LAB 4: DNA Barcoding of Fish larvae

A. Overview: Whose larvae?

Marine invertebrate and fish larvae are often difficult to identify at early stages of development. Many of the morphological characters used for identification of adults are absent in early life-history phases, and it can be difficult to identify individual larvae to species. For vertebrate animals like fish, later-stage larvae in some groups can be identified to the species level by experts in taxonomy and systematics. Here on Oahu, we are lucky to have one such person: Bruce Mundy, who works for the National Oceanic and Atmospheric Administration’s (NOAA) Pacific Islands Fisheries Science Center (PIFSC). Bruce has over 40 years experience in identifying larval fish from all over the world, and has generously provided us with copies of his identification guide representing all orders of bony (teleost) fishes. Characters used for identification include: overall body length to width ratio, shape and placement of the eye, the length and shape of the gut, the placement of the fins, pigment patterns, and number of muscle segments. Unfortunately, larvae of many fish family are not well studied and therefore hard to identify to the genus or species level, and many tropical fish family or genera contain complexes of closely related species that give us no morphological clues to their species identity.

Technical identification problems make the most basic ecological questions related to biology and ecology of marine larvae difficult to answer. For example, how can we understand how larval supply changes in time and space if we cannot identify larvae to species? How can we know how stocks of economically important fish populations are doing if we cannot measure larval production? A solution to this problem is to use information encoded in the DNA of every organism. We can now extract and sequence the DNA from even small samples like a fish egg quickly and cheaply, and use this information in combination with world-wide databases with similar sequence information (e.g., “GenBank” or BOLD) to make precise species identifications. In some cases, we can assign individuals to specific populations that may live in geographically isolated locations.

To barcode, or not to barcode…

The description and naming of new species requires years of formal training within a specific group of organisms. For example, there are “Goby” specialists, who focus only on the grouping and naming of goby species. These highly trained specialists are called “alpha taxonomists,” and it is they who actually give names to, and set the criteria for identification of species. Traditional methods of alpha taxonomy require an assessment of variation in morphological characters between hypothesized species and between populations within a species’ geographic range. Inexpensive sequencing technology has opened a door to the possibility that non-specialists (like us!) can identify species strictly from DNA sequences. Non-specialists might also be able to identify new species based on DNA sequence information alone! This is the essence of a new approach to alpha taxonomy called “DNA barcoding”. DNA barcoding draws on an analogy to universal product codes (UPCs). These are the codes used by store clerks to identify and check out goods. The clerk reads the UPC bar code from the package with a laser-based optical reader, and a computer translates the “bars” into a number, which references one product. Each UPC identifies a unique product that a retailer can use to price and inventory stock. DNA barcoding assumes that DNA acts in a similar fashion, with unique sequences identifying different species. The concept is simple to grasp and has intuitive appeal, but will it work for the often fluid boundaries of biological species? Questions as to the utility of DNA barcoding have been under intense debate and the outcome of this debate will certainly change the future of systematics and how we catalogue the diversity of life on earth. In this lab we will get a taste for how DNA barcoding is actually done, revealing it’s strengths and weaknesses.

Objectives
There are two major goals of this three-week lab. The first is to introduce you to the technology and methods of DNA sequence data applied to the specific problem of identifying the fish larvae sampled during the plankton cruises. We will extract, amplify, sequence, and analyze a segment of the mitochondrial protein-coding gene known as COI (Cytochrome oxidase I). This stretch of DNA (500-600 nucleotide base pairs, or bp) is typically highly variable between species, making it an excellent marker to identify species, and in some cases populations, and is used in DNA barcoding of most animal groups. We will use genomic tools associated with an international database of DNA sequence data called “GenBank” as well as the species identification tool on the Barcode of Life Data systems website “BOLD”. In some cases we will achieve very close matches, meaning that “GenBank” or “BOLD” contain reference sequences for our larval species. In others, we may only get similarity at the family or genus level because no one has sequenced this gene for that particular species yet. In a nutshell, we will be doing real scientific research, and making new discoveries with this lab.

A second goal is to get you thinking about how the DNA barcoding approach can be used to identify previously described species, and perhaps new species that are unknown to science. Is sequence data by itself likely to be informative for making claims to new species? Are there different cases where this approach is more useful? What are potential technical problems that might limit the application of these methods? By the end of this lab you should be able to answer these questions and you will have the opportunity to discuss them in your first formal report of this course.

