

DNA Metabarcoding of Pooled Specimens with Oxford Nanopore's Flongle

Introduction: This protocol describes our approach to DNA metabarcoding using the Nanopore Flongle. It covers brief background on Nanopore sequencing and DNA metabarcoding, the primers we designed for the metabarcoding, our PCR approach, the Nanopore library prep and sample loading, and the data analysis using the Ontbarcoder app.

Oxford Nanopore sequencing: Oxford Nanopore's sequencer¹ is a third generation real-time long-read sequencer that is becoming increasingly popular. It is relatively inexpensive (starts at \$1000), is small and portable, generates long reads (1000's of base pairs), and sequences in real-time, meaning that you can download and analyze sequence data as the sequencing reaction proceeds. Nanopore sequencing works by detecting the change in the electrical charge in nanopores on flow cells as DNA is pulled through the pore. DNA nucleotides (A, C, T, G) change the electrical charge in different ways as they go through the nanopores, so the machine can determine the sequence of the DNA strand based on how the pore charge changes.

The Flongle: Nanopore Flow cells come in two varieties, regular flow cells for large projects (cost ~\$1,000) and Flongle² flow cells for small experiments (cost ~\$90 per flow cell). Although the Flongle flow cell cost is not too high, it is still too much to sequence a single specimen. Regular Sanger sequencing costs \$2-6 per specimen! So, specimens have to be pooled for sequencing, that is, several to many specimens are loaded together on a single Flongle flow cell. To separate the specimens later, they need to be tagged with barcodes so that they can be identified.

DNA metabarcoding: DNA barcoding is the use of a reference sequence to identify species. One sequences a designated barcoding gene, traditionally the mitochondrial COI gene, and then compares the sequence obtained against a database of barcode sequences. DNA metabarcoding refers to the pooling of many individuals in a single sequencing reaction to identify species using DNA barcoding. The Nanopore sequencer can be used for DNA metabarcoding and generate sequences for many specimens in a single sequencing run.

DNA Metabarcoding in our lab: In our lab, we use the Nanopore sequencer with Flongle flow cells to perform DNA metabarcoding. DNA is extracted using phenol-chloroform³, Qiagen⁴, or even Chelex⁵ (insects) protocols. We then conduct PCR to amplify the COI DNA barcoding gene (other genes like 12S and 16S⁶ can also be used), run the products on an agarose gel to see how

¹ <https://nanoporetech.com/>

² <https://nanoporetech.com/products/flongle>

³ https://condor.depaul.edu/waguirre/1_dna_extraction_protocol_19.pdf

⁴ <https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/dneasy-blood-and-tissue-kit>

⁵ https://condor.depaul.edu/waguirre/chelex_dna.pdf

⁶ <https://via.library.depaul.edu/depaul-disc/vol12/iss1/2/>

the PCR worked, clean up the PCR products, quantify them using a Qubit, pool products from many different specimens, conduct the library prep, sequence them, and then analyze the data using the Ontbarcoder pipeline. We cover these steps in the rest of the protocol.

Primer design and logic: Special primers must be used for DNA metabarcoding that include barcodes to identify the individual specimens that are being sequenced in the specimen pool. Otherwise, you cannot link a specific sequence with a specific specimen, unless you already know the specimen’s DNA code. One could order primers with the barcodes already attached. This would be a simpler procedure but has the disadvantage that one would need to order each set of COI primers used with each barcode. To sequence 96 specimens in a pooled reaction, we have 12 forward barcodes that are combined with 8 reverse barcodes. Individual specimens can be identified by the particular combination of forward and reverse barcodes, and 12 x 8 gives 96 possible combinations for individual identification. So, we would need to order 20 barcoded primers for each different COI primer sequence that we use. Sometimes, a single primer pair may not work for all species and several sets must be tried. The alternative is to order the regular COI primers with a universal M13 tag that allows the combining of any barcode with any primer if both have the same M13 tag. The barcodes can then be ordered with the M13 tag so that a single set of 20 barcodes can be combined with any COI primers that have the M13 tags. This cuts the primer cost substantially. One can try many different COI primers, or primers for other mtDNA genes, by ordering them with the M13 tag and then using the universal barcode set that we already have in the lab.

