Biotechnology Explorer™

Fish DNA Barcoding Kit

Bioinformatics Guide

explorer.bio-rad.com

Catalog #166-5100EDU, Fish DNA Barcoding Kit
Catalog #166-5115EDU, DNA Barcoding Sequencing Module

Note: This document is for planning purposes only and may vary from the final product specifications.

For technical service, call your local Bio-Rad office or, in the U.S., call 1-800-424-6723.

BIO-RAD

Table of Contents

Instructor's Advance Preparation

Before the Barcode of Life Data Systems-Student Data Portal (BOLD-SDP) can be used to analyze student data, you must register for an account. The steps required to do so are outlined below. There is also a video tutorial on the BOLD-SDP website that demonstrates these steps. Please be aware that you need to register with the BOLD-SDP website 2–3 days before you want to have the students run through the bioinformatics lessons.

The procedure outlined below contains some steps that must be performed by the instructor before the students can enter the BOLD-SDP database. Other steps, such as creating specimen folders and uploading sequencing trace files, can be done by either the students or the instructor. Where there are wait times for calculations the BOLD-SDP software needs to perform, the amount of time is listed in the instructions to help with classroom time planning. Ultimately, the instructor can choose which steps are performed by whom and cater the course to teaching goals.

Setting Expectations

The BOLD-SDP portal was designed to serve as a workbench for those interested in generating data of sufficiently high quality to be considered for publication in the BOLD reference barcode database. The data contained within the BOLD database come from samples that were vouchered and have sequencing data of extremely high quality. In order to be able to contribute sequencing data to this database for use by other researchers, a very stringent process must be obeyed. Because of this, certain sequences students generate may not meet the requirements for assessment by the BOLD-SDP software and may receive trace file quality assessments of "low" or "fail". This does not preclude students from learning bioinformatics or analyzing the data they generate, however, and students should not be discouraged! If students receive sequencing data quality assessments that are not high enough to allow them to complete the use of all the software packages contained within the BOLD-SDP portal, they can still perform a BLAST (basic local alignment search tool) search using the GenBank database (instructions included in the first section of the student instruction manual) and see what level of homology they have to fish samples in that database.

GenBank is a commonly used and well known database and does not have the same stringent requirements as BOLD for submission of sample sequences to its database. Hence sample identification based on sequence matches with GenBank sequences may not be as trustworthy. For example, sequences for known fish samples have mistakenly corresponded to sequences from marine bacteria that had contaminated the original samples, and samples have been incorrectly identified in the first place. These errors are not just made by students, but also sometimes by researchers! This is an important teaching point on the handling of samples and looking at database data. Students can still work through a homology search and see if their sequences might match any in a sequencing database and at what level. An 86% match? A 99% match? How comfortable does your student feel with guaranteeing an identification of a fish if it has only 86% homology to a fish in the database? The more experience students get in looking at their data against data in different databases, the more they will improve their critical thinking skills. It is not merely a matter of "my sequence matches this database record by *x*% and therefore must be *y* fish."

Timeline for Required Advance Preparation

Steps	Work Time	Wait Time
Create an Instructor Account on BOLD-SDP • Receive an email from BOLD with instructor username and password • Receive an email from BOLD with five registration keys	5 minutes	• 1 hr • 1–2 days wait post registration
Register a new course with BOLD-SDP • Receive an email from BOLD with COURSE Username and password	No time	• 1 hr
Create specimen folders into which each student group will upload fish sample data • Create specimen folders • Upload data into folders • Optional: Assess quality of the data and determine workflow students will follow	Up to 3 hrs	• 1 hr • 1 hr, plus up to 24 hrs wait for quality data to be calculated • 1 hr

Instructor Required Advance Preparation Steps

Step 1: Registering an account

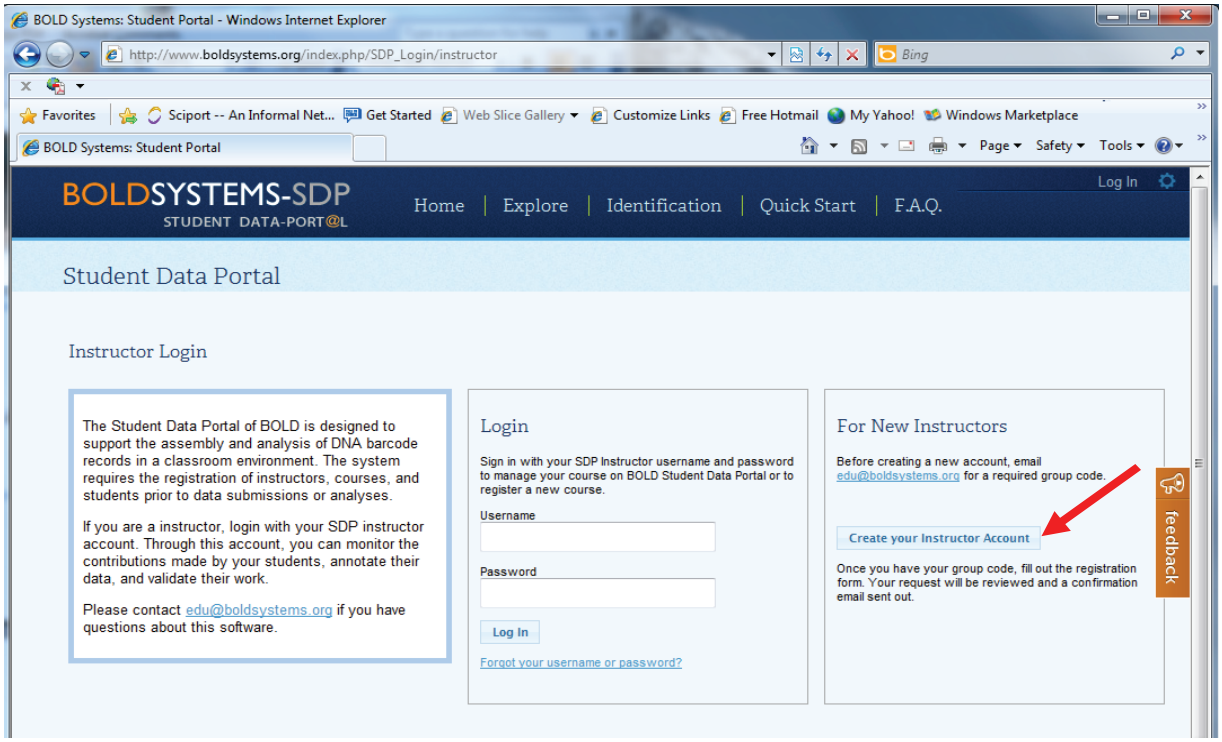
a. Access the homepage of BOLD-SDP at **http://www.boldsystems.org/index.php/SDP_Home**.

