1 NEON protocols

1.1 NEON Community standard generation and storage

1.1.1 Initial protocol

Background: NEON is developing a DNA community standard consisting of known macroinvertebrate and/or zooplankton taxa in 95% ethanol.

Background: NEON is developing a DNA community standard consisting of known macroinvertebrate and/or zooplankton taxa in 95% ethanol.

The community standard will be homogenized with a hand immersion blender.
The sample will be pulsed for 2 seconds at a time for a total of 30 seconds. The steel blender will be cleaned via dipping/mixing in Rneasy followed by 80% ethanol between samples. The standard will betransferred to sterile 50ml conical vials and spun down for 30 minutes at 4000 RPM. Supernatants will be poured off. 0.25 grams of pellet material will be loaded into 12 extraction tubes (3 grams of material total).

QA/QC – Samples are visually inspected to ensure sample was homogenized prior to freezing.
Pass: if sample was completely homogenized
Fail: if sample still looks heterogeneous in areas
Action if fail: Repeat step 1 until homogenous

Genomic DNA from samples will be extracted using the DNeasy Powersoil Kit (100) (Carlsbad, CA) according to the manufacturer's protocol. gDNA will be visually inspected. gDNA should be clear in color, indicating samples were mostly free of PCR inhibitors and debris. Yellow or brown colored samples need to be extracted again. Each well will also be examined to ensure that an appropriate volume of liquid is present.

QA/QC: gDNA is visually inspected. gDNA should be clear in color, indicating samples were mostly free of PCR inhibitors and debris. Each well is also examined to ensure that an appropriate volume of liquid is present.
Pass: if sample is clear and ~100ul
Fail: if sample is yellow or brown colored or less than 90ul
Actions if fail: Yellow or brown colored samples or low volume samples need to be extracted again.

The 12 sample will be pooled into a single vial, mixed, and DNA concentration be analyzed with a Qubit4 fluorometer.

gDNA Quantitation Protocol:

gDNA will be quantitated using a Qubit fluorometer. Briefly, 2ul of each gDNA sample is incubated with 198 ul of Qubit broad range reagent and incubate 2 minutes at RT. Read on a Qubit v4 system (Invitrogen).

QA/QC: gDNA concentration should be between 0.5 and 100 ng/ul.
Pass: if sample concentration is within the expected range of 0.5 and 100 ng/ul
Fail: if sample is lower than 0.5 ng/ul or higher than 101 ng/ul.
Actions if fail: Repeat extraction protocol on a new raw sample.

Pooled gDNA will then be aliquoted into 15 cryovials. Each cryovial is not to be thawed and used more than 10 times. The actively being used vial will be stored at -20°C and tracked for number of freeze-thaw cycles and the remaining aliquots will be either stored at -80°C on site or extra vials will be sent to NEON Headquarters for storage.

1.2 Sequencing of zooplankton and macroinvertebrates using CO1 primers

1.2.1 October 2018 protocol

Primer sets to use for metabarcoding analyses

Target gene	F primer (5'-3')	R primer (5'-3')	Approx length	Reference
CO1, F230 fragment	GGTCAACAAATCATA AAGATATTGG	CTTATRTTRTTTATICGIGG RAAIGC	314	F: Folmer et al.; R: Gibson et al. 2015
CO1, BE fragment	CCIGAYATRGCITTYCCICG	GTRATIGCICCIGCIARIAC	224	F&R: Hajibabaei et al., BMC Bioinformatics, 2012

Updated: 2018-11
Update notes: none

General Precautions

Gloves must be worn at all times and changed frequently. Surfaces must be cleaned with Rnase Away followed by 70% Ethanol prior to preforming any work and again after completing work.

Sample Prep

All samples are shipped in 97.5% ethanol.

1. Macroinvertebrate samples are blended with a hand immersion blender. Each sample is pulsed for 2 seconds for a total of 30 seconds per 250 mL of material in the sample, calculated in 250 mL intervals. Steel blender is cleaned via dipping/mixing in Rneasy followed by 80% ethanol between samples. Gloves are changed between samples.
2. Zooplankton samples are allowed to sit unbothered for 4 hours. Using a 25ml transfer pipette, most of the ethanol is decanted until ~50ml of liquid remains.

3. 50ml of each of the samples are transferred to sterile 50ml conical vials and spun down for 30 minutes at 4000 RPM.
4. Supernatants are poured off and pellet is frozen at -20 °C.

