U.S. Fish & Wildlife Service

Northeast Fishery Center
Conservation Genetics Lab

Protocol: Mitochondrial genome sequencing

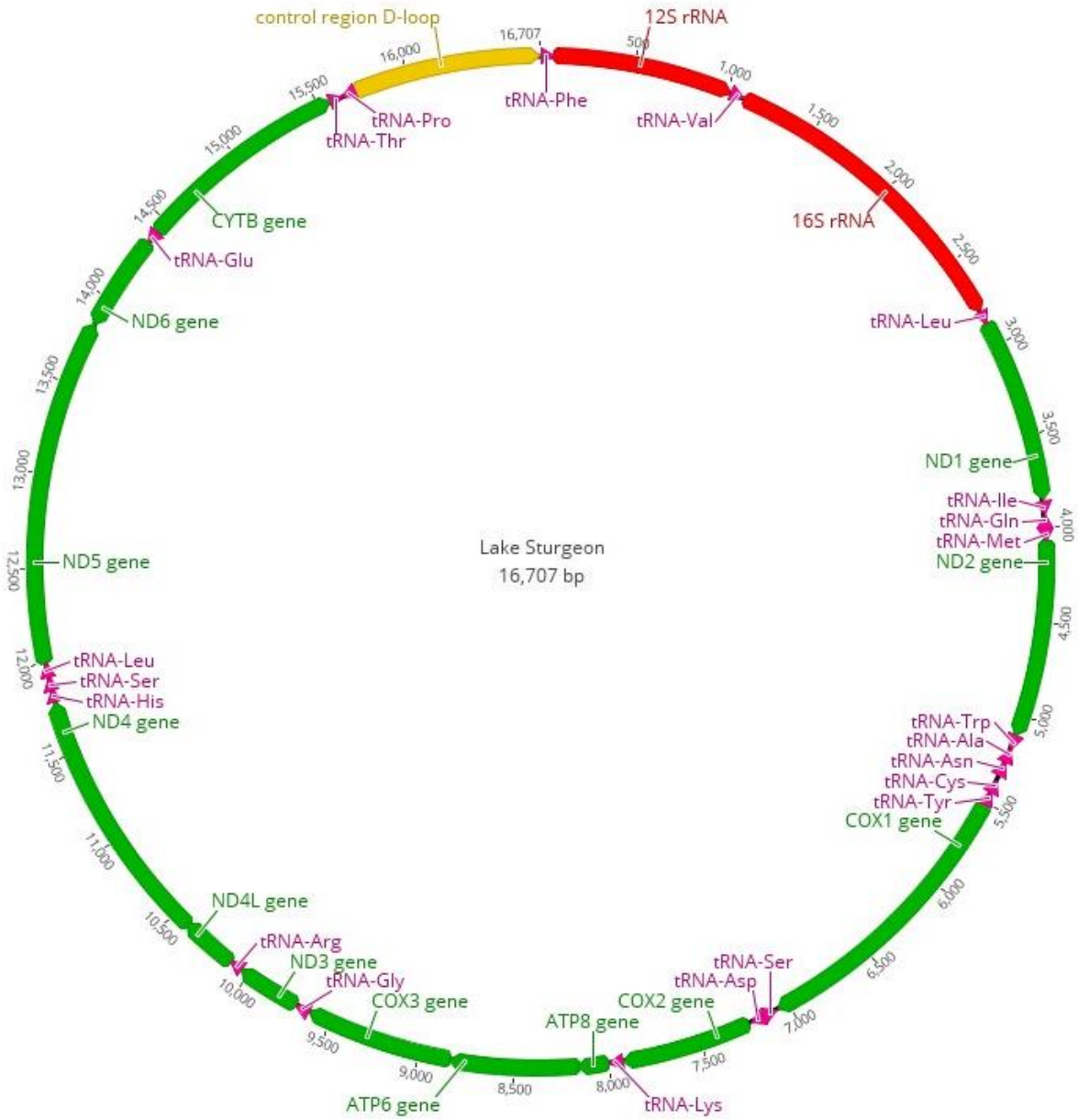

Table of Contents

Overview

Mitochondria are membrane bound organelles found within the cytoplasm of most eukaryotic cells and are responsible for oxidative phosphorylation. This process converts simple sugars and oxygen into adenosine triphosphate (ATP), the cell's main energy source. The number of mitochondrion per cell varies based on cell type and can range from zero (red blood cells) to >2,000 (liver), with an average range of 100-1,000 mitochondrion per cell commonly reported. Each mitochondria in turn contains 2-10 copies of mitochondrial DNA (mtDNA). Mitochondrial DNA is a circular molecule that within freshwater fish ranges in size between 16,000-17,000 bp in length. The generic fish mitochondrial genome contains two ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes and 13 protein coding genes, all of which are involved in the oxidative phosphorylation process. The origin of replication within the mitochondrial genome is located in a noncoding region known as the displacement (D-loop). Mitochondrial DNA is separate from the nuclear DNA and in most eukaryotic organisms is maternally inherited. There are, of course, exceptions to these basic generalizations.

Mitochondrial DNA has long been a genetic region of interest in phylogenetics and evolutionary biology for its usefulness in establishing genetic relationships within and among species. More recently, the emerging field of environmental DNA has targeted mtDNA for use in both single-species targeted assays and metabarcoding applications. Environmental DNA approaches are dependent on isolating and detecting trace amounts of DNA from samples which contain cellular material originating from a mixture of species. The probability of detecting a given species in environmental samples is substantially increased by designing genetic markers that target mtDNA. In the average cell, mtDNA will have hundreds to thousands more copies than a gene located within the nuclear genome. It's this combination of containing adequate genetic information and higher detection probability that makes mtDNA the target of choice for eDNA studies.

Environmental DNA studies require mitochondrial sequence data for marker design and as reference data against which taxonomic assignments for metabarcoding data are made. Current sequence coverage is sufficient only for a 650 bp region of the Cytochrome Oxidase Subunit I (COI) gene – this is the standard animal "barcode region." While this region has proven useful for targeted detection work and species ID, it represents only 4% of the overall data available in the mitochondrial genome. In addition, COI is generally not the most suitable region for metabarcoding studies due to the variation observed at the third base pair position within each codon. Many metabarcoding studies now target the 12S and 16S rRNA genes as they provide enhanced primer design locations over those of protein coding regions. Targeting

rRNA genes in metabarcoding studies relieves some of the primer design issues, but incomplete reference datasets hamper taxonomic assignment of metabarcoding sequencing reads.