If after reading these instructions you require additional guidance to create a user account, please note that there is a video tutorial that can be accessed by clicking the **Quick Start** tab or the **Quick Start Guide** on the BOLD-SDP homepage.

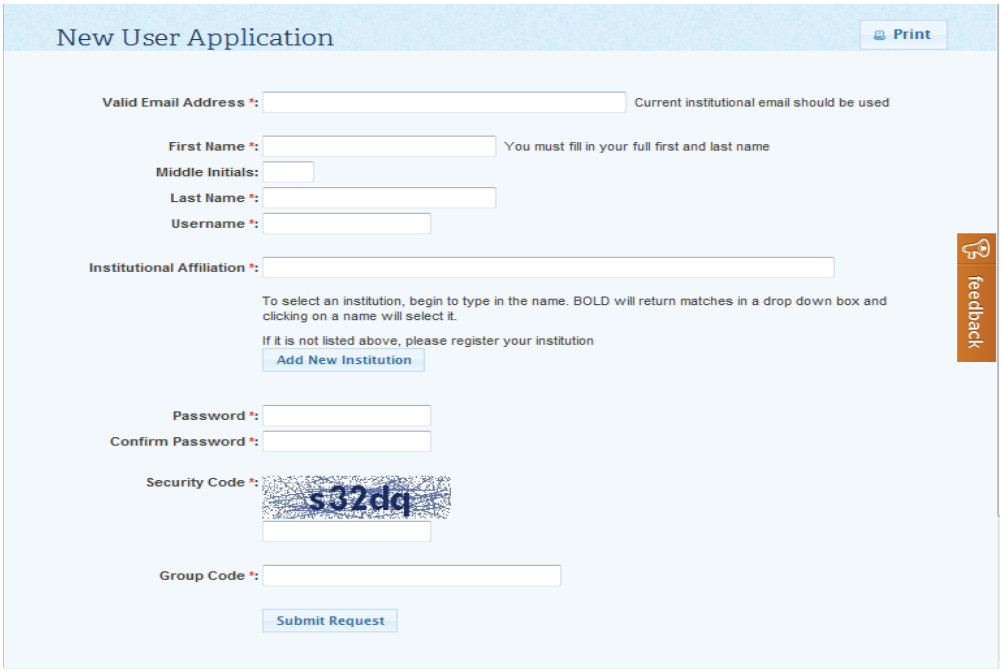
b. Click the **Instructors** icon at the bottom of the homepage.

c. If this is the first time you are using the BOLD-SDP website, you will need to create a new instructor account. Note that you can create an account without emailing BOLD Systems to request a group code because the required group code has already been obtained and is listed below in step 2.viii.

Click the **Create your Instructor Account** button.



This will open the New User Application page.



Fill in all the information and click **Submit Request**.

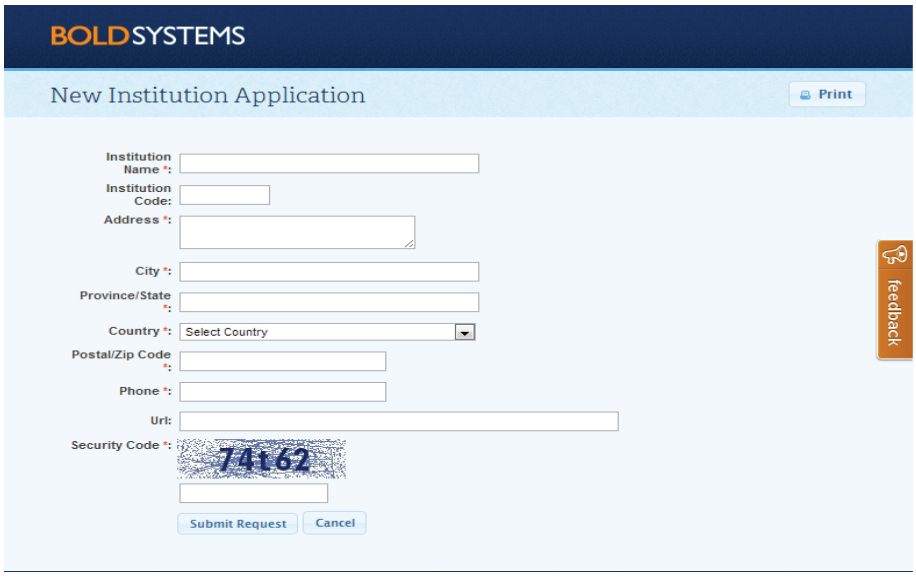
i. Enter your valid email address.

ii. Enter your first name.

iii. Enter your last name.

iv. Enter your desired username.

v. Enter your institutional affiliation. If your institution is not in the database, click **Add New Institution** and add the required information, then click **Submit Request**.



vi. Enter a password and then enter it again in the Confirm Password box.