B. Schedule of Lab Activities and Homework

!! Wearing close-toed shoes during weeks I and II will be enforced!!

Week I: The first lab requires you to work for ca. 1h in lab before your regular lab time (check schedule on Laulima when you signed up during lab the week before), in addition to your regularly scheduled afternoon lab period.

Reading before lab: Puncher et al. 2015 & Lab manual

1. Sketch 2 fish larvae under dissecting microscope
2. Photograph all your assigned fish larvae
3. DNA extractions of all your assigned samples
4. PCR amplification of COI

Week II: Regular lab period

Reading before lab: Hebert et al. 2015, Will & Rubinoff 2004, & lab manual

1. Agarose electrophoresis to confirm amplification of COI
2. Debate/Discussion on DNA Barcoding
3. Tour ASGBP sequencing facility in Snyder Hall
4. Run “exosapit” incubation to prepare for sequencing
5. Submit samples for sequencing

Week III: Regular lab period

Due: Sequence analysis exercise, Drafts of Introduction and Methods for Formal Report

1. Check quality and edit sequences
2. Analyze sequences using “GenBank” and “BOLD” databases
3. Compile species lists for samples with depth information
4. Data compilation for lab report

Week IV: Regular lab period
Due: Drafts of Result and Discussion for Formal Report

1. Peer review of Introduction, Methods, Results
2. Discussion of how to make sense of large data sets, including results of compiled class data
3. Draw graphic representation of Formal Report
4. Peer Review of Discussion

C. Methods

You will be working with the same lab partner for all three weeks of this lab. We have sorted all the fish larvae into individual samples. Your team will be responsible for sketching, extracting and sequencing a subset of the larvae. The class as a whole will be sequencing the entire data set. It is EXTREMELY important that you keep track of your sample numbers through all three labs. Use Data Sheet 1 to keep track of what you do with each sample and the results of each protocol. Use the “Notes” section if you need extra room. Make sure to record information such as who did a particular sketch, the use of a protocol different than that outlined in this lab manual and to note any mistakes such as accidentally switching tubes, losing a sample, or a pipetting error. Before the extraction, fish larvae need to be handled with care. Unnecessary rough treatment breaks down the sample, reducing DNA quality and quantity. **We are depending on your best laboratory technique to make these new discoveries. We think you are up to it!**

Notes on pipetting technique: Before starting, review and refresh your pipetting techniques. **If you have not taken Biol 275L or a similar course yet and have never used a micro pipet before or need refreshing, ask your TA.** We will be working with small volumes today (1 µl in some cases) so be sure you are comfortable picking up and ejecting these volumes. **Never re-use tips in this lab unless instructed to do so.**

*Note:* The pipets in lab are set by rotating the plunger button clockwise or counter-clockwise. The digital readout gives usually 1.0 µl increments (1 microliter = 1/1000 milliliter or ml), but this might depend on the brand and size of pipet. **Ask your TA for clarifications if you are unsure.**

C.1. Protocol for Week I:

*Reading before Lab: Puncher et al. 2015 & Lab manual*

I. Pre-Lab session (according to sign-up):

Check the schedule on Laulima and/or with your TA to confirm when you signed up before your regular Lab. The lab coordinator and/or a TA will be happy to get you started. It is very important that you show up. If you absolutely cannot make this time period, confirm with your lab partner that they can make it. We will have a sign-in sheet in. You will be given a subset of larvae to work with.

a. **Sketches and photos.** Use Data Sheet 2 provided in class for your drawings. Before starting, you will be given a specific pre-assigned set of tubes containing your fish larvae. Label a set of new **1.5 ml microtubes** with the **same assigned sample numbers.** Each larval sample will be placed in this new tube after you have finished your sketch of it and taken a photo.

**Sketch 2 of these larvae of your choice!** Begin sketching by carefully removing a single larva from its original tube, place it in a petri dish with a few drops of ethanol and examine it under a dissecting microscope.

**Take photos of all larvae assigned to you!** Take a photo of each of your larva to have an additional morphological reference to the DNA sequence. Follow directions by the dissecting
scopes with the digital cameras. **Take a picture of your sample numbers so we know which photo corresponds to which sample.** See instructions by scope for more details.