QA/QC – Samples are visually inspected to ensure sample was homogenized prior to freezing.
Pass: if sample was completely homogenized
Fail: if sample still looks heterogeneous in areas
Action if fail: Repeat step 1 until homogenous

Extraction Protocol
Pelleted samples are thawed at room temperature. A sterile swab is used to collect homogenized material and placed into a well of the extraction plate. Genomic DNA from samples are extracted using the MoBio PowerSoil Isolation Kit (Carlsbad, CA) according to the manufacturer's protocol. Note: samples extracted in tubes rather than plate format.

QA/QC: gDNA is visually inspected. gDNA should be clear in color, indicating samples were mostly free of PCR inhibitors and debris. Each well is also examined to ensure that an appropriate volume of liquid is present.
Pass: if sample is clear and ~100ul
Fail: if sample is yellow or brown colored or less than 90ul
Actions if fail: Yellow or brown colored samples or low volume samples need to be extracted again.

gDNA Quantitation Protocol
All extracted gDNA will be quantitated using a Qubit fluorometer. Briefly, 2ul of each gDNA sample is incubated with 198 ul of Qubit broad range reagent and incubate 2 minutes at RT. Read on a Qubit v4 system (Invitrogen).
QA/QC: gDNA concentration should be between 0.5 and 100 ng/ul.
Pass: if sample concentration is within the expected range of 0.5 and 100 ng/ul
Fail: if sample is lower than 0.5 ng/ul or higher than 101 ng/ul.
Actions if fail: Repeat extraction protocol on a new raw sample.

PCR Protocol
96 well PCR plate set up is performed in a biosafety cabinet. Each plate will contain 50 samples, the positive community control, 1 negative control (blank), and 1 non-template control. This set up will be replicated 6 times, 3 for each primer pair listed above. This allows for the PCR to occur in triplicate before pooling. Triplicate PCR minimizes stochasticity and decreases the likelihood of a failed amplification. A 2-step PCR system is used to amplify and index the amplicons. With this process, two portions of CO1 gene are PCR amplified from each genomic DNA sample using the primer pairs listed above. Both primers also contain a 5' adaptor sequence to allow for subsequent indexing and Illumina sequencing. Each 25 µL PCR reaction is mixed according to the Promega PCR Master Mix specifications (Promega catalog # M5133, Madison, WI) which included 0.4 µM of each primer and 1 µl of gDNA. DNA is PCR amplified

using the following conditions: initial denaturation at 95 °C for 5 minutes, followed by 40 cycles of 40 seconds at 94 °C, 1 minute at 46 °C, and 30 seconds at 72 °C, and a final elongation at 72 °C for 10 minutes. (Protocol was from Hajibabaei 2012 and Gibson 2015 papers listed above)

QA: 5ul of each PCR product is run on a 2% agarose gel. DNA bands of the desired size (either ~200 or ~300 bp) are scored on their intensity value on a scale of 0-3 with 3 being the brightest and thus containing the most PCR product.
Pass: All samples must have a band intensity of 2 and that appropriate size fragments are present. Each well is also examined to have sufficient volume of liquid.
Fail: If samples have a band intensity of 0 or 1 or do not have appropriate size fragments or volume of liquid.
Action if fail: PCR is repeated with a new aliquot of sample.

Further QA: Positive PCR should result from positive control, negative PCR should result from negative control. If not, troubleshoot accordingly (e.g. identify sources of contamination, re-extract or re-PCR as necessary). If NTC produces positive PCR result, troubleshoot source of contamination and re- run samples as necessary.

Post-PCR protocol:
Amplicons are cleaned by incubating amplicons with Exo1/SAP for 30 minutes at 37 °C following by inactivation at 95 °C for 5 minutes. A second round of PCR is performed to give each sample a unique 12-nucleotide index sequence. The indexing PCR included Promega Master mix, 0.5 µM of each primer and 2 µl of template DNA (cleaned amplicon from the first PCR reaction) and consisted of an initial denaturation of 95 °C for 3 minutes followed by 8 cycles of 95 °C for 30 sec, 55 °C for 30 seconds and 72 °C for 30 seconds.

QA/QC: 5µl of PCR products of each sample are visualized on a 2% agarose gel and scored on the same 0-3 scale as above. Each gel is run with a size standard and products are ensured to have proper size fragments.
Pass: if >75% of the PCR product is the correct amplicon size
Fail: if >25% of the PCR product is the wrong size or no PCR product is amplified
Action if failed: Repeat the PCR reactions of the failed samples. If failed twice, return to the raw sample and extract a new gDNA sample then follow with PCR. If still failing proceed with pooling and sequencing.