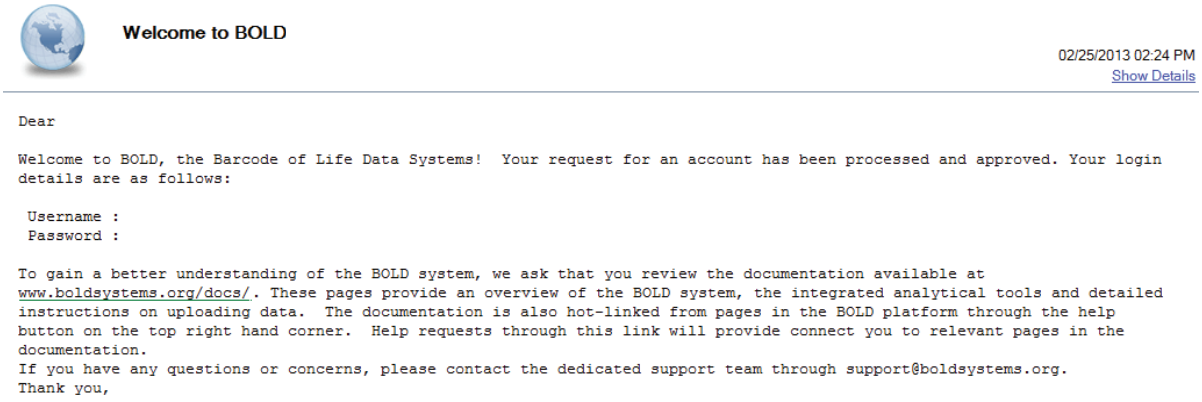
vii. Enter the security code as it is shown.

viii. Enter the group code **BOLD-EDU-SDP** and then click **Submit Request**.

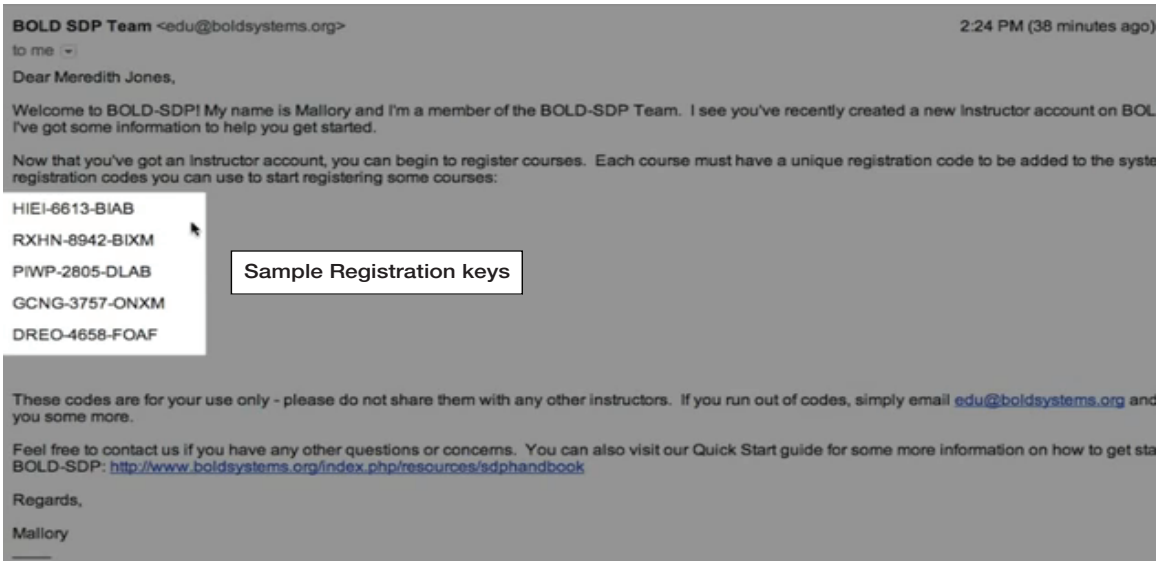
1-2 days

d. Within 1–2 days of submitting your request you will receive two separate emails from BOLD-SDP.

- Email 1 will say "Welcome to BOLD" in the subject line. This email will contain information on your **instructor username and password**



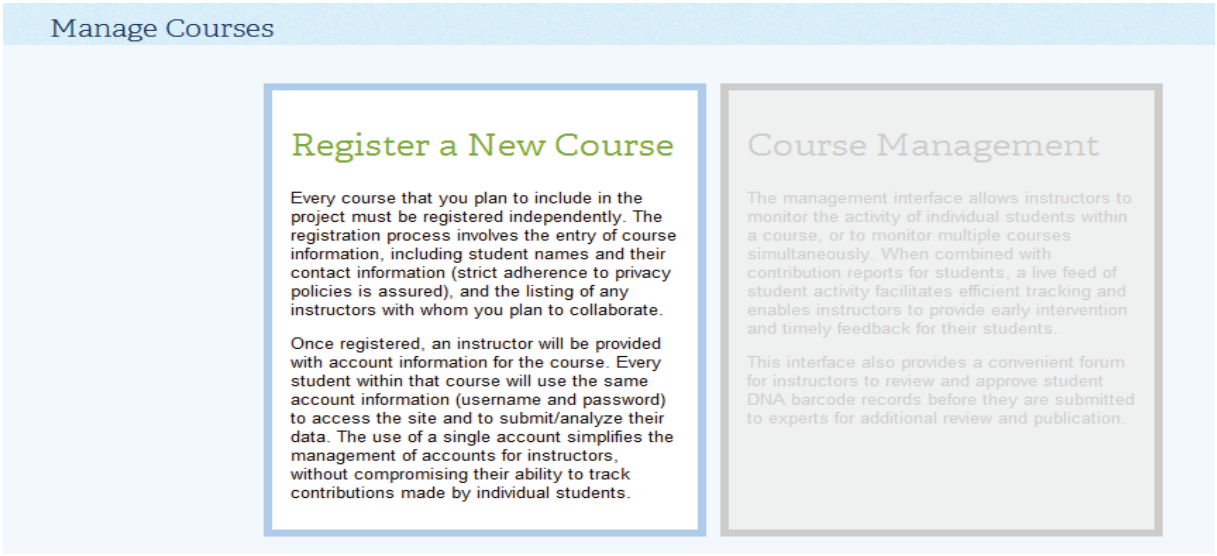
- Email 2 will arrive later with “Welcome to BOLD-SDP!” in the subject line and will include five registration keys. Each registration key can be used for a separate class. If you need more keys, please email **edu@boldsystems.org**