**Important:** sketch and photograph one larva at a time, so that you do not mix up samples, then place the just photographed larva in its newly labeled 1.5 ml microtube. Use the provided Data Sheet 1 to keep track of your samples throughout the three weeks.

Note the following in your sketch:

1. Make an assessment of the stage of development: **pre-flexion**, or **post-flexion**. Pre-flexion larvae have no well-developed fin rays. Post-flexion larvae have well-developed fin rays. Post-flexion larvae are older and are approaching the recruitment stage. Consult the handout: “The morphology and identification of fish eggs and larvae” by Bruce Mundy for descriptions.

2. Size and position of the eyes (if still present and intact).

3. **Pigment** patterns. External and internal spots will appear black. These spots can in some cases provide important characters to species.

4. Length and folding of gut.

5. **Fin** positions and ray counts.

b. **Begin DNA extraction.** We will be using a commercial product called a **Qiagen DNeasy DNA extraction kit**. These kits use a specially designed spin column to bind DNA and let the remainder of cellular debris filter through the column. Before we start working with the columns, we need to first break down the tissues into cells. We will start this protein digestion before your regular lab period.

Add the following reagents to each 1.5 ml microtube, each of which should contain one fish larva:

1. **180 µl** Qiagen ATL buffer.

2. **20 µl** proteinase K.

3. Mix briefly by **flicking with finger 5-10 times** or vortex.

4. Give all your samples to your TA or lab coordinator, who will place them in a waterbath set at 56°C for ca. 2 hours (or leave at room temperature for >20hrs). The buffer and proteinase K will break down all tissue to the cellular level.

II. **Afternoon session: 1:30-4:30 pm**

a. **Continue DNA extraction**

5. **Cell lysis.** Vortex for 15 seconds. Add **200 µl AL buffer**, mix thoroughly by vortexing, and incubate in **hot-block/water bath** for 10 minutes at 56°C.

6. **DNA precipitation.** Add **200 µl 100% ethanol** to the sample, mix thoroughly by vortexing.

7. **Label a set of DNeasy spin columns with your sample #s.** Label the top of the spin column, not the collection tube. You should have as many spin columns as samples.
8. Set the large P1000 pipette to 650 µl. Pipette mixture from step 6 into a correctly labeled DNeasy spin column placed in a 2 ml collection tube (you should have 600 µl after step 6; pipetting 650 µl assures that you transfer all of your sample).

9. Bind DNA to column. Centrifuge spin columns (with collecting tubes) using the mini-centrifuges in the lab. **Set RPM to 8000. Spin for 1 minute.** Remove column from collection tube, empty collection tube into appropriate waste container, and place the spin column in a new collection tube. Discard the collection tube in your “used tips” waste container.

10. Wash DNA once. Add **500 µl of AW1 buffer.** Centrifuge for **1 minute at 8000 RPM.** Remove column from collection tube, and place the spin column in a new collection tube. Discard flow-through and collection tube as in the previous step.

11. Wash DNA again. Add **500 µl of AW2 buffer.** Centrifuge for **3 minutes at 14,000 RPM** to dry the membrane. Label a new set of 1.5 ml microtubes with sample numbers. Place each spin column in the correctly labeled microtube. Discard flow-through and collection tube.

Be careful removing the collection tube from the centrifuge, and **DO NOT TIP OR SPILL ANY AW2 BUFFER ONTO THE MEMBRANE.** Carefully remove the spin column from the collection tube, again assuring the buffer does not touch the membrane. The membrane is the white material that makes up the “floor” of the spin column. If the membrane gets wet at this stage, spin again (without any new buffer) at 14,000 for 3 minutes in an empty collection tube. Empty collection tube into appropriate waste container, and place used collection tube in recycling bin.

12. DNA elution. This step will release the DNA from the membrane in the column. **Add 200 µl of AE buffer directly to the membrane.** Make sure you pipette the AE buffer directly over the membrane, not on the sides of the column, but make sure not to puncture the membrane with your pipette tip.

**Now let stand at room temperature for 2 minutes.**

Then **spin at 8000 RPM for 1 minute.** In the microtube you should now have all the genomic DNA (nuclear and mitochondrial) from your fish larva.