Continued development of eDNA applications is dependent on the expansion of mtDNA reference data beyond the standard COI barcode region. Sequencing efforts should focus on obtaining whole mitochondrial genomes to increase both species diversity and capture within species variation that occurs across a species geographical range. There are two primary strategies used to obtain mitochondrial genome information using Illumina next generation sequencing. The first method relies on long-range PCR. In this approach, PCR is used to amplify the entire mitochondrial genome in two to four overlapping fragments prior to pooling PCR product and sequencing. This PCR amplification step enriches the sample to a point where very little genomic DNA is present in the sample during sequencing. This allows for a high degree of multiplexing in the sequencing run and a low per sample cost. However, the need to design and optimize multiple primer sets for the initial PCR amplification and increased bench time are also necessary to consider. This is in contrast to a PCR free approach that uses whole genome shotgun sequencing. Sequencing without prior enrichment results in a high percentage (>99%) of genomic in the resultant sequencing reads. Despite mtDNA being in a much higher copy number per cell, the small size of the mtDNA is vastly outweighed by the quantity of genomic DNA present. Whole genome shotgun sequencing require much lower levels of multiplexing to obtain higher read counts per sample, resulting in a higher cost per genome. This higher cost of whole genome shotgun sequencing is partially offset by labor savings realized by eliminating the long-range PCR steps.

Despite the approach taken, library preparation is nearly identical and requires only an additional pooling step to be added to accommodate long-range PCR. However, the sequencing data output of the two approaches necessitates different bioinformatics strategies to efficiently handle mitochondrial genome assembly. The remainder of this protocol details the steps necessary to obtain mitochondrial genomes using long-range PCR (Part 2) and whole genome shotgun sequencing (Part 3). Each part largely focuses on the bioinformatics steps

necessary to assemble, annotate and submit high quality mitochondrial genomes to GenBank. This protocol is to be used in conjunction with Illumina Nexter XT library preparation guides as detailed within.

Part 1: Mitochondrial genome sequencing using long-range PCR

Primer Design – Eight primers are designed to amplify the entire mitochondrial genome in four overlapping fragments (In some instance six primers are used to produce three fragments). Each primer set should amplify a fragment of approximately 4-7,000 bp with a minimum overlap of 500bp. Primer design protocol described below utilizes Geneious R10.

1. Download all available mitogenomes for chosen order/family group from Genbank and import them into Geneious
   - This can generally be restricted to NC sequences to obtain a non-redundant species list
2. Highlight all genomes to be aligned by ticking the box at the left of each sequence name
   - Ensure that bp 1 is set at the start of tRNA-Phe for all sequences, numbering can be changed under "Sequence" → "Change Residue Numbering" (See annotation section for more details)
3. Select "Tools" → "Align/Assemble" → "Multiple Alignment" and choose the "MAFFT Alignment" option using the default settings
   - The MAFFT algorithm is available as a plugin download
4. Select the new alignment file and open the "Display" tab on the right hand side of the screen. In the display tab, select "Consensus" and set the "Threshold" at desired level (usually 90-100%)
5. Open the "Graphs" tab and tick the "Show Graphs" box
6. Areas of 100% identity will appear green in the identity graph and tend to be located in the 12S and 16S rRNA genes and tRNAs
7. Using the mouse, highlight an area of conservation in the consensus sequence that will serve as a potential priming site
8. Select "Tools" → "Primer" → "Design New Primers"
   a. Tick box for either "Forward Primer" or "Reverse Primer"
      - Forward and reverse primers are designed independently
   b. Tick box for "Include Region"
      - Number range should reflect the region highlighted in step 7
   c. Set "Number of pairs to generate" to 5
   d. Set desired primer characteristics
   e. Tick box "Allow Degeneracy" and set to desired level (generally ≤5)

f. Set “Design primers on:” to “Consensus”
g. Click “OK”

- Step 8 is repeated for each new potential primer location

A set of five primer candidates are displayed if suitable sites were found in the selected region. Continue mapping forward and reverse primers to potential locations along the entirety of the consensus sequence. Match forward and reverse priming locations to identify segments that can be pieced together in four large overlapping fragments. A list of all primer locations and relevant information can be exported to Excel for final primer selection.

Tissue storage and DNA Extraction – Long range-PCR is dependent on relatively long (2,000 – 7,000 base pair) fragments of mitochondrial DNA for successful amplification, resulting in a method that is highly dependent on DNA quality. Care should be taken throughout tissue collection, storage, and DNA extraction to ensure high quality DNA is obtained. Fresh tissue samples (fin clips or other tissue) should be stored at -80°C until DNA extraction. Tissues directly frozen and those preserved in 95% EtOH and stored at -80°C have produce consistent results after 3 years of cold storage. Ethanol preserved tissue stored at room temperature is not recommended for long-range PCR applications. For older samples and those stored at room temperature, whole genome shogun approaches are more appropriate. Numerous DNA extraction protocols are likely compatible with long-range PCR. NEFC routinely uses DNeasy Blood &Tissue kit (Qiagen) with favorable results.

Long-Range PCR – Long range PCR for mitochondrial genome sequencing should be carried out using a high fidelity Taq polymerase. Two enzymes are routinely used: Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs) and HiFi Hot Start DNA Polymerase (Kappa Biosystems). Q5 and Kappa master mix includes dNTPs. Each of the four fragments are amplified in independent PCR reactions as follows:

Master Mix	
Reagent	Concentration
Master mix	1X
10µM F primer	0.5µM
10µM R primer	0.5µM
Template	variable
NF H_2O	variable
Total volume 25µl	

Thermocycler			
3 min	@	98°C	
20 sec	@	98°C	35 cycles
20 sec	@	See Table 1	
3 min	@	72°C	
7 min	@	72°C	

* It is essential that the amplification of each fragment is confirmed on an agarose gel (or other means) prior to pooling. Check that the amplicon is of the correct size and that the amplification is reasonably specific to the desired product. A low level of nonspecific amplification and dimers is acceptable. Sequences of non-mitochondrial origin will be discarded during assembly.

PCR Product Pooling and QC – Pooling is done on a per specimen basis. Each pool contains the four overlapping PCR fragments that make up the entire mitochondrial genome of a specimen.

1. Obtain a concentration (ng/µl) for each PCR amplicon using a Qubit dsDNA BR assay
2. Calculate the DNA concentration in nM based on the size (Table 1) of the PCR amplicons as follows:

$$\frac{\left(amplicon\ concentration\ in \frac{ng}{\mu l}\right)}{\left(660 \frac{g}{mol} * amplicon\ length\right)} * 10^6$$

3. Based on the above calculation, pool the four fragments in equimolar proportions
4. Obtain a concentration (ng/µl) for each pool using a Qubit dsDNA BR assay
5. Dilute pools to a final concentration of 0.2ng/µl
    - The diluted pools are the "Input DNA" for Nextera XT library preparation

Sample Sheet Preparation – Prior to sequencing and library prep, a sample sheet must be created that provides the run instructions and sample information for the MiSeq. Completing the sample sheet prior to library prep will ensure compatible Index combinations are being used. This is completed in the application Illumina Experiment Manager and is available free through Illumina's website.