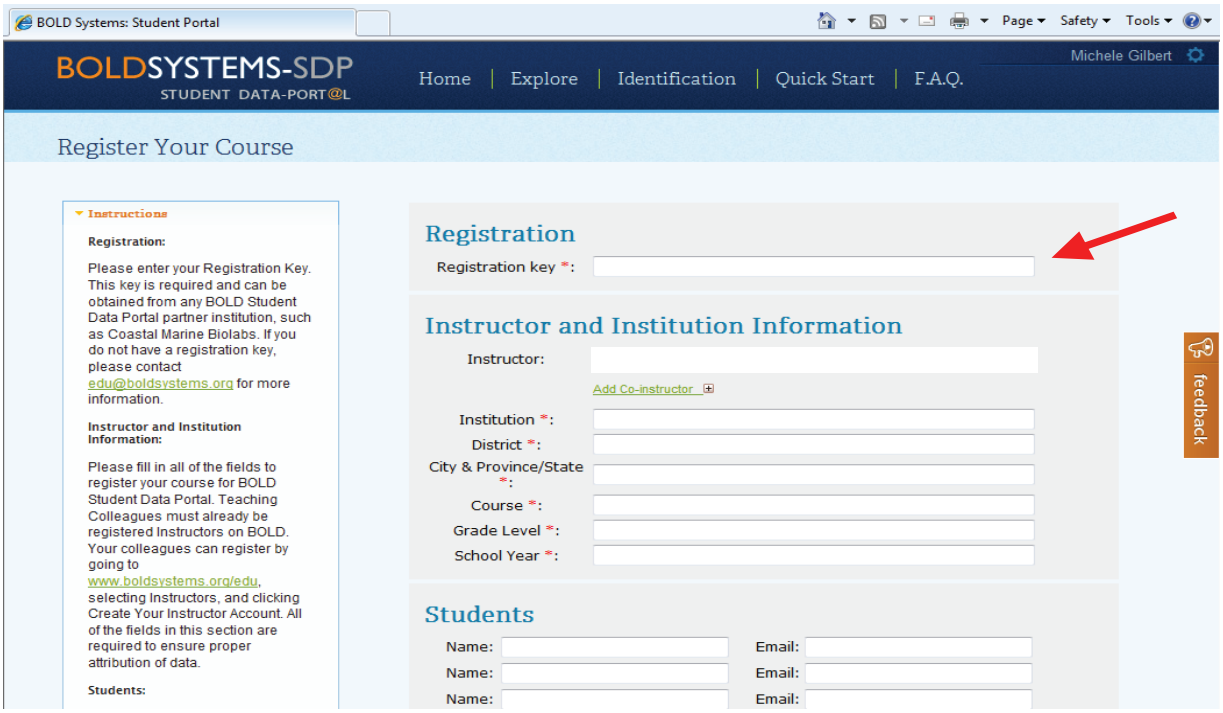
e. At this point you can register your class on BOLD-SDP. From the BOLD-SDP homepage, click the **Instructors** button.

f. Log in with the **instructor username and password** that was emailed to you, then click **Log In**.

g. You will now see a page where you can register your class information. Click **Register a New Course**.



Type in or copy and paste one of the five registration keys that was emailed to you.