**b. Set-up polymerase chain reaction (PCR).**

1. **Prepare for PCR.** Each PCR reaction occurs in a single tube in a chain of 8 strip tubes of 25 µl volume. **Each tube equals 1 reaction.** Label a set of strip tubes with your sample numbers. Add two additional reactions for your positive control (i.e., a known DNA extraction from fish) and a negative control (i.e., a blank reaction with no DNA; use AE buffer or distilled water instead). The positive control is a check to see if you have setup your PCR mix correctly. The negative control makes sure that there is no non-fish contaminating DNA in your PCR. **So if you did 8 larval DNA extractions, you would label 10 tubes total.** You can use scissors to cut strip tubes into smaller groups if necessary, or use a strip of 8 tubes and 2 single tubes.

2. Each single PCR reaction will contain the following ingredients:


   **However we will use a commercial PCR mix that already includes 1.-4. and 7.**
3. Calculate **PCR Mastermix.** For each reaction you will be using **25 µl reaction volumes.** This reaction volume contains 24 µl of PCR ingredients and 1 µl of **DNA template.** The DNA template comes from each larval extraction. We mix up all the ingredients except the DNA template beforehand in the PCR Mastermix. If your PCR includes 8 larval samples, you would need to make enough Mastermix for: 8 larval extractions + 2 controls + 1 = 11 reactions. We add one extra reaction volume to cover pipetting error. Some of this error is equipment, not user!

Use Table 1 to calculate the volumes for each ingredient that will go into the Mastermix. **Check with your TA if these are the correct concentrations and amounts in case there were some last minute adjustments.**

**Table 1. Calculating the PCR Mastermix**

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration</th>
<th>Volume (in µl)</th>
<th># reactions + 1</th>
<th>Total volume for Mastermix = (Volume)*(# reactions + 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>d H₂O</td>
<td>-</td>
<td>8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MyTaq Red Mix</td>
<td>2x</td>
<td>12.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fwd COI Fish2 Primer</td>
<td>10µM</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rev COI Fish2 Primer</td>
<td>10µM</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA template</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>Do not add to Mastermix</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>-</td>
<td><strong>25</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. **Assemble the Mastermix.** Add each quantity in the table above to a 1.5 ml microtube in the order according to Table 1, beginning with the largest volumes first. Mix by flicking tube or vortex briefly.

5. **Pipette Mastermix into individual strip tubes.** Add **24 µl of Mastermix** to each strip tube. Each tube should be labeled with a sample number or control. **You should have some residual Mastermix left over.** If not, you did not calculate the Mastermix or pipet correctly. See your TA.

6. Now carefully add **2 µl of DNA template to each individual tube.** Make sure you are adding the correct DNA template to the correct tube. In other words, your labels match. **Watch the pipette tip to make sure you are picking up this small quantity. It should contain no bubbles, and you should see the DNA template liquid enter the tube.**

Place your strip tubes on ice, until your TA is ready to start the class PCR in the PCR machine. Fill out the datasheet to ensure we know where your tubes

7. **PCR protocol.** For this mitochondrial gene we will be using the following PCR protocol. **A PCR machine,** or thermocycler, is capable of changing temperatures rapidly, and holding temperatures very precisely. The PCR protocol has the following steps (each step indicates a temperature and time):

1. 95°C for 5 minutes
2. 95°C for 30 seconds
3. 47°C for 30 seconds
4. 72°C for 60 seconds
5. Repeat steps 2-4 for 35 times
6. 72°C for 10 minutes
Each step in the main part of the cycle (steps 2-4) accomplishes a specific task in replicating the DNA sequence of interest.

**a. Denaturation:** 95°C for 30 seconds. In this step the complimentary DNA strands in the template DNA separate from one another.

**b. Annealing:** 47°C for 30 seconds. During this step the short primers (an oligonucleotide strand ~20 bp long; we are using primers designed to amplify ca. 600bp of Cytochrome oxidase I in bony fish we call “COI Fish2 primer”) find their priming sites and bind to them.

**c. Extension:** 72°C for 60 seconds. The Taq polymerase binds the individual nucleotides to the 3’ end of each primer, and extends the sequence. Individual nucleotides (Adenine, Thymine, Cytosine, and Guanine) are in the **dNTP mix** (deoxynucleotide triphosphates).