1. Open Illumina Experiment Manager
2. Select "Create Sample Sheet"
3. Select "MiSeq" → "Next"
4. Select "Other" → "FASTQ Only" → "Next"
5. On the Workflow Parameters page:
    a. Enter Reagent Cartridge Barcode (found on the sequencing reagent cartridge)
    b. Select library prep kit "Nextera XT"
    c. Select Index Reads "2"
    d. Enter experiment name, investigator and description information
    e. Select Read Type "Paired End"
    f. Set Cycle Read 1 and Cycle read 2 to "XXX"

- Cycle reads will depend on the sequencing kit used. For a 300 cycle kit use 151; for a 500 cycle kit use 251; for a 600 cycle kit use 301

g. Tick box "Use Adaptor Trimming"

6. On Sample Selection page:
   a. Select "Add Blank Row" (add one row for each specimen being sequenced)
   b. Enter Sample ID and Sample Name (This can be copied and pasted from Excel) for each specimen
   c. For each specimen, select the corresponding Index 1 (I7) and Index 2 (I5) from the drop down menu that was used in the indexing step of library preparation
   d. Select "Finished"
7. Name and save the corresponding CSV file to a thumb drive
8. Copy the new sample sheet to the "Sample Sheets" folder on the Miseq desktop

Nextera XT DNA Library Prep – Follow the Illumina protocol "Nextera XT Library Prep Reference Guide" (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-reference-guide-15031942-02.pdf) and the "Denature and Dilute Libraries Guide" (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-01.pdf) for library preparation.

Sequencing – The sequencing reagents need to be completely thawed and the flow cell cleaned before loading samples on the MiSeq.

Prepare sequencing reagents and flow cell

1. One hour before the sequencing run is to begin, remove the sequencing reagent cartridge (Box 1 of the MiSeq Reagent Kit) from the freezer and thaw in approximately 2 inches of water. The sequencing cartridge has a submersion line on the side that is not to be exceeded.
2. Remove the incorporation buffer and flow cell from the refrigerator (Box 2 of the MiSeq Reagent Kit)
3. Remove the flow cell from the plastic storage container and discard the storage buffer
4. Rinse the flow cell with DI water and blot dry with clean lens paper (needs to lint-free paper)

5. Clean the flow cell with lens paper and 80% ethanol to ensure the glass surface is streak-free and dry (be mindful not to damage the rubber injection ports)
6. After cleaning, the flow cell can be temporarily stored in the plastic flow cell container which has been emptied of storage buffer

Starting the MiSeq Sequencing Run

1. Shut down the windows operating system and turn the power off to the Miseq for 10 seconds and reboot the entire system
2. On the MiSeq touch screen select the "sequence" button
3. Optional: Select the Base Space option and enter your MyIllumina account information (sequencing data and real time run metrics will be available by logging into your MyIllumina account – data is by default also stored locally on the D drive)
4. Select "next"
5. Load the freshly cleaned flow cell (the flow cell can only be loaded in one orientation, if the door does not close with minimal effort, check the orientation/alignment and try again) and select "Next"
6. Load the incorporation buffer and lower the sipper into place. Ensure the waste reagent reservoir is empty
7. Pierce the foil of the "Load Sample" well on the sequencing reagent cartridge with a clean pipet tip and transfer the entire sample to be sequenced (this is the DAL from Nextera XT library prep protocol) into the well
8. Slide the sequencing reagent kit into the MiSeq and select "Next"
9. Click the sample sheet button and direct the Miseq to the sample sheet created in the Sample Sheet Preparation step above and select "Next"
10. After the MiSeq has completed its pre-run check, select the "Sequence" button to begin the sequencing run
11. After sequencing is complete, perform a Post-Run wash with the optional Template Line Wash

Retrieve Sequence Data – Sequencing data is stored locally on the MiSeq in the D drive and sent to Base Space or local server depending on the setup. To retrieve data stored locally:

1. Select "Start Menu" → " Computer" → " D drive" → "Illumina" → "MiSeq Output" → [folder with corresponding run name] → "Data" → "Intensities" → "Base Calls"
2. In the Base calls folder, a list of the R1 and R2 FASTQ files with corresponding specimen ID names should be present (these are the files to be imported into Geneious)
    - All sequencing reads are sorted based on indexes and trimmed of adaptor and indexes as part of the onboard processing on the MiSeq

Genome Assembly (de novo) – Raw sequencing reads from the MiSeq must be quality- filtered, merged, and normalized before assembly into contigs. This basic workflow generally produces a complete high quality mitochondrial genome. However, this is just one strategy and may not work well with all data inputs.

1. Create a folder with specimen ID name in Geneious
2. Drag and drop the corresponding R1 and R2 FASTQ files into the new folder
3. Tick the box at the left of the sequence name to highlight both R1 and R2 files
4. Select "Sequence" → "Set Paired Ends"
    a. Tick box "Pairs of sequence lists"
    b. Tick box "Forward/Reverse (inward pointing, e.g. Illumina paired end)
    c. Set Expected Distance / Insert Size to 500
    d. Set Read Technology to "Illumina" and "Paired End"
    e. Select "OK"
        - A new file of interleaved paired reads has been created
5. Select " Annotate & Predict" → "Trim using BBDuk" (available as a plugin download)
    a. Tick box "Trim Low Quality"
        - Set drop-down boxes: Trim = Both Ends; Minimum Quality = 20
    b. Tick box "Discard Short Reads"
        - Set drop-down box: Minimum length = 25bp
    c. Tick box "Keep original order"
        - Set option: memory use to 1,000 MB; Custom BBDuk Options: minavgquality=20
    d. Select "OK"
        - Low quality bases have now been trimmed from each read, short reads discarded, and reads with < Q20 average quality discarded
6. Select "Sequence" → "Merge Paired Reads"
    a. Set Merge Rate to "Normal"
    b. Set maximum memory use to 5,000MB
    c. Select "OK"
        - Two files are created, one containing all the newly merged reads and the other with reads that were unmerged
7. Tick the box to highlight the newly-created "Merged Trimmed [specimen name]" read file
8. Select "Sequence" → "Error Correct & Normalize Reads"
    a. Tick box for "Error Correction"
        - Default setting are fine for remaining options

b. Tick box for “Normalization”
    i. Set Target Coverage Level to 60
    ii. Set Minimum Depth to 4
c. Default settings are fine for remaining options
d. Select “OK”
    - A new file of “Normalize & Error Corrected Merged [specimen name]” reads is created