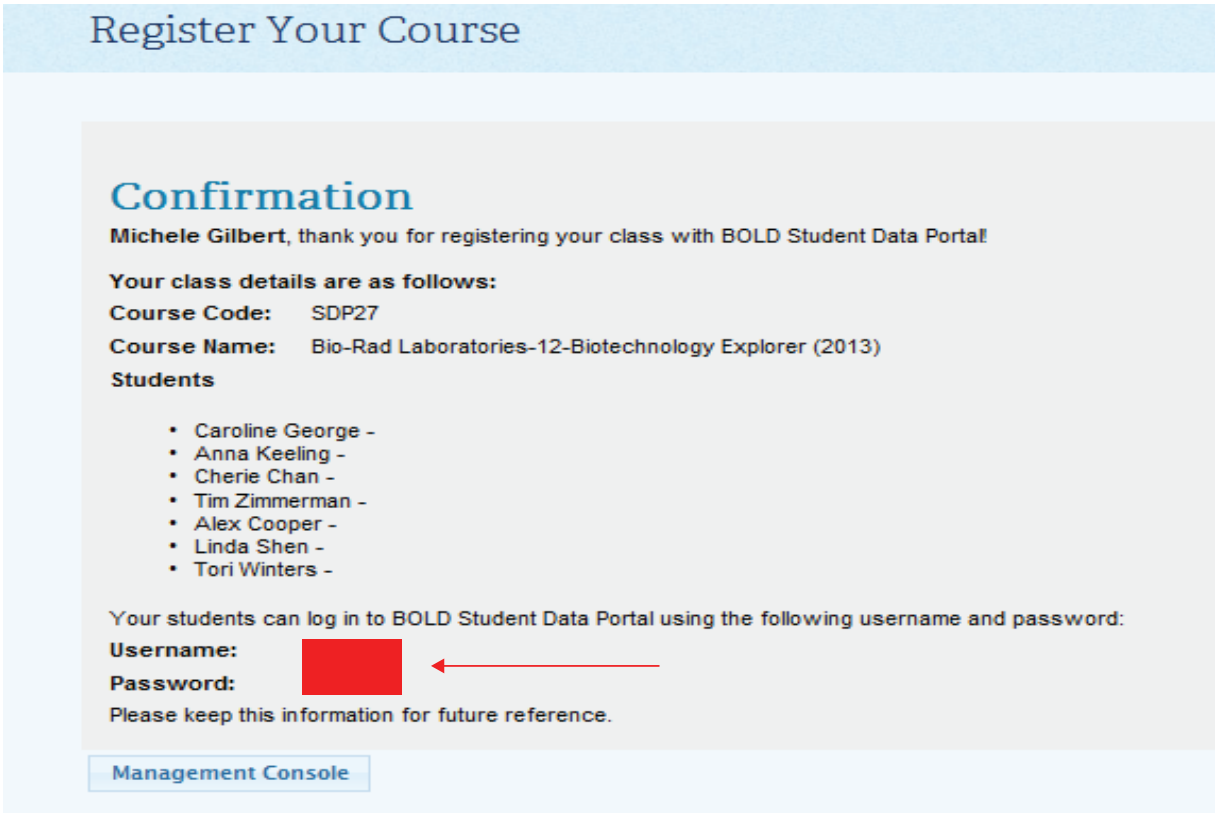


h. Your name should already be in the Instructor box. If you have a co-instructor, click **Add Co-Instructor**. (Co-instructors must also have their own instructor usernames and passwords in BOLD-SDP in order to access your account's data.) Please note that only the instructor can add more co-instructors. However, all other instructor privileges will be shared by anyone designated a co-instructor.

i. Now enter your institution, district (or county), city, province/state, course name, grade level (9–16 entered as a numeral), and school year in the appropriate boxes.

j. Enter the names of your students. Email addresses are not required, but can be entered if desired.

k. Click the **Submit** button. This will open a page where the **course username and password** will be shown. You should also receive an email with the same information. **Make sure to record your course username and password, as these will be required to work on your data within BOLD-SDP.**



Course Username: _____

Course Password: _____

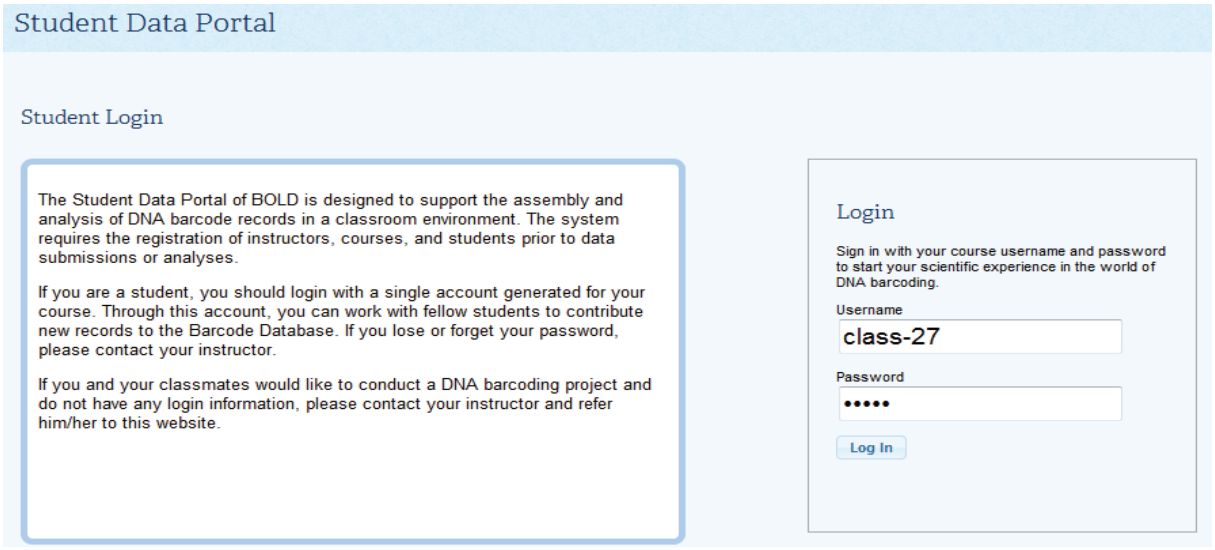
Step 2: Creating specimen folders into which student groups will upload data

At this point, you have registered yourself, the instructor, with BOLD-SDP. You have received an instructor username and password and five registration keys, each of which can be used to generate a separate course username and password for each class, and you have registered your students and received your first course username and password.

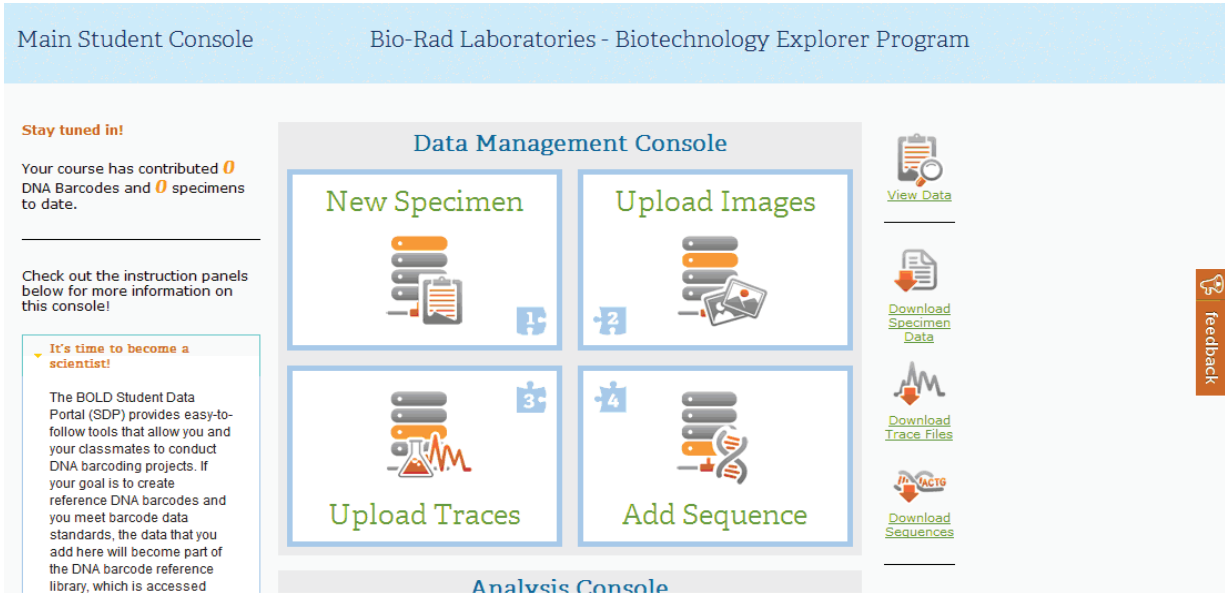
The following steps outline how to create specimen folders so that forward and reverse sequence data can be uploaded to BOLD-SDP. This step can be performed either by the instructor or by the students, but must be performed before sequencing trace files can be uploaded and analyzed in BOLD-SDP.