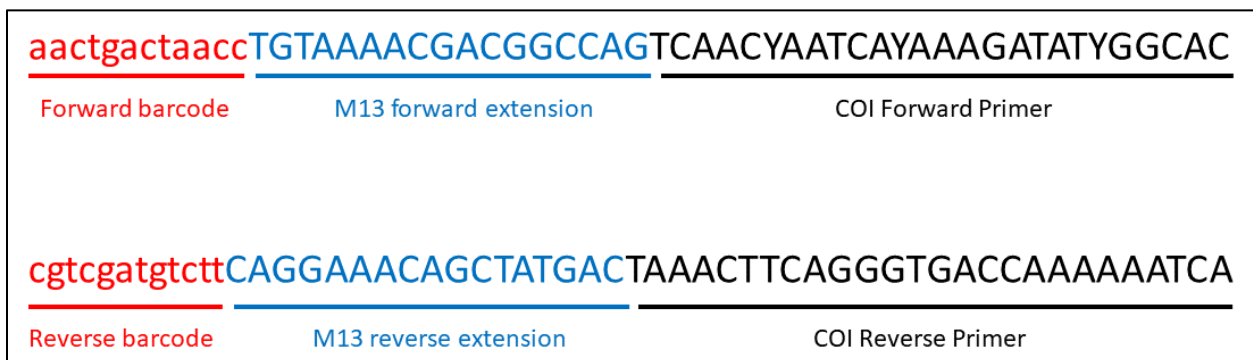


Figure 1. Visual depiction of COI primers with M13 tags, and barcodes. We have 12 different forward barcodes and 8 reverse barcodes created using the Barcode Generator program.

The structure of the primer, tag, and barcode in the final product to be sequenced is displayed in Figure 1. This structure is achieved over two rounds of PCR. In the first round, PCR is performed with the COI primers that have the M13 tags. This amplifies the COI gene and incorporates the M13 tag into the PCR products. In the second PCR reaction, the barcodes + M13 tags are used as primers. The M13 tag in the barcode+M13 complex will bind to the complement of the M13 tag sequence that was incorporated into the PCR products in the previous round. This 2nd round of PCR product makes additional PCR product that now has the barcodes incorporated into them. A different combination of forward and reverse barcodes+M13 tags are used for each specimen so that these can be individually identified after

the sequencing is performed. Table 1 lists the two sets of COI + M13 primers that we have in the lab as of August 2023 and the different forward and reverse barcodes + M13 tags. The barcodes were designed using the Barcode Generator app, are 12 bases long with a GC content 40-60%, and differ from each other by at least 6 bases. They were also examined for self-binding.

Table 1. List of COI primers, M13 tag sequences, and forward and reverse barcodes that the lab has. We have two sets of COI primers with M13 tags for testing with different species. We also have 12 forward barcodes and 8 reverse barcodes. In combination, these allow for the individual labeling of 96 specimens or samples (12 X 8 = 96). Use of the M13 tags allows for the combination of any primer with any barcode. The barcodes were designed with the Barcode Generator⁷. The barcodes are 12 bases long, GC content 40-60, and differ from each other by at least 6 bases. They were also examined for self-binding.