9. Tick the box to highlight the newly-created “Normalized & Error Corrected…” file
10. Select “Tools” → “Align/Assemble”→ “De Novo Assembly”
    a. Set Assembler to “Geneious”
    b. Set Sensitivity to “Custom Sensitivity”
    c. In the results settings
        i. Tick box “Save contigs” and “Maximum” – set maximum to 5
    d. In the advanced settings:
        i. Untick the box “Don’t merge variants with coverage over approximately”
        ii. Change “Maximum Mismatches Per Read to 5%
        - All other settings can be run under the default parameters
    e. Select “OK”
        - A new subfolder is created containing an assembly report, each contig found, and a list of consensus sequences
11. Select the new folder “Normalized & Error Corrected Merged Trimmed [specimen name] Assembly” in the file tree.
    a. Select the contig of the longest length to view the consensus sequence, coverage map, identify graphic and read assembly
    b. If this contig represents a complete mitochondrial genome, it should:
        i. Be between 16,000-17,000 bp in length (most species of fish fall within this range but a few exceptions exist)
        ii. Be circular
            ▪ On the right and left margins of the assembly map, reads are marked with a small directional triangle if they connect the circular molecule
    c. In the “Display” tab on the right of the contig view, tick the box “Consensus” and set to “0% - Majority”
    d. Click on the word “Consensus” in the contig view – this should highlight the entire consensus sequence
    e. Select the “Extract” button in the contig view window
        i. Name the new sequence
        ii. Select “OK”

- The newly-created file is the putative mitochondrial genome and has a green circular file icon to indicate that it is a circular sequence

** The new putative genome must go through a series of QA/QC measures and be annotated prior to submission to GenBank. Proceed to Part 4: Processing consensus sequences for GenBank Submission.

Part 2: Mitochondrial genome sequencing using whole genome shotgun sequencing

Tissue storage – Whole genome shotgun sequencing is not as sensitive to DNA quality as the long-range PCR approach. However, it is still advisable to handle and store tissue in a way that maximized DNA quality. The source tissue is also important. While whole genome shotgun sequencing can be used on any tissue type, our preliminary experiments suggest muscle issue contains higher copy number mitochondrial DNA than do fin clip samples. This difference can be compensated for by ensuring a lower level of multiplexing when running fin clips versus muscle samples.

DNA Extraction – Any routine DNA extraction is likely sufficient for this method, but may contain varying ratios of mitochondrial to genomic DNA. It is possible at the DNA extraction stage to enrich your sample for mitochondrial DNA. At NEFC we use Qiagen Miniprep kits designed for isolation and purification of bacterial plasmid DNA for this purpose. Our preliminary experiments have shown approximately a 10 fold increase of mitochondrial DNA relative to Qiagen Blood and Tissue extractions.

Sample Sheet Preparation – Prior to sequencing and library prep, a sample sheet must be created that provides the run instructions and sample information for the MiSeq. Completing the sample sheet prior to library prep will ensure compatible Index combinations are being used. This is completed in the application Illumina Experiment Manager and is available free through Illumina's website.

1. Open Illumina Experiment Manager
2. Select "Create Sample Sheet"
3. Select "MiSeq" → "Next"
4. Select "Other" → "FASTQ Only" → "Next"
5. On the Workflow Parameters page:
    a. Enter Reagent Cartridge Barcode (found on the sequencing reagent cartridge)
    b. Select library prep kit "Nextera XT"
    c. Select Index Reads "2"
    d. Enter experiment name, investigator and description information
    e. Select Read Type "Paired End"
    a. Set Cycle Read 1 and Cycle read 2 to "XXX"
        - Cycle reads will depend on the sequencing kit used. For a 300 cycle kit use 151; for a 500 cycle kit use 251; for a 600 cycle kit use 301
    f. Tick box "Use Adaptor Trimming"

6. On Sample Selection page:
    a. Select “Add Blank Row” (add one row for each specimen being sequenced)
    b. Enter Sample ID and Sample Name (This can be copied and pasted from Excel) for each specimen
    c. For each specimen, select the corresponding Index 1 (I7) and Index 2 (I5) from the drop down menu that was used in the indexing step of library preparation
    d. Select “Finished”
7. Name and save the corresponding CSV file to a thumb drive
8. Copy the new sample sheet to the “Sample Sheets” folder on the Miseq desktop

Nextera XT DNA Library Prep – Follow the Illumina protocol “Nextera XT Library Prep Reference Guide” (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-reference-guide-15031942-02.pdf) and the “Denature and Dilute Libraries Guide” (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-01.pdf) for library preparation.

Sequencing – The sequencing reagents need to be completely thawed and the flow cell cleaned before loading samples on the MiSeq.

Prepare sequencing reagents and flow cell

1. One hour before the sequencing run is to begin, remove the sequencing reagent cartridge (Box 1 of the MiSeq Reagent Kit) from the freezer and thaw in approximately 2 inches of water. The sequencing cartridge has a submersion line on the side that is not to be exceeded.
2. Remove the incorporation buffer and flow cell from the refrigerator (Box 2 of the MiSeq Reagent Kit)
3. Remove the flow cell from the plastic storage container and discard the storage buffer
4. Rinse the flow cell with DI water and blot dry with clean lens paper (needs to lint-free paper)
5. Clean the flow cell with lens paper and 80% ethanol to ensure the glass surface is streak-free and dry (be mindful not to damage the rubber injection ports)
6. After cleaning, the flow cell can be temporarily stored in the plastic flow cell container which has been emptied of storage buffer

Starting the MiSeq Sequencing Run

12. Shut down the windows operating system and turn the power off to the Miseq for 10 seconds and reboot the entire system
13. On the MiSeq touch screen select the "sequence" button
14. Optional: Select the Base Space option and enter your MyIllumina account information (sequencing data and real time run metrics will be available by logging into your MyIllumina account – data is by default also stored locally on the D drive)
15. Select "next"
16. Load the freshly cleaned flow cell (the flow cell can only be loaded in one orientation, if the door does not close with minimal effort, check the orientation/alignment and try again) and select "Next"
17. Load the incorporation buffer and lower the sipper into place. Ensure the waste reagent reservoir is empty
18. Pierce the foil of the "Load Sample" well on the sequencing reagent cartridge with a clean pipet tip and transfer the entire sample to be sequenced (this is the DAL from Nextera XT library prep protocol) into the well
19. Slide the sequencing reagent kit into the MiSeq and select "Next"
20. Click the sample sheet button and direct the Miseq to the sample sheet created in the Sample Sheet Preparation step above and select "Next"
21. After the MiSeq has completed its pre-run check, select the "Sequence" button to begin the sequencing run
22. After sequencing is complete, perform a Post-Run wash with the optional Template Line Wash

Retrieve Sequence Data – Sequencing data is stored locally on the MiSeq in the D drive and sent to Base Space or local server depending on the setup. To retrieve data stored locally:

1. Select "Start Menu" → " Computer" → " D drive" → "Illumina" → "MiSeq Output" → [folder with corresponding run name] → "Data" → "Intensities" → "Base Calls"
2. In the Base calls folder, a list of the R1 and R2 FASTQ files with corresponding specimen ID names should be present (these are the files to be imported into Geneious)
    - All sequencing reads are sorted based on indexes and trimmed of adaptor and indexes as part of the onboard processing on the MiSeq

Genome Assembly (map to reference) – This basic workflow will generally produce a complete mitochondrial genome if sufficient read depth is available. However, it is just one strategy and

may not work well with all data inputs. It is common practice to use multiple assembly strategies to create draft genomes before settling on a final version.