Normalization and Pooling

Final indexed amplicons from each sample are cleaned and normalized using SequalPrep Normalization Plates (Life Technologies, Carlsbad, CA). 25 µl of PCR amplicon is purified and normalized using the Life Technologies SequalPrep Normalization kit (cat#A10510-01) according to the manufacturer's protocol. 5ul of each normalized sample is then pooled together. Each pooled sample will consist of 1 positive control, 1 negative control, 1 no-template control, and no more than 93 samples.

Pre-Sequencing QC

Pooled libraries are quantified via Qubit using a high sensitivity assay reagents. Amplicon size of the pooled library is sized by running on a TapeStation using 1000bp high sensitivity reagents. Size and concentration are used to determine the molarity of the library.

QA/QC-Library should be free of 200bp amplicons. All 200bp fragments are purified out using AmpPure beads at a 0.7X ratio.
Pass: if >95% of the pool is >201bps on TapeStation report
Fail: in >5% of overall sample is the 200 bp or smaller
Action if fail – purify pools with 0.7X ratio of AmpPure beads to remove the smaller amplicons. Repeat TapeStation and bead cleanup protocols until pool is acceptable in size.

Sequencing protocol

10pM of library with a 15% spike in of PhiX is sequenced on an Illumina MiSeq (San Diego, CA) in the CU Boulder BioFrontiers Sequencing Center using the v2 500-cycle kit (cat# MS-102-2002).

QA/QC – Runs are monitored via BaseSpace
Pass: if Q score is >20 and run passed internal QC
Fail: if Q score is <20 or fails Ilumina internal QC
Action if Fail – Rerun the library pool on new MiSeq and/or with a new kit. If still failing, repeat all PCRs and pooling protocols until sequencing run passes

Post-sequencing QC

Sequencing data will be processed to assess the following:

1. Verify sequencing depth success rates. At least 90% of samples should produce high-quality (e.g. post-filtering procedures) sequence data at the target sequencing depth of 50k reads. Action if fail: If this success rate is not achieved, discuss with the NEON project team the strategy to troubleshoot failed samples. This may include re-processing and re-analysis of failed samples.
2. Ensure minimum sequencing depth. Minimum 3000 sequences per sample; minimum PHRED-equivalent quality score of 20. No more than 1 ambiguous base call in a sequence. Action if fail: Re-run sample. If sample fails upon re-run, communicate to technical representative for final decision.
3. Ensure data quality. Sequences with quality scores equivalent to Phred scores < 20 over a 10-bp window will be removed, primer adapter regions are removed and sequences demultiplexed. Action if fail: remove data for failing samples.
4. Positive PCR should result from positive control, negative PCR should result from negative control. If not, troubleshoot accordingly (e.g. identify sources of contamination, re-extract or re-PCR as necessary).
5. If NTC produces positive PCR result, troubleshoot source of contamination and re- run samples as necessary.

6. Positive control sample must successfully pass all QA/QC checks and produce sequence data meeting the criteria for 50k reads per sample.

1.3 Traceability chain for sample analysis

1.3.1 Initial protocol

Updated: November 7, 2018

Sample receiving

1. Only trained lab personnel will receive samples.
2. Outer box is visually inspected for damage prior to accepting the package
3. Shipments are opened and inspected for visual damage or leakage. Any leaking or damage are noted on the sample ingest form. Leaking or damage that compromises the quality of the sample (i.e. possible contamination) must be reported to supervisor immediately. Samples must be left in the box until further instructions.
4. Samples are removed from the box and cross-referenced to the included packing list (or ingest form). Once contents are confirmed, the 'Jonah Sample Document' should be updated with PI, number of samples, date, responsible person, and primers requested. If sample count/IDs do not match ingest form, supervisor must be contacted immediately.
5. Ingest forms/packing lists are stored in filing cabinet for future reference.
6. Samples will be stored at 4° C prior to processing

Sample Processing

1. Barcode reader will be used to log each sample.
2. Each sample will be processed gDNA extracted as described in the SOPs. Each gDNA sample will be added to a 96-well plate.
3. Excel spreadsheets are used to track the location of each sample.
4. Each 96 well plate of gDNA will be barcoded and stored at -20 °C in numerical order
5. gDNA plates will be used as template for PCR plates. Location of each PCR reaction will also be logged into the excel tracking sheets and barcoded
6. Samples will be pooled and again tracked on the excel tracking spreadsheet.
7. Mapping files will be created in which each sample is tracked by PI name, sample barcode, and unique indexing sequence.

