Standard Operating Procedures

Jinliang Li

Title: NEON INVERTEBRATE METABARCODING

Author: Jinliang Li

Version: 1.0

Authorized By: Jinliang Li

Authorized Date: 07/25/2023

NEON protocols

1.0 Sequencing of zooplankton and macroinvertebrates using CO1 primers

PCR primer sets to use for metabarcoding analysis. Both primer set will be used for all samples (i.e., both macroinvertebrate and zooplankton are analyzed with both the F230 and BE primers)

Target gene	F primer (5'-3')	R primer (5'-3')	Approx length	Reference
CO1, F230 fragment	GGTCAACAAATC ATAAAGATATTGG	CTTATRTTRTTTATICG IGGRAAIGC	314	F: Folmer et al.; R: Gibson et al. 2015
CO1, BE fragment	CCIGAYATRGCITTYCC ICG	GTRATIGCICCIGCIARI AC	224	F&R: Hajibabaei et al., BMC Bioinformatics, 2012

General Precautions

Gloves must be worn at all times and changed frequently. Surfaces must be cleaned with 10% Bleach followed by 75% Ethanol prior to preforming any work and again after completing work. All work is performed with sterile techniques and materials.

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. Steel blender is cleaned via dipping/mixing in 10% Bleach followed by 80% ethanol between samples. Blender is allowed to dry, and gloves are changed between samples.
   Note: Samples may be transferred to a secondary container for homogenization if the original container mouth is not wide enough for the blender.
2. Zooplankton samples are allowed to sit unbothered for 4 hours. Using a 25ml transfer pipette, most of the ethanol is decanted until ~15 ml of liquid remains.
3. 50ml of each of the samples are transferred to sterile 15 ml centrifuge tube and spun down for 30 minutes at 4000 RPM.
4. Supernatants are poured off and pellet is frozen at -20 °C in racks in sequential orders according to the sample sheet.

QA/QC – Samples are visually inspected to ensure each sample was homogenized after blending.

Samples are considered homogeneous if the blended particles are uniform in size within the ethanol.
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. Under a laminar flow hood, sterile cotton swabs (Fisher, cat# 22-363-173) were coated with sediment matter, and the swabs were placed in the corresponding extraction plate or tube. Genomic DNA from samples will be extracted using Laragen SOP.02.043 - Microbial DNA Extraction From Fecal Samples. Briefly, the samples mechanically shake with Zirconia Beads in lysate buffer with Proteinase K. Genomic DNA is extracted with a proprietary magnetic beads-based protocol using a Kingfisher Flex. A blank control will be used as negative extraction control.

Note: samples extracted in 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 consists ~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 for 2 minutes at RT. Read on a Qubit v4 system (ThermoFisher).

QA/QC: gDNA concentration should be between 1.0 and 100 ng/ul.
Pass: if sample concentration is within the expected range of 1.0 and 100 ng/ul
Fail: if sample is lower than 1 ng/ul or higher than 101 ng/ul.
Actions if fail: Repeat extraction protocol on a new raw sample or dilute to <100 ng/ul

gDNA storage

gDNA will be stored in a 2 ml screw cap tube in a 80°C ultra freezer in a cardboard box. The gDNA tubes should be arranged sequentially according to the gDNA sample PCR Protocol

PCR Protocol

96 well PCR plate set up is performed in a PCR workstation. Each plate will contain 93 samples,

1 positive control (NEON DNA Community Standard), 1 negative control (blank), and 1 non-template control. 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 below Both primers also contain a 5' adaptor sequence to allow for subsequent indexing and Illumina sequencing. Each 30 µL PCR reaction is mixed according to the Thermo Fisher PCR Master Mix specifications (Thermo Fisher # C571B001, CA) which included 0.5 µM of each primer and 2 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 et al., 2012 and Gibson et al., 2015).

QA: 5ul of each PCR product is run on a 2.5% agarose gel. DNA bands of the desired size (either 224 or 314) 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 approximately 30µl 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 a positive control, negative PCR should result from a 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 Laragen SOP.02.006 - Direct PCR Product Purification RevD. A second round of PCR is performed to give each sample a unique 8-nucleotide index sequence. The indexing PCR included Thermo Master mix, 0.5 µM of each primer (Illumina Nextera Index primer) and 2 µl of template DNA (cleaned amplicon from the first PCR reaction) and consisted of an initial denaturation of 95 °C for 10 minutes followed by 8 cycles of 95 °C for 10 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 2.5% 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 (either 224 or 314).

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 based on the agarose gel and DNA concentration. 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 reagent. Amplicon size of the pooled library is sized by running on an ABI 3730XL sequencer. Size and concentration are used to determine the molarity of the library. Refer Laragen SOP.02.029 - MiSeq Library Sizing and Quantification for detailed protocol.