File import and quality control

1. Create a folder with specimen ID name in Geneious
2. Drag and drop the corresponding R1 and R2 FASTQ files into the new folder
    a. Geneious will prompt you to select the Read Technology and how the R1 and R2 files are to be imported
        - Reads can be either be imported as paired ends (select, "Paired End (inward pointing)") or unpaired (select, "Don't pair")
        - Either option appropriate – I prefer to use unpaired reads
        - Click "OK"
3. Two files will appear in Geneious (If the reads were paired the files are combined into one interleaved file – the remaining assembly is the same for this option, just select the one input file). Tick the box at the left of the sequence name to highlight both R1 and R2 files.
4. Select "Annotate & Predict" → "Trim using BBDuk"
    a. Tick box "Trim Low Quality"
        - Set drop down: Trim = Both Ends
        - Set drop down: Minimum Quality = 20
    b. Tick box "Discard Short Reads"
        - Set drop down: Minimum Length = 20
    c. Click "OK"
        - Two new files are created with file names indicating that the sequences have been trimmed. The trimmed files are used as input for mapping sequences to a reference genome.

Mapping reads to a reference sequence

The number of reads required in whole genome shotgun sequencing generally precludes de novo assembly of sequencing reads. It is far more efficient to first map the sequencing reads to a reference genome to filter out sequences of mitochondrial origin. The reference genome should be from the species you are trying to assemble or a closely related species.

1. Obtain an appropriate reference genome. This genome can be downloaded from GenBank.

a. Go to GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and search the species name you are looking for
b. In the left column of the results page, filter by “mitochondrion” and “RefSeq”
    - Any species with a mitochondrial genome should have a reference sequence. These are denoted within GenBank with an accession number that begins NC_
    - If no reference sequence is available for your species, choose a closely related species (preferably same genius, but family should work) until one is found
c. Select and open the reference sequence
d. In the top right, select he drop down menu “Send to”
    - Select = Complete record
    - Choose Destination = File
    - Format = GenBank (full)
    - Click “Create File”
        - By default this file should be saved in you downloads, drag and drop the reference file into the Geneious folder that contains your sequencing read data. A new file should appear with a green circle icon indicating that the file is a circular sequence

2. In the tick box next to the file names in Geneious, select the new reference genome and the “Trimmed [specimen name]” R1 and R2 files.
    a. Select “Tools” → “Align/Assemble” → “Map to Reference”
        - In the new Map to Reference Dialogue box ensure the file path in the “Reference Sequence” box in pointed to the correct reference sequence. A new file path can be chosen if necessary.
    b. Click “OK”
        - Default parameters are acceptable for this draft assembly
    c. A new subfolder is created with a file name that indicates the sample ID and the sequence it was assembled to
        - This file folder contains the assembly (contig) and a list of all the sequencing reads used in the assembly. The “used reads” file contains the mitochondrial reads which are now isolated from those originating from genomic DNA
    d. Open the new subfolder and select the “contig” file
        - This is the first graphical view of the draft genome
    e. If the contig represents a complete mitochondrial genome, it should:
        i. Have 100% coverage across the reference genome

- The coverage map at the top of the alignment provides a graphical display of coverage across the genome – check for spots of low or no coverage
- If using a more distantly related genome for mapping, it is possible to have the appearance of less than 100% even though your genome is complete

ii. Be circular
- On the right and left margins of the assembly map, individual reads are marked with a small directional triangle if they connect the circular molecule

d. Extract you draft genome
   i. In the "Display" tab on the right of the contig view, tick the box "Consensus" and set to "0% - Majority"
   ii. Click on the word "Consensus" in the contig view – this should highlight the entire consensus sequence
   iii. Select the "Extract" button in the contig view window and name the new sequence
      - The newly-created file is the "mapped to reference" draft mitochondrial genome and should have a green circular file icon

e. This represents a first draft genome. Depending on the situation, there could be several ways to proceed. Below are two that should work well for most samples.
   i. If the draft genome was assembled against the same species and is of sufficient coverage, complete a de novo assembly of sequences contained in the "used reads" file (see below for de novo assembly instructions)
      - The used reads are only the sequences identified as of mitochondrial origin. If less than 10,000 reads are available, de novo assembly will be time efficient. Greater number of reads can be normalized prior to assembly to reduce computation time (the normalization process is described in the long-range PCR assembly section)
   ii. If the draft genome was assembled against a distantly related species, use the draft as the reference genome and map the trimmed reads to it
      - This is done by repeating the mapping reads to a reference sequence instructions. Aligning to the first draft will generally identify additional sequences and produce a higher quality alignment. A de novo assembly of the used reads can be completed as above.

De novo assembly of used reads – Mapping to a reference genome is necessary to identify the mitochondrial reads and isolate them from sequences that originated from genomic DNA. This process can occasionally introduce bias into the consensus sequence, especially in species that have varying numbers of tandem repeats in the D-loop. This bias can be minimized by obtaining a de novo assembly of the mitochondrial reads.

- Highlight the "used reads" file that you would like to de novo assemble
- Select "Tools" → "Align/Assemble"→ "De Novo Assembly"
    f. Set Assembler to "Geneious"
    g. Set Sensitivity to "Custom Sensitivity"
    h. In the results settings
        i. Tick box "Save contigs" and "Maximum" – set maximum to 5
    i. In the advanced settings:
        i. Untick the box "Don't merge variants with coverage over approximately"
        ii. Change "Maximum Mismatches Per Read to 5%
    - All other settings can be run under the default parameters
    j. Select "OK"
    - A new subfolder is created containing an assembly report, each contig found, and a list of consensus sequences
        - Select the new folder "used reads Assembly" in the file tree
    k. Select the contig of the longest length to view the consensus sequence, coverage map, identify graphic and read assembly
    l. If this contig represents a complete mitochondrial genome, it should:
        i. Be between 16,000-17,000 bp in length (most species of fish fall within this range but a few exceptions exist)
        ii. Be circular
            - On the right and left margins of the assembly map, reads are marked with a small directional triangle if they connect the circular molecule
    m. In the "Display" tab on the right of the contig view, tick the box "Consensus" and set to "0% - Majority"
    n. Click on the word "Consensus" in the contig view – this should highlight the entire consensus sequence
    o. Select the "Extract" button in the contig view window
        i. Name the new sequence
        ii. Select "OK"
            - The newly-created file is the putative mitochondrial genome and has a green circular file icon to indicate that it is a circular sequence

- As mater of curiosity and informal quality control, the original draft genome (draft assembled the original reference) and the new de novo draft can be aligned. Any differences should be investigated. To align sequences, the orientation and sequence origin will need set (see below).