a. Access the homepage of BOLD-SDP at: **http://www.boldsystems.org/index.php/SDP_Home**. Click the **Students** button to access the Student Login page.

b. Enter the course username and password into the appropriate spaces and click **Log In** to enter the Main Student Console page. Please note that the password is case sensitive!



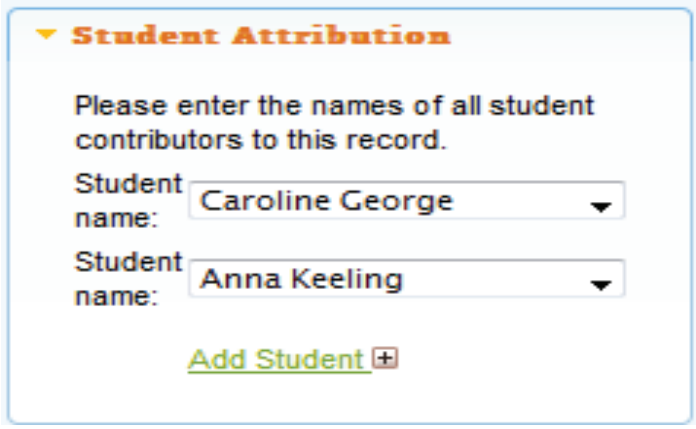
c. You should now see the Data Management Console.



d. Click the **New Specimen** icon.

The New Specimen page will open. For each student group, you will need to enter multiple pieces of information.

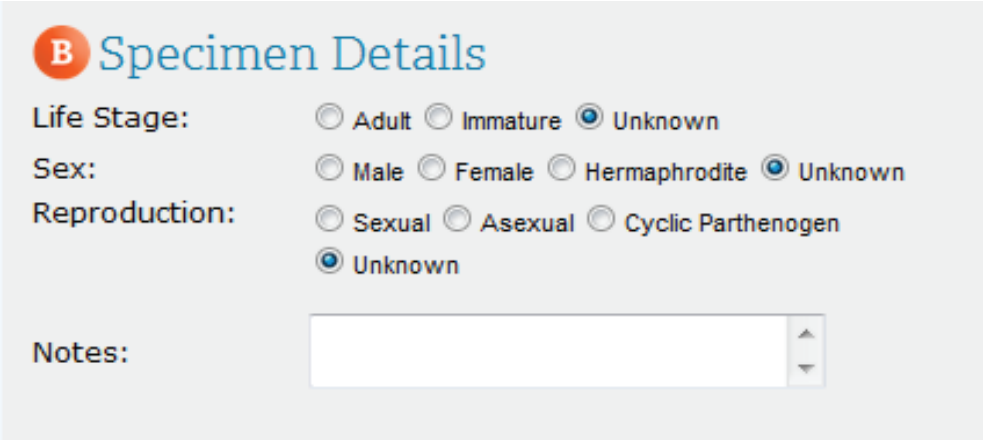
First, using the Student Attribution box, choose the student(s) who worked on this fish sample. If more than one student worked on this sample, click the **Add Student** button and select his/her name from the dropdown menu. Repeat the **Add Student** process until all students who worked on this sample have been added. This is important in order to later assign credit to the students who did the work for each sample in various steps of the process.



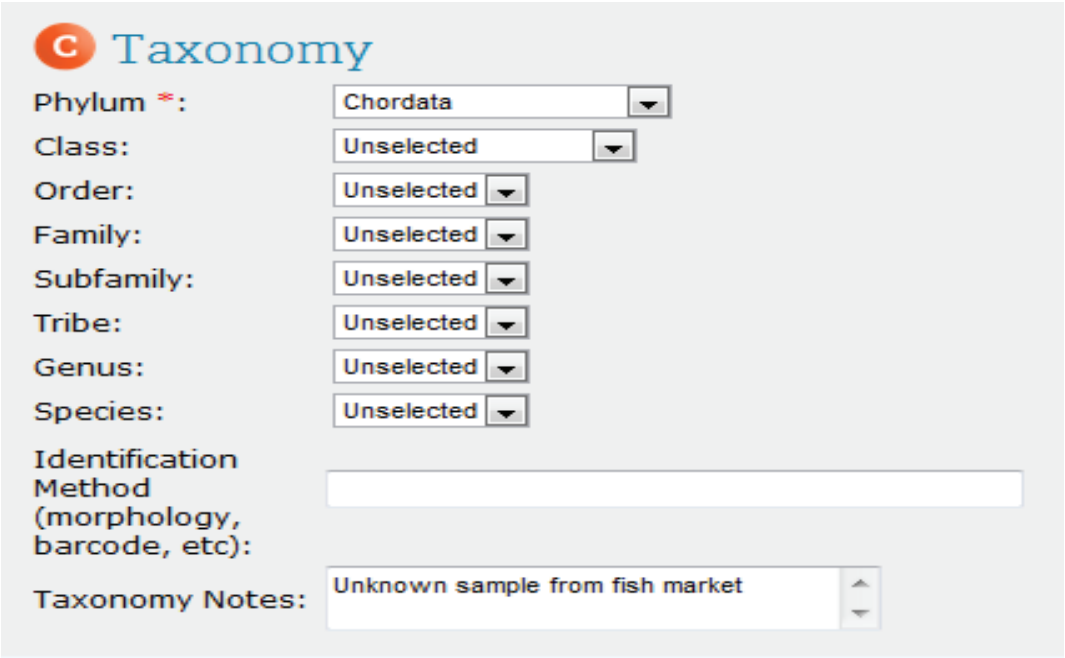
In step A, **Specimen Identifier**, enter a name to uniquely identify the sample you are testing. The specimen identifier needs to be a name that is unique not only for the samples tested by your class, but also unique within the entire BOLD database. **The best way to ensure a unique specimen identifier name is to include the following information in the name: 1) year, 2) institution, 3) group name, and 4) some information about the fish sample**. After entering a name, press the **Tab** key. If the name is unique, a green arrow will appear on the right-hand side of the sample ID name. An example of a unique specimen identifier name is shown below.