Bioinformatics

1. Samples will be tracked via unique indexing sequence. Sequencing and bioinformic processing will be tracked via 'Jonah Sample log sheet'

Data storage

1. All sample tracking sheets are stored on the cloud via Dropbox and Google docs.
2. AWS is used to store sequencing data

Sample storage

1. Unprocessed samples in ethanol will be stored at 4° C after receipt
2. Processed raw samples will be stored in 50 ml conical vials in the 'VWR' -20 freezer
3. gDNA will be stored in barcoded plates in the 'Imperial' -20 °C freezer
4. PCRs will be stored in barcoded plates in the 'Imperial' -20 °C freezer
5. PCR plates and pooled amplicons will be discarded 3 months after sequencing run is complete

Sample delivery to transport to sequencing facility

1. Samples need to be transported to the CU Boulder Sequencing facility located at 3415 Colorado Avenue Boulder, CO 80303 which is approximately 2 miles from Jonah Ventures
2. Samples are packing in insulated containers to protect from UV and excess heat.
3. Trained Jonah Employees use personal vehicles to transport the samples to the sequencing center.
4. Upon arrival, samples are transferred directly to a member of the CU Sequencing Center Staff who then accepts custody of the library.

1.4 Data return

1.4.1 Initial protocol

Overview

Contractor's data will be paired with NEON metadata for posting to the publicly available MG-RAST sequence data repository. The NEON project team will then create the final package and post to MG-RAST.

Raw Data files to BOX filesharing folder

A BOX filesharing folder will be set up by Battelle and dedicated to this project for return of raw sequence files

1. Only NEON data will be accepted. Data from any non-NEON samples must be removed from raw data files prior to delivery to NEON. Only NEON samples will be present on a run.
2. Raw sequence files. These are pre-processed, unfiltered fastq files and any metadata files (e.g. 'mapping' files) that are required for downstream processing. Demultiplexed fastq files are also acceptable and should be named according to the conventions below. Data should be provided as compressed files, with each file containing the fastq sequences from no more than 1 sequencing run. Maximum file size = 5GB. Raw files may be split in an organizationally coherent way to ensure files do not exceed the size limit. For example, the forward sequence read files (e.g. R1 data) from a single gene fragment can be compressed into a single tar.gz (or .zip) file. Name files according to the convention: ▪ Lab identifier + locally unique sequencing run identifier + gene identifier + read direction. Ex: Lab1.run20180201.CO1BE.R1.fastq.tar.gz Only alphanumeric

characters, periods, and underscores are allowed in file names. Only the designated file types are allowed.

Processed, quality-checked sequence data

1. Demultiplexed, merged, and quality-filtered sequence files (e.g. fasta), with each sample in a separate file. These should be uploaded to a dedicated Box folder. As with the raw files, files should be no greater than 5GB in size. Name files according to the convention:
   ▪ Lab identifier + locally unique sample identifier + gene identifier. Ex: Lab1.NEONuid.CO1BE.fasta. Only alphanumeric characters, periods, and underscores are allowed in file names.

Completed data ingest sheets (Attachment 2) uploaded to NEON Data Portal

Contractor is responsible for uploading completed data ingest sheets to the NEON Data Portal. Instructions for use of the data portal upload system are included in Attachment 2a.
Data ingest sheets, Attachment 2b, need to be completed following the field definitions, prescribed units, and use of choices from options list where defined options must be used as outlined in Attachment 2b. Adherence to the formatting requirements is essential for the data to be uploaded to the NEON project database. Each of the three tabs in Attachment 2b (dnaExtraction, PCR, sequencing, and rawDataFiles) must be saved as separate csv files for loading to the NEON Data Portal. General formatting requirements are as follows:

- File Type = csv (in MS Office Excel, save as “Comma Separated Values, .csv”)
- Encoding = UTF-8
- Null convention = leave cells without information empty (do not use “NA”, “NULL”, etc.)
- Dates (and Times): Report local dates
- Date Format = YYYYMMDD or YYYY-MM-DD
- Max characters within a cell = 255
- If removing entire rows of data, the entire row should be deleted and not removed by “clear contents” or backspacing to delete. Clearing contents and backspacing to delete results in an empty data row (“ghost rows”) that won’t load to the database. “Ghost rows” become visible by opening up the csv file in a text editor such as Notepad and appear as rows of commas. They can be removed by deleting the entire row.
- Text strings need to be placed within double quotes unless the csv file is being created using Microsoft Excel. If using Excel, allow Excel to do its own formatting of text strings when the file is saved as .csv, i.e., do not purposefully add double-quotes as Excel will do this automatically where needed.