists of the following stages:

**A] Sequence preparation and merging**

1. Convert fastq file format to fasta file format
2. Identify and isolate putative fungal ITS1 and ITS2 regions from R1 and R2 reads
3. Add SampleID, runID and "sample=" information to the sequence headers
4. Concatenate all sequences per sequencing run into a single file

**B] Sequence analysis**

1. Quality screening, zotu/ASV calling and sequence mapping
2. Classify and remove flipped sequences
3. Replace arbitrary ASV ID's with the sequence itself in the table index

**C] Following analysis, all data is combined to give a single dataset, by the following steps:**

1. Merge tables into a single table
2. Remove controls from the abundance tables, to create separate samples and control datasets
3. Make a fasta from the abundance table
4. Classify Sequences

**Software Required**

The following software is used in the steps below:

1. ITSx (Bengtsson-Palme et al., 2013)
2. Mothur (Schloss, et al., 2009)
3. USEARCH (Edgar 2010)
4. Seqtk (<https://github.com/lh3/seqtk>)
5. Fastx (<http://hannonlab.cshl.edu/fastx_toolkit/>)
6. Python3.x

**A] Sequence preparation and merging**

**Identify and isolate ITS1 and ITS2 regions**

Illumina fastq R1 and R2 files are first converted to fasta file format using SeqTk. In addition, SeqTk is used to generate the reverse complement of R2 reads.

ITSx (Bengtsson-Palme et al., 2013) is the used to identify and isolate fungal ITS1 and ITS2 regions from neighbouring ribosomal genes (SSU, 5S and LSU rRNA sequences). Arguments used for ITSx are as follows,

-t F --complement F --preserve T --partial 100 --save\_regions ITSn --detailed\_results T

Following merging, the merge quality is manually checked by examining the FLASH log file for the percentage of reads that were merged. Plates with low merge rates (< 70%) are manually checked to see if the alignments can be improved.

R1 and R2 reads not identified as ITS by ITSx are discarded

**Rename files**

File names are formatted to the following format:

sampleID\_plateID.fasta

The sampleID of the mock communities and negative controls on each plate are standardised to the following naming convention (all listed for completeness):

* 16S Bacteria mock communities: **BACMOCK**
* A16 Archaeal mock communities: **ARCMOCK**
* 18Sv4 Eukaryote mock communities: **EUKMOCKv4**
* 18Sv9 Eukaryote mock communities: **EUKMOCKv9**
* ITS Fungal mock communities: **FUNMOCK**
* Negative control: **NEG**
* STAN: **STAN**

**Add sample and file name to the sequence headers**

Sample identifiers are added to the header of each sequence for downstream processing. As each fasta is now named SampleID\_plateID.fasta we simply add the file name with the extension to the sequence header. At this stage we also add any other information and delimiters that downstream programs will likely require. For Usearch we add "sample=" and ";", for Qiime we add "\_".

#!/bin/bash

#this initiates the loop for all files with extension .fasta in the list generated by ls -l

for file in `ls -1 \*.fasta`

do

#tell me which file you're working on

echo working on $file

#cut the appropriate string from the name of the file.

VAR1=`echo $file | cut -d "." -f 1`

#tell me what you've cut and called VAR1

echo VAR1 is $VAR1

#use VAR1 to add the sample name to the sequence id. Substitute > for >VAR1\_ in all files. The variable (VAR1) needs to be in quotation marks to be passed from bash to perl, the perl script needs to be embedded in the bash script with the ` ` marks

`perl -pi -e 's/\>/\>sample='$VAR1';\_ITS\_/' $file;`

done

**Concatenate all sequences per sequencing run into a single file**

All of the files for each plate are concatenated into a single file for zotu/ASV calling. The resulting file name is standardized to the format: plateID\_all\_ITSn.fasta

Where ITSn represents the ITS region being analysed (ITS1 or ITS2)

**B] Sequence analysis**

**Quality screening, zotu/ASV calling and sequence mapping**

* The first step removes sequences that have ambiguous bases, or have more than 12 homopolymers.

mothur "#set.dir(modifynames=F);summary.seqs(fasta=plateID\_all\_ITSn.fasta, processors=10); screen.seqs(fasta=current, maxambig=0, maxhomop=12, processors=10); summary.seqs()"

* Next reads are dereplicated

usearch64 -fastx\_uniques plateID\_all\_ITSn.good.fasta -fastaout plateID\_all\_ITSn.good\_uniques.fasta -sizeout

* Sort unique reads by abundance

usearch64 -sortbysize plateID\_all\_ITSn.good\_uniques.fasta -fastaout plateID\_all\_ITSn.good\_sorted\_uniques.fasta -sizeout