**d. Repeat:** These three steps are repeated 35 times to produce an exponentially increasing set of copies of the target gene.

The PCR run will take about 2 hours and is fully automated. The lab coordinator will store your completed reactions in the fridge until the next lab.

### C.2. Protocol for Week II

**Reading before Lab:** Hebert et al. 2015, Will & Rubinson 2004, & Lab manual

Recall that last week you extracted the genomic DNA from a set of larval fish samples from Hawaii. You then used PCR to make many copies of a mitochondrial gene so that we can directly sequence it. In this week’s lab, we will verify your DNA extraction and PCR amplifications from last week by using agarose gel electrophoresis. After you verify that you have amplified a potential mitochondrial gene, you will use an enzyme system called “Exosapit” to clean up the PCR reaction and prepare it for direct sequencing.

**Make sure to keep track of which of your samples worked and what you did with them on Data Sheet 1.**

**Background information for gel electrophoresis**

In gel electrophoresis we run our DNA strands through a super fine agarose matrix. Think of a sponge with teeny tiny holes, all of the same size. We move the DNA through this matrix by subjecting it to an electrical field. The DNA moves (migrates) from the region of negative charge to the region of positive charge at a rate that depends on its length. Longer fragments move slower whereas shorter fragments move faster.

**Figure 1.** Top view of a DNA agarose gel, stained with ethidium bromide. In this figure the current moves from the top of the gel (region of - charge) to the bottom of the gel (region of + charge). Note this gel has two rows of “lanes.” Each sample is loaded into the small rectangle at the top of the row. The sample moves downward depending on the size of the DNA strands. Small fragments move further. The left lane in each row has a DNA size standard, or “ladder”. On this gel, the size standard has 100 bp increments. If you count the bars upward, you should see that the PCR products in each lane are all of similar size, about 600 bp.
At one point today you will be given a tour of one of UH’s high throughput sequencing facilities: The Advanced Studies of Genomics, Proteomics, and Bioinformatics “ASGPB” (www.hawaii.edu/microbiology/asgpb/) located in Snyder Hall. This Center will be sequencing your samples this week. They have the capability to sequence thousands of genes every day (they will sequence our samples in a few hours!), and we would like to give you an idea of the kinds of technology and the capabilities of this facility.

I. Casting your Agarose gel. Your agarose gel will be cast for you on the gel tray in the gel-chamber with the casting gates. We will use a 1.5% agarose gel. 1.2 gr of agarose will be added to 80ml of 1xTAE Buffer in an Erlenmeyer flask, swirled to mix, and heated to melt the agarose in the microwave. Subsequently the fluorescent dye GreenGlo will be added before pouring the still warm and liquid agarose into the gel tray. Wait till the gel hardens to a jello consistency.

II. Loading your samples and running your agarose gel. Each team will load their PCR reactions onto their prepared agarose gel. You need one more lane than you have numbers of PCR reactions for the DNA ladder (e.g. if you have 10 PCR reactions you need 11 lanes). The DNA ladder should go at the beginning or at the end, best in lane 1. Fill out Data Sheet 3 with the lanes and numbers of your samples. It is crucial that you keep track of your samples and lanes. Lane 1 is always the lane to the far left of the gel. Lanes 2 to X proceed from left to right.

Important: We will be using a fluorescent DNA dye that is supposed to be less a health hazard compared to the often used Ethidium Bromide. However, for safety reasons still wear gloves while handling gels at all times.

You will use Data Sheet 3 to score your gel for successful amplification, and to record whether the PCR product is potentially ready to sequence.

a. Load DNA ladder: If you have never loaded an agarose gel before, see your TA to practice on a separate gel before loading your samples. Load 5 µl of the 100 bp DNA ladder into lane 1 of your gel.

b. Load Samples: Load 5 µl of your sample into each sequential lane on the gel. (You do not need to add any loading dye to your samples, as the dye is already in your PCR Buffer.)

c. Run Gel: Place the cover on the gel rig and turn the power pack on. Make sure the cover is on the correct position so that the DNA runs toward the red power plug (check for bubbles). Run at 100 Volts for ca. 30 min.