QA/QC-Library should be free of 200bp amplicons. All 200bp fragments are purified out using AmpPure beads at a 0.8X ratio.
Pass: if >95% of the pool is >201bps on ABI 3730XL Genemapper plot view.
Fail: if >5% of overall sample is the 200 bp or smaller
Action if fail – purify pools with 0.8X ratio of Laragen beads to remove the smaller amplicons. Repeat sizing and bead cleanup protocols until pool is acceptable in size.

Sequencing protocol

8.0 pM of library with a 15% spike in of PhiX is sequenced on an Illumina MiSeq (San Diego, CA) in Laragen facility using the v2 500-cycle kit (cat# MS-1022003). Percentage of PhiX spike can vary and is determined using QC measures and the known diversity of the library pool itself. Stated above is the Illumina protocol for PhiX which can be just adjusted depending on QC results. The library was loaded to MiSeq with a 500 cycle flow cell following Laragen SOP.02.030 - MiSeq Ready-to-Run Loading and Operation RevB.

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 reprocessing 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.
7. Sequencing negative control should have <3000 reads. If the negative control has more than 3000 reads, troubleshoot the contamination and re-run samples if necessary.
8. Neon community standard control should pass all QA/QC and produce the sequence data meeting the criteria outlined in sections 9,10, 11.
9. 80% Sequence read length should be long than 100 bp.
10. At least 70% of total bases in Sequence read run have sequences with Q30 or higher.
11. The median number of sequences per samples in each sequencing run must be at least 45,000.
12. Notify NEON team if sections 9,10,11 did not pass the QA/QC.

2.0 Traceability chain for sample analysis

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 and in good condition, an electronic receipt form emailed with the shipping information will be completed and uploaded to the NEON Data Portal. If sample count/IDs do not match ingest form or were received in a compromised condition, Battelle with be notified by responding directly to the shipping notification email using “reply-all.”Ingest forms/packing lists are stored in filing cabinet for future reference.
5. Samples will be stored at 4° C prior to processing

Sample Processing

1. Barcode reader will be used to log each field sample and the homogenized, pelleted samples.
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 bioinformatic processing will be tracked via 'NEON Sample log sheet'
2. The sequences will be analyzed with Qiime2 using COI database as described in Laragen SOP.02.053 - COI Metagenomics Qiime2 Analysis.
3. The F230 and BE primer datasets will be analyzed separately.

Data storage

1. All documents and files shared with NEON will be uploaded to the NEON Data Portal and Box.

3.0 Sample storage

1. Unprocessed samples in ethanol will be stored at 4° C after receipt
2. Processed raw samples will be stored in 15 ml conical vials in a -20 freezer
3. gDNA will be stored in 2ml screw capped tubes in a -20 °C freezer
4. PCRs will be stored in barcoded plates in a -20 °C freezer
5. PCR plates and pooled amplicons will be discarded 3 months after sequencing run is complete.
6. For long term storage, genomic DNA and community standard DNA are stored in 2.0 ml screw capped tubes in a-80^{o}C ultra freezer.
7. Genomic DNA will be returned to NEON upon request.

4.0 Data return

Data return will consist of the following. Only NEON samples will be present on a run so any raw data files will only include NEON data:

1. NEON specific ingests in .csv format will be uploaded to the NEON Data Portal. Attachment 2c_inv_dnaExtraction.csv must be uploaded first before any other NEON data ingests can be returned.
2. Demultiplexed raw sequence files will be uploaded to a BOX filesharing folder specified by NEON.
3. Multiplexed raw sequence files will be uploaded will be uploaded to a BOX filesharing folder specified by NEON.
4. FastQC or equivalent sequence quality files will be uploaded to a BOX filesharing folder specified by NEON.

5.0 References

1. Folmer, O., et al. DNA primers for amplification of mitochondrial Cytochrome C oxidase subunit I from diverse metazoan invertebrates. 1994 Molecular Marine Biology and Biotechnology 3(5):294-9
2. Gibson, J. F., et al. Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. 2015 PLoS ONE 10(10):e0138432
3. Hajibabaei M., et al. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. 2012 *BMC Ecology* **12:28**
4. Laragen SOP.02.006 - Direct PCR Product Purification RevD
5. Laragen SOP.02.029 - MiSeq Library Sizing and Quantification
6. Laragen SOP.02.043 - Microbial DNA Extraction From Fecal Samples
7. Laragen SOP.02.030 - MiSeq Ready-to-Run Loading and Operation RevB
8. Laragen SOP.02.053 - COI Metagenomics Qiime2 Analysis.

6.0 Version History

Version 1.0 – This is the first version of this document.