Note: It is important to take extra care when entering your desired sample ID, as this information cannot be modified later. If a mistake is made, the entire specimen record can be deleted, but this option is accessible only when logged in as the instructor.

In step B, **Specimen Details,** if you know the life stage, sex, and reproduction mode of the fish sample you are using, you can select those parameters. Otherwise, select the Unknown circle.



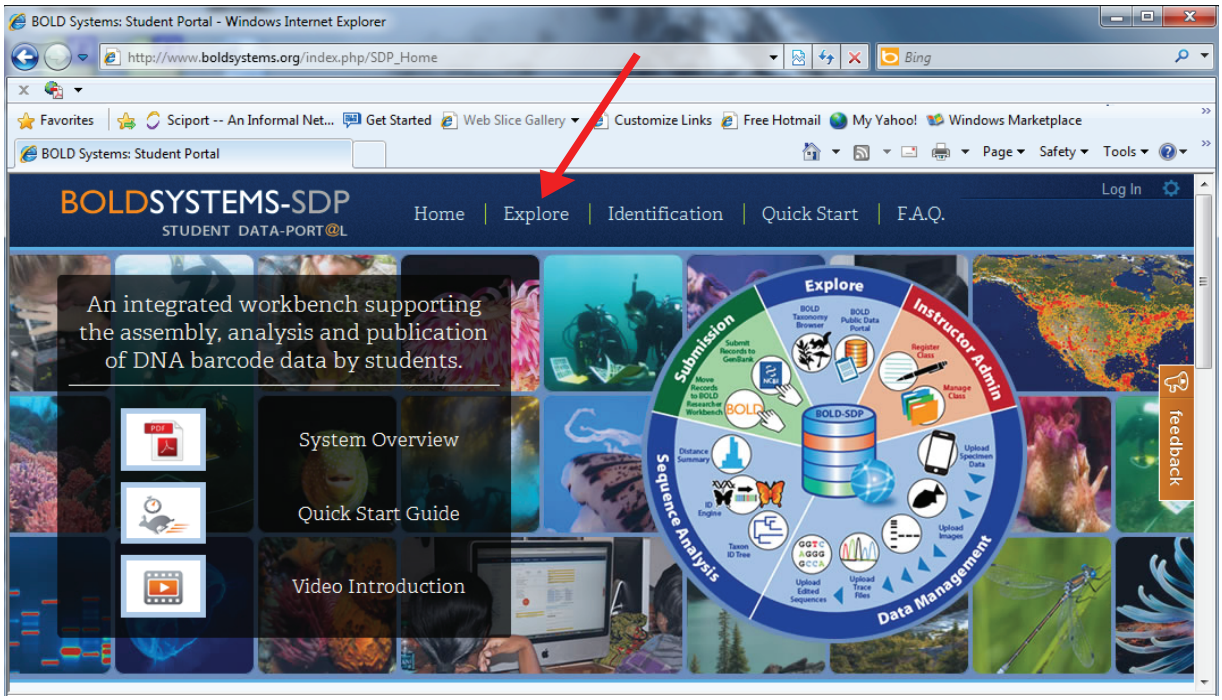
In step C, **Taxonomy**, fish are members of the phylum Chordata. If you are using fish samples, use the dropdown menu to select **Chordata**. If you are using other marine life such as shrimp, squid, mussels, etc., find out the appropriate phylum and choose that from the dropdown menu. Also, you can type comments in the Taxonomy Notes to specify where you collected your sample. Only the phylum data field is required — other taxonomic data are optional.



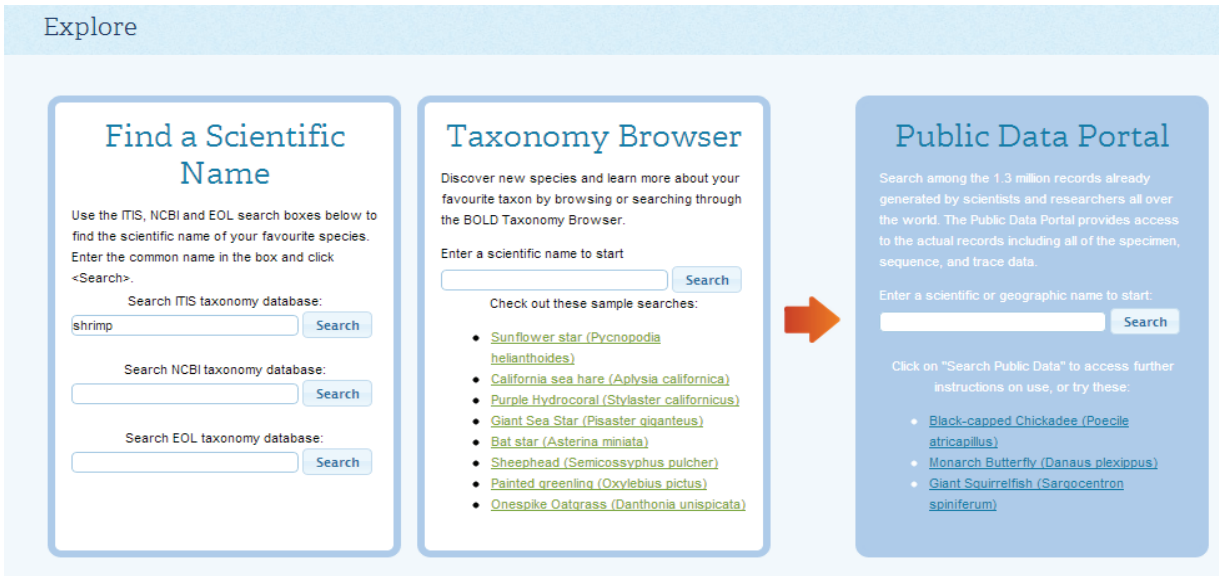
If you do not know the phylum of your sample, you can use a search browser in BOLD-SDP to find it.