III. Visualizing your Gel:

We will visualize the gel on a mini-photo system. Take a picture with your phone for your records.

IV. Analyze PCR results, prepare PCR products for sequencing.

a. Examine your gel image for single, dark bands in sample lanes. Your positive control lane should have a single band, and the negative control lane should be blank. Mark all sample lanes on Table 2 that have a single dark band as “OK for sequencing”. Provide an explanation of your gel results in the notes section of Data Sheet 3. Have your TA verify your analyses of PCR products.
b. **PCR Clean-up.** Label a set of new strip tubes with all samples that are OK for sequencing from Step a above. **Add 9 µl of your PCR reactions** to the accordingly labeled strip tube. **Add 1.1 µl of enzyme** provided by your TA. The enzyme is a mixture of ExoI (Exonuclease I from *E. coli*) and FastAP (thermosensitive Alkaline Phosphatase). Place strip tube in PCR machine. When all samples are in the machine, your TA will turn on the following incubation profile:

- **Incubation:** 37°C for 30 min.
- **Denaturation:** 85°C for 15 min.

c. **Sequencing.** We will take 5 µl of each of your samples to the ASGPB facility for sequencing at the end of the week. ASGPB will e-mail us the sequences for analyses in the final installment of this lab.

### D. Analyses: Week III

**Due: Sequence analysis exercise, Drafts of Introduction and Methods for Formal Report**

In this last lab of a three-part series you will get the opportunity to use a popular bioinformatics tool “**BLAST**” to search a library of sequence data called “**GenBank**” as well as the species identification tool on the Barcode of Life Data systems website “**BOLD**”.

We will first check the sequences returned from the ASGPB automated sequencing facility for quality and if needed edit them using a simple editing program called “**Sequence Scanner v1.0**.” We will then submit edited files of good quality sequences via the web and analyze the results.

**I. Editing Sequence Files:** The sequencing results from each numbered sample were returned as chromatogram files (also called trace files). These files are saved on Laulima in the folder “DNA”. You can open and view these files with the program “**Sequence Scanner**” that has been installed on all lab laptops. This program will allow you to view the chromatograms, edit the DNA sequences, and save the edited DNA sequences for further analyses. Find the sequences assigned to you by your TA in the folder and open them in “Sequence Scanner”. Sequence files are named as follows: e.g., file “TH1_F2-C33-A02.ab1” is sample # C33.

a. **Opening files with Sequence Scanner:** First open the program by double-clicking on the shortcut on your desktop. On the left you’ll see a blue bar called “Trace Manager”. Click on “Import Traces” under “File Tasks”. Select all sequence files assigned to you in the Shared folder and click “Add Selected Traces”, then click “OK”. A list of your sequences will appear in the “Details” view. Select one sequence at a time to edit. To edit your first file double click on it and the chromatogram/trace file will appear in a window below your list. The trace file will display a series of colored peaks, each corresponding to a nucleotide. If the PCR amplified a single band with little or no contaminating sequence you will get a “clean read”: a series of clearly distinguishable single peaks with scored nucleotides given as a sequence across the top of the screen (bars above nucleotides are tall). PCR amplifications of multiple genes or non-specific DNA template lead to “messy reads”. Messy reads have multiple peaks, which are difficult to score. Scoring nucleotides, or “base calls”, are ambiguous because one peak does not dominate. These sequences will contain many “Ns”, or nucleotide positions that cannot be determined with certainty (the bars above the nucleotides are short). We call these “ambiguous base calls” because there is too much noise in the data to get a clear picture of the nucleotide at that position. Keep in mind that the beginning (< 20 base-pairs) and end of a clean sequence are always “messy,” and these Ns should not be
included when trying to find matches in online sequence databases. For now, note whether the middle of your sequence is clean or messy. You will edit those sequences that have a long, clean middle section and try to find matches in two online databases (GenBank and BOLD).

b. **Edit good DNA sequences in Sequence Scanner:**

1. **Cut off the messy beginning and end.** Find the light blue marker at the beginning and end of your sequence file. Drag it to the last nucleotide you do not want to include in your sequence (usually an N). You will see that the colored trace and base calls now are light grey. **Keep track of the length of your sequences on Data sheet 4.**

2. Nucleotides can be edited at any position by highlighting the base and typing a new nucleotide or N.

c. **Export your edited and cut off sequence:**

1. Click on “Export Traces” under “File Tasks” in your “Trace Manager” bar. Under “File Format” choose “*.fsta” and “Post-trim Sequence Only”. Save the file under the same name on your desktop.