* ASV/Zotus are called by UNOISE3, from sequences that have => 4 representatives

usearch64 -unoise3 plateID\_all\_ITSn.good\_sorted\_uniques.fasta -zotus plateID\_all\_ITSn.good\_sorted\_uniques\_zotus.fasta -ampout plateID\_all\_ITSn.good\_sorted\_uniques\_ampout.fasta -tabbedout plateID\_all\_ITSn.good\_sorted\_uniques\_unoise3.txt -minsize 4

* Map reads to zotus to generate abundances

Reads are mapped against the zotus using USEARCH. Note the termination conditions on the mapping run

(-maxaccepts 0), this seems to be required to ensure the best match is found and to produce consistent results when adding multiple plates together, as we do later.

usearch64 -otutab plateID\_all\_ITSn.fasta -zotus plateID\_all\_ITSn.good\_sorted\_uniques\_zotus\_renamed.fasta -otutabout plateID\_all\_ITSn.good\_sorted\_uniques\_zotutab\_MA0.txt -mapout plateID\_all\_ITSn.good\_sorted\_uniques\_zmap\_MA0.txt -maxaccepts 0 -threads X

**Classify and remove flipped sequences**

A final QC step is performed to remove likely erroneous sequences. The ASVs are classified, with those that do not align to the UNITE SH ITS database in the correct orientation being removed. As we know the that R1 sequences correctly orientated and the reverse complement of the R2 also puts it into the correct orientation, sequences that need to be "flipped" to a new orientation to obtain the best alignment to the database are most likely errors.

**1. classify the seqs against UNITE ITS database**

classify.seqs(fasta=${plateID}\_all\_ITSn.good\_sorted\_uniques\_zotutab\_relabelled\_MA0.fasta, reference=UNITEv7\_sh\_dynamic\_s.fasta, taxonomy=UNITEv7\_sh\_dynamic\_s.tax, cutoff=60, probs=T)

**2. use the \*acnos.flip list to remove sequences from the table**

The above produces the final abundance table, with sequences as index for each sequencing run. These tables are then combined as below to produce a single dataset

**Replace table indexes with ASV sequence**

Currently the tables have an arbitrary OTU number as the index, replace this with the sequence that the arbitrary number represents. These sequences are unique strings and allow tables to be merged etc. downstream easily. They also negate the need to maintain a dictionary of ASVs and the sequences they represent.

**C] Prepare the single dataset**

Now we have an ASV abundance table for each plate, with ASV's as row and SampleID\_runID as column headers. To prepare this data for ingest into the AMD database the following steps are carried out:

1. Each table is converted from short to long format (from rectangular to 3 column, with columns ASV, sampleID, Abundance)
2. All of these 3 column tables are concatenated into a single table
3. Controls and samples are split into separate tables
4. sequencing run ID's are removed from the column headers and any sample sequenced more than once has the abundances from these runs summed to give a single abundance per sample
5. A fasta file is created from all ASV's in this final table
6. Sequences are classified.

**Classify the sequences**

Sequences are classified to provide taxonomies relative to the UNITE ITS database as below:

mothur "#classify.seqs(fasta=seqs\_listSET.fasta, reference=UNITEv7\_sh\_dynamic\_s.fasta, taxonomy=UNITEv7\_sh\_dynamic\_s.tax, cutoff=60, probs=FALSE)"

**References**

Bengtsson-Palme, J., Veldre V., Ryberg M., Hartmann M., Branco S., Wang Z., Godhe A., Bertrand Y., De Wit P., Sanchez M., Ebersberger I., Sanli K., de Souza F., Kristiansson E., Abarenkov K., Eriksson K.M, Nilsson R.H.(2013). ITSx: Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for use in environmental sequencing. Methods in Ecology and Evolution, 4: 914-919.

Schloss, P.D., et al., (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75(23):7537-7541.

Edgar, R.C., (2010), Search and clustering orders of magnitude faster than BLAST, Bioinformatics 26(19): 2460-2461.

Seqtk available at: <https://github.com/lh3/seqtk> (last accessed 23 Jan 2019).

Fastx available at: (<http://hannonlab.cshl.edu/fastx_toolkit/>) (last accessed 23 Jan 2019).

Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Kõljalg U, Abarenkov K. (2018). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research, DOI: 10.1093/nar/gky1022

* 1. METAGENOMICS

Bioinformatic processing and analytic workflows for metagenomics will be made available on the Australian Microbiome Initiative GitHub: <https://github.com/AusMicrobiome>

Workflows that have been used for the Marine Microbes project the the processing of raw sequence data processing to reference gene catalogue and sequence binning can be found at: <https://github.com/martinostrowski/metagenome>

* 1. METATRANSCRIPTOMICS

Bioinformatic processing and analytic workflows for metatranscriptomics will be made available on the Australian Microbiome Initiative GitHub: <https://github.com/AusMicrobiome>