** The new putative genome must go through a series of QA/QC measures and be annotated prior to submission to GenBank. Proceed to Part 4: Processing consensus sequences for GenBank Submission.

Part 3: Processing consensus sequences for GenBank submission

Set strand orientation and sequence origin – Mitochondrial genomes should be presented in the heavy strand orientation and the residue numbering standardized so they all have the same origin.

1. Download a mitochondrial genome from GenBank to use as an annotation reference
   - The annotation reference needs to be stored in a separate file folder (only one reference per file folder)
   - At this point any somewhat closely-related fish species can be used – we will do the proper annotation in later step
2. Tick the box next to the newly-created mitogenome you want annotate
   - The sequence should now be highlighted and the "Sequence View" window active
3. In the "Sequence View" window, select the "Live Annotate & Predict" tab on the right side of the window
   a. Tick the "Annotate From…" box
   b. Click the "Select a folder" button
   c. In the "Select Features Folder" window, direct Geneious to the folder created in Step 1 that contains the genome to be used as an annotation reference
   d. Select "OK"
      - Some annotations should now be present on your new genome and visible in the sequence view window
   e. Select the "Similarity" slider and move to around 70%
      - The sequence should now be displaying annotations (<u>Do Not</u> apply these annotations)
4. Ensure the new mitogenome sequence is displaying the heavy strand

- The majority of genes are encoded on the heavy strand and annotations should be in the 5´ to 3´ orientation, i.e. annotations should have a pointed end on the right hand side and a blunt end on the left (If true, continue to step 5)
    a. To change the orientation of the sequence, click the "RC" button in the Sequence View window
        i. In the "Reverse Compliment" window, select the "Reverse compliment the entire sequence" option
        ii. Select "OK"
        - The sequence should now have flipped direction and be displaying the heavy strand
5. Change the residue numbering so that base pair 1 is at the start of tRNA-Phe
    a. Locate tRNA-Phe and click in on it
    b. Zoom in using the controls at the right of the sequence view window until the 5´ end is in view
    c. Click to locate the curser just before the first base pair in tRNA-Phe
    d. With the curser in place, select "Sequence" → "Change Residue Numbering"
        i. In the Change Numbering window, click "OK"
            - The first base pair of tRNA-Phe should now be residue number 1
    e. Select "Save" in the "Sequence View" window

Species ID quality check – It is important that our new mitochondrial genome matches the species ID of the source material. Field identifications can sometimes be wrong and sample tubes can be mislabeled during processing.

1. With the temporary annotations visible, right click the COI gene and copy the sequence data
2. Open you web browser and go to the Bar Codes of Life Data System identification page (http://www.boldsystems.org/index.php/IDS_OpenIdEngine)
3. Select the "Species Level Barcode Records" option
4. Paste the COI sequence in the dialog box and click submit
5. Results will display a species level match or indicate that multiple species match your query sequence
    a. This information can be used to confirm the field ID and ensure integrity of the lab processing procedures. Any sequences that produce unexpected results need further scrutiny before submission to GenBank
6. If you are comfortable with the identity of the mitochondrial genome, rename the sequence and copy it to a new file folder

- This new file folder will be used to collect all the newly obtained genomes that have a standardized direction, origin and COI species confirmation – we will later batch annotate and prepare these for GenBank submission

Annotation – After assembly, mitochondrial genomes need to be annotated, presented in the heavy strand orientations and the residue numbering standardized. Annotations can be cross referenced with those obtained using the MitoFish annotator (http://mitofish.aori.u-tokyo.ac.jp/annotation/input.html)

1. Download a mitochondrial genome from GenBank to use as a annotation reference
- The annotation reference needs to be stored in a separate file folder (only one reference per file folder)
- The reference should be the same species as you are annotating or something closely related
2. Go the new folder created in step 6 of the species ID quality check step
3. Tick the box next to the sequence(s) you want to annotate (Sequences can be annotated in batches of closely related species)
- The sequence should now be highlighted and the "Sequence View" window active
4. In the Sequence View window, select the "Live Annotate & Predict" tab on the right side of the window
    a. Tick the "Annotate From…" box
    b. Click the "Select a folder" button
    c. In the "Select Features Folder" window, direct Geneious to the folder created in step 1 that contains the genome to be used as an annotation reference
    d. Select "OK"
    - Some annotations should now be present on your new genome and visible in the sequence view window
    e. Select the "Similarity" slider and move to around 90%
    - With few exceptions, fish species will include 2 rRNA genes (red annotations), 13 protein-coding genes (green annotations), 22 tRNAs (pink annotation) and the control region D-loop (orangey annotation) – an additional 13 coding DNA sequence (CDS) annotations (yellow) are included with GenBank submissions
    - The control region will generally be the lowest similarity match, if it is present all other annotations have likely been mapped

- Setting the similarity threshold to low risks creating multiple annotations in the same location – especially in the tRNAs

f. Select "Apply" to transfer the annotations to your new genome

5. Select "Save" in the Sequence View window to save the newly-transferred annotations

Quality Assurance (Annotations) – The majority of annotations that are copied across from the reference genome are correct, but some corrections will still be needed. The start and stop codons in the protein-coding genes need to be checked for accuracy. In addition, the translation exceptions in CDS annotations often need some adjusting. It is also helpful to know the stop and start codon sequences for the vertebrate mitochondrial genome.

Vertebrate mitochondrial codons			
Start	att	Stop	taa
	atc		tag
	ata		aga
	atg		agg
	gtg		

1. Tick the box next to each genome to be evaluated so that they are highlighted and visible in the "Sequence View" window
2. Click the display tab on the right side of the Sequence View window
    a. Click the "Nucleotides" box
    b. Click the "Translation" box
        i. Set Translation Options: Frame = By selection or annotation; Genetic Code = Vertebrate Mitochondrial (transl_table 2)
3. Scroll through the selected sequences and ensure each protein-coding gene starts with a start codon, has no internal stop codons, ends with a stop codon or that a translation exception is present if part of the stop codon is not encoded and is instead generated post-transcriptionally by polyadenylation of the mRNAs.
    - Hovering the curser over the gene annotation will bring up relevant information

Quality Assurance (Species Origin) – Prior to submission to GenBank, it is essential to ensure that the sequences are representative of the species that are linked to the specimen ID. We already checked that individual sequences have a COI barcode that matches the expected results. COI barcodes in some closely related species are not discriminatory. We can provide an additional quality check using the complete mitochondrial genome by performing a basic cluster analysis. Any sequence producing unexpected results needs further validation before GenBank submission. Additional mitochondrial genomes from GenBank can be downloaded and used in this analysis to provide better resolution

Check species ID by cluster analysis:

1. Align all new mitogenomes and additional genomes (closely related and a few distantly-related species obtained from GenBank) using MAFFT
    a. Tick the box next to each sequence to be aligned so that they are highlighted
    b. Select "Tools" → "Align/Assemble" → "Multiple Alignment"
    c. In the "Alignment" window, select "MAFFT" → "OK" (Default parameters are fine)
        - A new alignment file will be created
2. Tick the box next to the new alignment file to highlight it
3. Select "Tools" → " Tree"
4. In the "Tree" window, select:
    a. Genetic Distance Model = "Tamura-Nei"
    b. Tree Build Method = "Neighbor Joining"
    c. Outgroup = "No outgroup"
    d. Select "OK"
5. Select the new Tree file
    - Note – this is not a proper phylogenetic tree, but rather a quick way to flag any sequence that may be labeled incorrectly or potentially chimeric. Species should cluster with the same species and/or closely-related species. Sequences that do not cluster on a species-specific basis need further evaluation prior to submission to GenBank.