Primer Name	Primer Sequence	Description
M13F-COI-Fish F1	TGTA AACGACGGCCAGTCAACYAATCAYAAAGATATYGGCAC	M13 tag + Primer
M13F-Ward-FishF1	TGTA AACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC	M13 tag + Primer
M13R-COI-Fish-R1	CAGGAAACAGCTATGACTAAACTTCAGGGTGACCAAAAATCA	M13 tag + Primer
M13R-Ward-FishR1	CAGGAAACAGCTATGACTAGACTTCTGGGTGGCCAAAGAATCA	M13 tag + Primer
BCF01_M13F	aactgactaacTGTAAACGACGGCCAG	Barcode + M13 tag
BCF02_M13F	aagtgttcaTGTAAACGACGGCCAG	Barcode + M13 tag
BCF03_M13F	aattcgactgTGTAAACGACGGCCAG	Barcode + M13 tag
BCF04_M13F	acataacgtccTGTAAACGACGGCCAG	Barcode + M13 tag
BCF05_M13F	agcctcacgtTGTAAACGACGGCCAG	Barcode + M13 tag
BCF06_M13F	agggttccagaTGTAAACGACGGCCAG	Barcode + M13 tag
BCF07_M13F	atccactgcttTGTAAACGACGGCCAG	Barcode + M13 tag
BCF08_M13F	atcgaggtctgaTGTAAACGACGGCCAG	Barcode + M13 tag
BCF09_M13F	attggtgggtTGTAAACGACGGCCAG	Barcode + M13 tag
BCF10_M13F	caggtgctgtaTGTAAACGACGGCCAG	Barcode + M13 tag
BCF11_M13F	cccatagactttTGTAAACGACGGCCAG	Barcode + M13 tag
BCF12_M13F	cggcatggagttTGTAAACGACGGCCAG	Barcode + M13 tag
BCR01_M01R	cgtcgatgtctCAGGAAACAGCTATGAC	Barcode + M13 tag
BCR02_M13R	ctctatcggccaCAGGAAACAGCTATGAC	Barcode + M13 tag
BCR03_M13R	gcactccaagtaCAGGAAACAGCTATGAC	Barcode + M13 tag
BCR04_M13R	gcgagaagacacCAGGAAACAGCTATGAC	Barcode + M13 tag
BCR05_M13R	ggaggaactaggCAGGAAACAGCTATGAC	Barcode + M13 tag
BCR06_M13R	gttcgataagcaCAGGAAACAGCTATGAC	Barcode + M13 tag
BCR07_M13R	tcgaaagaggagCAGGAAACAGCTATGAC	Barcode + M13 tag
BCR08_M13R	tgagaggagccaCAGGAAACAGCTATGAC	Barcode + M13 tag

⁷ http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator

PCR and gel electrophoresis: The phase I PCR reaction is a normal PCR reaction in which the COI primers + M13 tag are used as the primers. All specimens get the same primer pair. The PCR reaction is conducted under normal conditions to a 30 μ l volume. In tests (Fig. 2A), an annealing temperature of 56 $^{\circ}$ C produced crisper products than 50 $^{\circ}$ C, so I recommend an annealing temperature of 56 $^{\circ}$ C. 10 μ l of PCR product are run on a 1.5% agarose gel and the PCR product is quantified using the Qubit (although this step can be skipped if product looks good on the gel). A crisp PCR product band should appear for all specimens.

The phase II PCR reaction is conducted using a 1:10 dilution of the phase one PCR products as the DNA template. The barcode+M13 tags are used as the primers and each individual specimen gets a different combination of forward and reverse barcodes for bioinformatic individual identification. PCR reactions can be run in 30 μ l volumes again. Since there are 8 reverse barcodes and 12 forward, I recommend having a different reverse barcode in each of 8 tubes in a standard strip tube, and have each different strip tube have a different forward barcode. So, the reverse barcode is repeated in the same position in each strip tube and the forward barcode is the same for all tubes in a strip, changing from strip to strip. Up to 12 strip tubes can be combined in a single sequencing reaction using this design. Using the same design across experiments will also make standardizing the data analysis easier. In tests (Fig. 2B), diluting phase I PCR products 1:10 worked much better than using undiluted PCR products as template, and an annealing temperature of 56 $^{\circ}$ C worked better than 50 $^{\circ}$ C.