2. Open this file with “Notepad” (Double click on the file on your desktop and Choose “Notepad” as application to open it with).

d. **Conduct a BLAST search with your DNA sequence in GenBank:**

1. Open the web page [https://blast.ncbi.nlm.nih.gov/Blast.cgi](https://blast.ncbi.nlm.nih.gov/Blast.cgi). GenBank is an open access sequence database administered by the National Center for Biotechnology Information (NCBI) where all sequences used in scientific publications need to be deposited before publication. BLAST (Basic Local Alignment Search Tool) is a search algorithm that compares your submitted sequence to sequences in the GenBank database to find best matches.

2. Click on “nucleotide blast”. This will redirect you to the actual query page of BLAST.

3. Copy the content of your sequence with “Notepad” opened sequence file into the query box where it says “Enter accession number, gi, or FASTA sequence”. For “Database” select “others”, and at the bottom of the page where it says “Optimize for” under “Program selection” select “Somewhat similar sequences (blastn)”. Leave the rest as is. Click BLAST.

4. You may have to wait a few seconds before the results of the search come up. Once they do, start scrolling down to look at them.

5. Your results are presented as a) a Graphic Summary (colored bars), b) Descriptions (in a list) of your matches followed by some scores and numbers and a “GenBank” accession #, and c) a section with alignments of your sequence to those in the list above.

6. **Graphical Summary.** This plot has two kinds of information. First, the length of your sequence, “the query,” is given in red across the top with numbered base pairs. Next you will see a series of bars of different colors below the query sequence. These bars represent GenBank sequences with the highest alignment scores indicated by color, also ranked from best to worst match as you go down the page. **Red bars have scores > 200= a really good match.** Other colors give lower matches. The length of these bars represent how much of the GenBank sequence aligns with your query.
sequence. This has nothing to do with the quality or identity of the match, but only with the length of the sequences that match.

7. Descriptions. Each line consists of a description of a specific sequence on GenBank, starting with the genus and species of the organism the sequence originates from. You can click on it to get more information. Then there are a serious of numbers that have to do with the match of your sequence with those found on GenBank according to the search algorithm BLAST. The most important one for you is “Ident” which gives you the nucleotide identity between the two sequences in %. Anything above 95% is a good match (above 98/99% is better). The exact percentage to get species-level identification may however depend on your species. At the end is the GenBank accession # which is a unique number attributed to each sequence submitted to GenBank.

8. Alignments. The last part of the search result shows DNA alignments between the query sequence and the sequence found on GenBank. Important information to look for at the top is “Identities” between this sequence and your sequence (in numbers of nucleotides and %). You can inspect the two sequences to see where they differ. How many sites do the two sequences differ at? Does this mean they are different species? Here you can also click on the GenBank Accession Number. This will redirect you to the page of this sequence with some more information. If you click on the species you will be redirected to a page with some more information on this species and with some external links to more info. Explore some of the species you are curious about!

e. Conduct a BOLD search with your DNA sequence:

1. Open the web page www.barcodinglife.org/. Click on “Identification”.

2. Ensure you are in the “Animal Identification [COI]” Tab. Copy the content of your sequence file opened with “Notepad” into the query box where it says “Enter sequences in FASTA format:”. For “Search Databases” select “Species Level Barcode Records”. Click “submit”. If that does not give you any results, try to use “All Barcode records on BOLD” instead.

3. The main part of the Result page is a list of top 20 matches to your sequence with their taxonomic classification and a specimen “Similarity (%)”. This % similarity corresponds to the % sequence identity (“Ident”) in GenBank. Anything above 95% is a good match (above 98/99% is better). The exact percentage to get species-level identification may however depend on your species.

4. If your top 20 matches do not have any species ID (e.g. just say “Chordata”), change the number of results to 50 or 99. You can also redo the search selecting under “Search Database” “Species Level Barcode Records”.