Quality Assurance (submit to GenBank) – Batch submission of mitochondrial genomes to GenBank can be done within Genious using the Submit to GenBank tool. This submission tool will also do an automated quality check on your sequences and generate errors and warnings with information regarding the potential problem. It is best to provisionally "Submit" sequence one at a time for quality checking. You can then edit any errors present. These errors are generally due to annotation issues and not the actual sequences. Once all your sequences are free of errors and you are confident in their quality, they can be formally submitted in a single batch to GenBank. Other sequence submission tools such as Sequin can also be used for batch submission (https://www.ncbi.nlm.nih.gov/Sequin/)

Initial Quality Check

1. Tick the box next to the first sequence you wish to provisionally submit for quality check
2. Select "Tools" → "Submit to GenBank"
3. In the "Submit to GenBank" window:

    a. Enter a Submission Name and select “Edit Publisher Details…”
        i. Complete the publisher details as appropriate (This data will be used to popular the GenBank flat file)
        ii. Select “OK”
            - This data only has to be entered once and can be used for all sequences being submitted within a particular project
4. Tick “Save a local file”
5. In the “Fields” box, complete the dropdown fields:
    a. Project Name = your choice
    b. Specimen Voucher = column heading from genome file that contains corresponding information
    c. Molecule Type = Genomic DNA
    d. Genetic Location = Mitochondrion
    e. Sequence ID = column heading from genome file that contains corresponding information
    f. Organism = column heading from genome file that contains corresponding information
    g. Genetic Code = Vertebrate Mitochondrial
        - This data will be the same for each sequence checked and can be used during the final submission
6. Tick the box “Include Features/Annotations”
7. Select “OK”
8. If a “Genome Reference” window appears asking about database reference fields… select “Remove”
9. If errors and warning are present, a new “Submission Warnings” window will appear and provide information regarding the problems
10. Note the problem and feature each error or warning is associated with
11. Repeat these steps for each sequence to be submitted

Fixing Errors and Warnings – Several errors and warnings are common and easily fixed. Below are examples of the most common problems and how to fix them

1. Error Missing stop codon FEATURE
    a. In the “Sequence View” window of the sequence in question, navigate to the position that should contain the stop codon and zoom in so the residue number is visible
    b. Note the residue number for the first base pair of the stop codon (This is usually a partial codon with only a T of TA present)

c. Double Click the CDS annotation to bring up the Edit annotation window
d. Highlight “Transl_except:…” and then select the “Edit” button
e. In the “Edit transl_except” window, change the “Value” field to reflect the correct residue position for the first stop codon residue (only edit the position number)
f. Select “OK”
g. Select “OK”
h. Save changes made

2. Warning Unparsed transl_except qual (but protein okay)
    a. In the “Sequence View” window of the sequence in question, navigate to the position that should contain the stop codon and zoom in so the residue number is visible
    b. Note the residue number for the first base pair of the stop codon (This is usually a partial codon with only a T of TA present)
    c. Double click the CDS annotation to bring up the “Edit annotation” window
    d. Highlight “Transl_except:…” and then select the “Edit” button
    e. In the “Edit transl_except” window, change the “Value” field to reflect the correct residue position for the first stop codon residue (only edit the position number)
    f. Select “OK”
    g. Select “OK”
    h. Save changes made
3. Warning Circular topology without complete flag set BIOSEQ – This warning is always present and should not affect the GenBank submission

GenBank Submission –Now that all the errors and warnings (save the circular topography warning) have been resolved, samples can be submitted in a large batch.

1. Tick the box to highlight each sequence you want to submit to GenBank
2. Select “Tool” → “Submit to GenBank” to open the “Submit to GenBank” window
    - At this point all the data entered in to the Publisher Details and Fields (as completed when checking the quality of individuals sequences above) should still be as entered
    - Additionally, data in the Fields section should have remained unchanged and be as described above
3. Ensure Submission Type = “Batch Submission”
4. Tick the “Upload New Submission” option
5. Under Account, Tick the “Submit using the Geneious BankIT FTP” account

6. Select “OK”
7. The submission Warnings page will appear detailing the circular topography warning (Proceed if this is the only remaining warning for each sequence)
8. Select “Submit to GenBank”
    - An email will be sent confirming the receipt of the sequences and GenBank accession numbers are generally assigned within a week

Primer sets for use in long-range PCR

Long-range primers used to amplify complete mitochondrial genomes. Primers are designed to amplify all species within an Order or Family group. *Esociformes and Gobiidae primer sets have not yet been fully validated and their overall utility remains questionable.

	Primer	Sequence (5´-3´)	Length (bp)	Annealing Temp °C	Annealing Location	Approximate Size (bp)
Order Acipenseriformes	ACI_R1-F	GTTGTTAATTCAACTATAAAAACC	24	57	tRNA-Glu	4285
	ACI_R1-R	TTCATTTAAAAGACAAGTGATTAC	24		16S rRNA	
	ACI_R2-F	AACCTAACGAGCCTAGTAATAG	22	58	16S rRNA	5804
	ACI_R2-R	GTCTTGGAATCCTAATTGTG	20		COII	
	ACI_R3-F	ATCCTACAAAATCTTAGTTAAC	22	56	tRNA-Asn	6614
	ACI_R3-R	AGAATTAGCAGTTCTTAGTG	20		tRNA-Ser	
	ACI_R4-F	ATTTCGGCTCAACTAATTAT	20	57	tRNA-Arg	5633
	ACI_R4-R	AGTTTAATGTAGAATCTTAGCTTT	24		tRNA-Pro	
Family Ictaluridae	ICT_R1-F	TCAGACCCACCTAGAGGAGC	20	65	12S rRNA	6595
	ICT_R1-R	GCCGCGTCTTGGAATCCTAG	20		COII	
	ICT_R2-F	AGATGAGAAGGCCTCGATCCT	21	65	tRNA-Asn	6690
	ICT_R2-R	TTGGTTCCTAAGACCAAYGGATGA	24		tRNA-Leu	
	ICT_R3-F	AAGACCTCTGATTTCGRCTCAGA	23	67	tRNA-Arg	5524
	ICT_R3-R	TCTCCGGATTACAAGACCGG	20		tRNA-Thr	
	ICT_R4-F	TAACCAGGACYAATGACT	18	65	tRNA-Glu	3213
	ICT_R4-R	CTTACCATGTTACGACTTG	19		12S rRNA	
Family Salmonidae	SAL_R1-F	CTATATACCACCGTCGTC	18	55	12S rRNA	4845
	SAL_R1-R	AATGTCTTTGTGGTTGG	17		COI	
	SAL_R2-F	AAGTCCCCTCAATTCTAG	18	55	tRNA-lle/tRNA-Gln	4244
	SAL_R2-R	GCTTARTGTCATGGTCAG	18		ATP6/ATP8	
	SAL_R3-F	GTTAGCCTTTTAAGCTAAAG	20	55	tRNA-Lys	6451
	SAL_R3-R	GTGGTTTTTCAAGTCATTA	19		tRNA-Glu	
	SAL_R4-F	GCRCAATTTGGACTTC	16	55	ND5	4829
	SAL_R4-R	TAGAGAATGTAGCCCATT	18		12S rRNA	