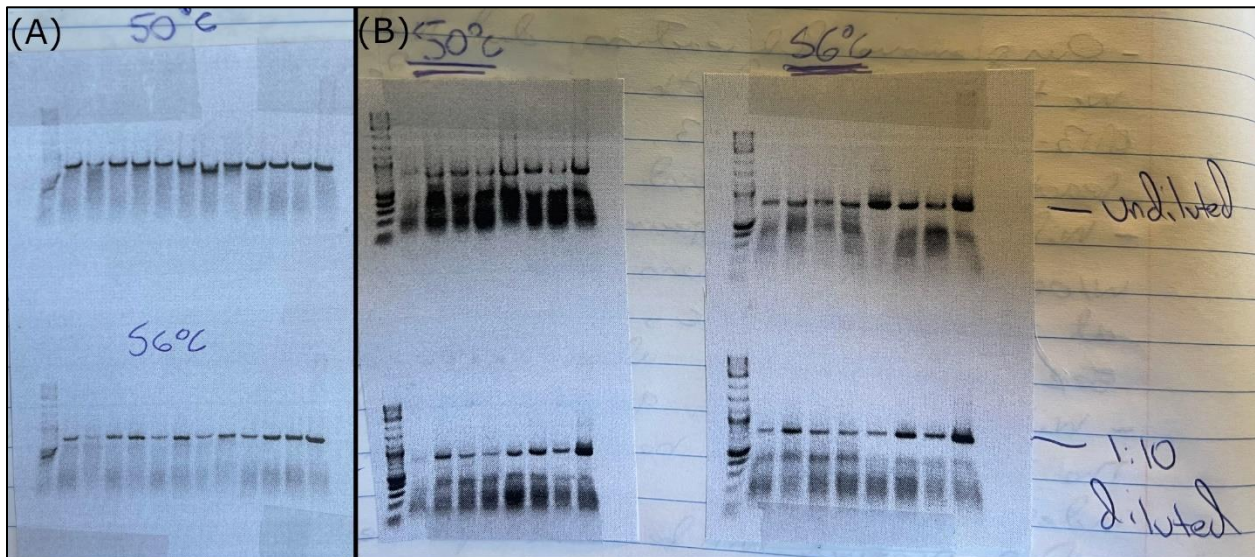


Figure 2. Scans of lab notebook showing results of (A) phase I and (B) phase 2 PCR tests. In phase I tests, an annealing temperature of 56 $^{\circ}$ C gave crisper PCR products. In phase II tests, diluted (1:10) phase I products used as template run at 56 $^{\circ}$ C gave the best results.

PCR product quantification and cleanup: Quantification of the phase II PCR products can be done with the Qubit following our protocol⁸. A clean up of the phase II PCR products should be conducted prior to library prep for Nanopore sequencing. This was not done in our experiments

⁸ https://condor.depaul.edu/waguirre/Qubit_4_Protocol.pdf

and the sequencing worked but only a low percentage of reads passed the filtering steps for the Ontbarcoder pipeline. Cleanup of PCR products will likely result in an increase of the read percentage that passes filtering. Srivathsan et al. (2021)⁹ suggest using SPRI bead-based clean-up with Ampure beads (Beckman Coulter). After cleanup, quantify PCR products again using the Qubit. The Flongle protocol calls for 500ng of DNA in a 25µl volume, so basically the final pooled PCR product concentration should be 20ng/µl. Standardize concentrations and pool PCR products to yield 500ng of DNA in 25µl volume.

Library preparation and Nanopore sequencing: Oxford Nanopore provides protocols and kits for library preparation for sequencing. We used the Genomic DNA by ligation kit for the Flongle (SQK-LSK109). The kits contain many but not all the products needed. A companion kit from New England Biolabs is required, as are Ampure XP beads for purification. These are in the Nanopore protocol. We follow the Nanopore protocol and have gotten good results.

Once the library is ready, it can be loaded on the Flongle flow cell. The sequencer will begin generating reads within a few minutes and data should appear in the output folder within the hour. The sequences are recorded in files of 1000 reads. These files accumulate over time and can be downloaded periodically. The Fastq files (genomic equivalent to Fasta files) are created from files with the electrical signal data, so they lag relative to the sequencing reaction. At the end of the process, however many Fastq files you have can be multiplied by 1000 to get the number of sequencing reads that are generated. I recommend using the MinION_{MK1C} sequencer which has an integrated computer. The base MinION has to be connected via USB to a computer and problems with the computer having the required specs can result in less reads being generated than one would hope for.

Ontbarcoder pipeline for data analysis: The Ontbarcoder pipeline^{10,11} works well for DNA metabarcoding based on pooled samples of specimens. It just requires a Fastq file with the Nanopore sequence data and a demultiplexing file that includes the barcode expected length, and the primer and barcode information for each specimen. See the paper and manual for more details. To generate the Fastq file, the individual sequence files that are generated can be combined into a single file using the following command line script in Linux: Cat *.fastq>>all.fastq . I use a Linux virtual machine on my laptop. To run this script, open a command line shell, enter the folder with the Nanopore sequence data, and then type the script above in the command line. The file all.fastq will have all of the sequence data from the run. The demultiplexing file can be generated in Excel and saved as a .csv file for use in Ontbarcoder. Once you open the program, the fastq and demultiplexing files can be dragged into the program window and the program run with default settings. The reverse primers and barcodes do not have to be reverse complemented, and the program can recognize Y's in the primer sequences. The sequence data should be all capitals though. The output includes a very useful summary Excel file and folders with reads by specimen.

⁹ <https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-021-01141-x>

¹⁰ <https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-021-01141-x>

¹¹ <https://github.com/asrivathsan/ONTbarcoder>