5. To visualize your results in a more graphical way you can click on “Tree-based Identification” at the top of your result list. Click on View tree and you will get a phylogenetic tree based on the relationship of the nucleotide sequences. Sequences with a high similarity (i.e. closely related to each other) will be grouped together. Find your sequence (it’s the one in red labeled as “unknown specimen” and see if it clusters with other sequences. If it does, it is a good match to these sequences, which hopefully represent sequences from a single species.

6. To get more info on the specie of your top match you can also click on “Species page” on your Result page.
II. Record your GenBank and BOLD results in Data Sheet 4. Use the Notes section for anything unusual you notice. Enter all your data into the team Excel spreadsheet (in the “Shared Biology Folder”) and email it to your TA before leaving. Your TA will post the compiled data on the class website.

III. Interpreting your results.
   a. Good match: Search scores >500 for sequences that can be aligned over 200 bp or greater and sequence identities above ca. 90/95% indicate that there is at least one sequence in GenBank that is a close match to the sample sequence. Look up the matched sequence and research the distribution of this fish species. Compare this identification with Bruce Mundy's morphological IDs for the samples. Do the genus and species (or at least family) identified from morphological characters match those from the molecular identification?
   
   b. Poor match: Low Search scores (<500), short alignments and low sequence identities indicate a poor match in the respective database. This indicates there are no submitted sequences for the larval species and the mitochondrial gene we sequenced. To find out if any genetic information exists for the potential species of your sample, search the family and genus (and species, if available) identified morphologically by Bruce Mundy under the webpage: http://www.ncbi.nlm.nih.gov/. Search under “nucleotide” and copy your genus or species name (or just the family) into the “for” search window.
      i. How many entries are returned?
      ii. What geographic regions do these entries represent?
      iii. What might this result say about the database?

   Use this information to complete Data Sheet 5.

   c. Once all the class data is compiled you can summarize some of the results and compare results among morphological identification and the two databases using DNA Barcoding. How many larvae got identified reliably/with confidence at the family, genus, species level with which method?

   Use Data Sheet 6 or make your own data sheet to summarize the class data.

Week IV

Due: Drafts of Results and Discussion for Formal Report

1. Peer review of Introduction, Methods, Results
2. Discussion of how to make sense of large data sets, including our compiled class data
3. Discussion of cohesiveness of Formal Report
4. Peer Review of Discussion
E. Formal Lab Report I

The data from this lab forms the basis for your first formal lab report. Use the structure outlined in “Guidelines for Writing a Scientific Paper” in Appendix B as a guide for your writing. The references discussed in class and chosen before as well as some others of your choice (see below for more examples) should be used to provide background in your introduction, and to help you interpret your results in the discussion. In your results section you need to include your sketches of the fish larvae (include your original sketches), as well as the compiled class results from GenBank and BOLD. Try to make some sense of this big data set by looking at some patterns and summarizing the data in some numerical or graphical way. Also, discuss the congruence between morphological and molecular identifications (you can make a table summarizing your information from data sheet 5, use data sheet 6 or make your own). As usual, if you refer to any summarizing data or table, refer to them in the text and include them in your formal report. References should be formatted as in the journal “Marine Biology”.

Ensure also to check the grading rubric when writing your report, as that is what your TA will use to grade your formal report.

F. References

This list is just a small subset of adequate publications. You are not limited to this list, but can do your own search.


Identification of fish larvae:


**Webpages:**
- The Barcode of Life: phe.rockefeller.edu/barcode
- Consortium for the Barcode of Life: [www.barcoding.si.edu/DNABarCoding.htm](http://www.barcoding.si.edu/DNABarCoding.htm)
- Barcode of Life Datasystems (BOLD): [www.barcodinglife.com](http://www.barcodinglife.com)
- Marine Barcode of Life: [http://www.marinebarcoding.org](http://www.marinebarcoding.org)
- Fish Barcode of Life Initiative: [www.fishbol.org](http://www.fishbol.org)
- Fishbase: [www.fishbase.org](http://www.fishbase.org)
- General good info on reef fish (incl. pics of larval stages): [www.coralreeffish.com](http://www.coralreeffish.com)
- California Academy of Sciences Catalog of Fishes: [http://research.calacademy.org/ichthyology/catalog](http://research.calacademy.org/ichthyology/catalog)
- American Fisheries Society, Section: Early Life History: fisheries.org/sections