*Esociformes primer have not been fully validated

	Primer	Sequence (5´-3´)	Length (bp)	Annealing Temp °C	Annealing Location	Approximate Size (bp)
Family Centrarchidae and Percidae	CP_R1-F	GCAATCACTTGTCTTTTAA	19	56	16S rRNA	5794
	CP_R1-R	CTTAAAAGGCTAACGCTA	18		tRNA-Lys	
	CP_R2-F	TTACCGCTCTGTCACT	16	56	tRNA-Ser	4822
	CP_R2-R	AGTTTTTGGTTCCTAAGAC	19		tRNA-Leu	
	CP_R3-F	CCAAGGAAAGATAATG	16	53	tRNA-Gly/ND3	6354
	CP_R3-R	AATAGTTGTCCCTCAC	16		D-Loop	
	CP_R4-F	CCCAAAGCTAGGATTCTA	18	55	tRNA-Pro	3528
	CP_R4-R	TAGATAGAAACTGACCTGGA	20		16S rRNA	
Family Cyprinidae	CYP_R1-F	TAAAACTCGTGCCAGCCACC	20	61	12S rRNA	4959
	CYP_R1-R	TTGTAGGATCGAGGCCTTCC	20		tRNA-Asn	
	CYP_R2-F	AAGCTTTCGGGCCCATACC	19	65	tRNA-Met	6048
	CYP_R2-R	TCTGAGCCGAAATCAGAGGTC	21		tRNA-Arg	
	CYP_R3-F	AGCCCATGACCMCTAACCGGA	21	65	tRNA-Gly/ND3	4752
	CYP_R3-R	DGTTTTTCGTAGGCTTGCCAT	21		CytB	
	CYP_R4-F	TTGGTCTTAGGAACCAAAAACTCT	24	65	tRNA-Leu	5272
	CYP_R4-R	CCGTCAGGTCCTTTGGGTTT	20		12S rRNA	
*Order Esociformes	ESO_R1-F	TACACATGCAAGTCTCCGCA	20	65	12S rRNA	5065
	ESO_R1-R	GTGTCTGGGTTGCATTCAGA	20		tRNA-Ala/tRNA-Asn	
	ESO_R2-F	AAGCTTTYGGGCCCATACCCC	21	65	tRNA-Met	4823
	ESO_R2-R	ATGCGTGTGCTTGGTGKGCC	20		COIII	
	ESO_R3-F	AAGGRAGGAATTGAACCCCCAT	22	65	tRNA-Ser	7323
	ESO_R3-R	CAACGGTGGTTTTTCAAGTCAT	22		tRNA-Glu	
	ESO_R4-F	TGGTCTTAGGAACCAAARACTCT	23	65	tRNA-Leu	5453
	ESO_R4-R	GAACAGGCTCCTCTAGGTGG	20		12S rRNA	

*Gobiidae primer have not been fully validated

	Primer	Sequence (5´-3´)	Length (bp)	Annealing Temp °C	Annealing Location	Approximate Size (bp)
Family Clupeidae	CLU_R1-F	ACCAAAAGTTTAACGGCCGC	20	65	16S rRNA	4980
	CLU_R1-R	GGGGTTCGATTCCTCCCTTT	20		tRNA-Ser	
	CLU_R2-F	CCTTCAAAGCTCCAAGCAGG	20	65	rRNA-Trp	4917
	CLU_R2-R	CTGAGCCGAAATCAGAAGTCT	21		tRNA-Arg	
	CLU_R3-F	TACGTCTCYATCTACTGATGAGGATC	26	65	tRNA-Gly	6080
	CLU_R3-R	GCTTTGGGAGTTAGAGGTGGA	22		tRNA-Pro	
	CLU_R4-F	YGAAAAACCACCGTTGTTATTCAA	24	65	tRNA-Glu	4888
	CLU_R4-R	GAACCCTTAATAGCGGCTGC	20		16S rRNA	
Family Catostomidae	CAT_R1-F	CCCGTCACTCTCCCCTGTTA	20	65	12s rRNA	6196
	CAT_R1-R	AAGGAAGTGGCAGAGTGGTT	20		tRNA-Ser	
	CAT_R2-F	CTCTGTCTTCGGGGCTACAA	20	65	tRNA-Tyr	6517
	CAT_R2-R	TTGCACCAAGAGTTTYTGGTTCC	23		tRNA-Leu	
	CAT_R3-F	AAGACCTCTGATTTCGRCTCAGA	23	65	tRNA-Arg	5611
	CAT_R3-R	CAGGGGTGGGAGTTAAAATCT	21		tRNA-Pro	
	CAT_R4-F	TGAAGAACCACCGTTGTYATTCA	23	65	tRNA-Glu	3792
	CAT_R4-R	TAGGCAACCAGCTATCACCA	20		16s rRNA	
*Family Gobiidae	GOB_R1-F	AGCCACCGCGGTTATACG	18	65	12S rRNA	4941
	GOB_R1-R	TTTGTAGGATCGAGGCCTTCC	21		tRNA-Asn	
	GOB_R2-F	GCTTCCACTACACCACTTCC	20	65	tRNA-Glu	7978
	GOB_R2-R	CCAAGAGTTTTTGGYTCCTAAGACC	25		tRNA-Leu	
	GOB_R3-F	HAAAATATTTGATTTCGGCTCAAAAG	26	50	tRNA-Arg	7121
	GOB_R3-R	AGAACAGGCTCCTCTAGG	18		12S rRNA	

Contact information

Aaron Maloy
U.S. Fish & Wildlife Service
Northeast Fishery Center
227 Washington Avenue
Lamar PA, 16848
570-726-4247 x157
aaron_maloy@fws.gov