1. Open a new browser page and go to the BOLD-SDP homepage at: **http://www.boldsystems.org/index.php/SDP_Home**. You do not need to be logged in to your class account in order to perform the following searches.
2. Click **Explore**.

This will open the Explore page.



Type the common name of your sample (for example, shrimp) into the Search ITS taxonomy database box, then click Search.



A new page will open with multiple species and subspecies. Find your sample on the list and click it to get its entire taxonomical tree, including the phylum to which it belongs — in the case of shrimp, Arthropoda.

ITIS Search Results

Home | About ITIS | Data Access and Tools | Get ITIS Data | Submit and Update Data

Results of: Search in every Kingdom for Common Name containing 'shrimp'

Kingdom Animalia

abalone visored shrimp – Species: Betaeus harfordi (Kingsley, 1878) – valid
Adonis shrimp – Species: Parapenaeopsis venusta De Man, 1907 – valid
Aesop shrimp – Species: Pandalus montagui Leach, 1814 – valid
African mud shrimp – Species: Solenocera africana Stebbing, 1917 – valid
african spider shrimp – Species: Nematocarcinus africanus Crosnier and Forest, 1973 – valid
akiami paste shrimp – Species: Acetes japonicus Kishinouye, 1905 – valid
Alabama cave shrimp – Species: Palaemonias alabamae Smalley, 1961 – valid
Alachua fairy shrimp – Species: Branchinella alachua Dexter, 1953 – valid
alamang shrimp – Species: Acetes sibogae Hansen, 1919 – valid
Alaska bay shrimp – Species: Neocrangon alaskensis (Lockington, 1877) – invalid
Alaska coastal shrimp – Species: Heptacarpus moseri (M. J. Rathbun, 1902) – valid
Alaskan pink shrimp – Species: Pandalus eous Makarov, 1935 – valid
Aleutian coastal shrimp – Species: Heptacarpus maxillipes (M. J. Rathbun, 1902) – valid
alkali fairy shrimp – Species: Branchinecta mackini Dexter, 1956 – valid
american grass shrimp – Species: Palaemon pandaliformis (Stimpson, 1871) – valid
American grass shrimp – Species: Periclimenes americanus (Kingsley, 1878) – valid
anchialine pool shrimp – Species: Palaemonella burnsi Holthuis, 1973 – valid
anchialine snapping shrimp – Species: Metabetaeus lohena A. H. Banner and D. M. Banner, 1960 – valid

ITIS Report

Home | About ITIS | Data Access and Tools | Get ITIS Data | Submit and Update Data

Go to Print Version

Pandalus eous Makarov, 1935
Taxonomic Serial No.: 621110

Download data (Download Help) ***Pandalus eous*** TSN 621110

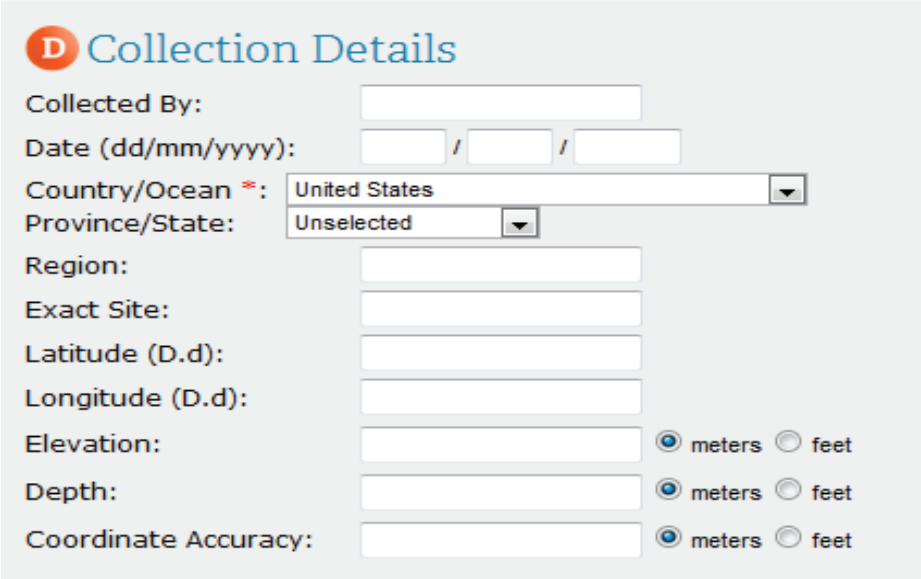
Taxonomy and Nomenclature

Kingdom:	Animalia
Taxonomic Rank:	Species
Synonym(s):	Pandalus borealis eous Makarov, 1935
Common Name(s):	Alaskan pink shrimp [English]
Taxonomic Status:	
Current Standing:	valid
Data Quality Indicators:	
Record Credibility Rating:	verified - standards met

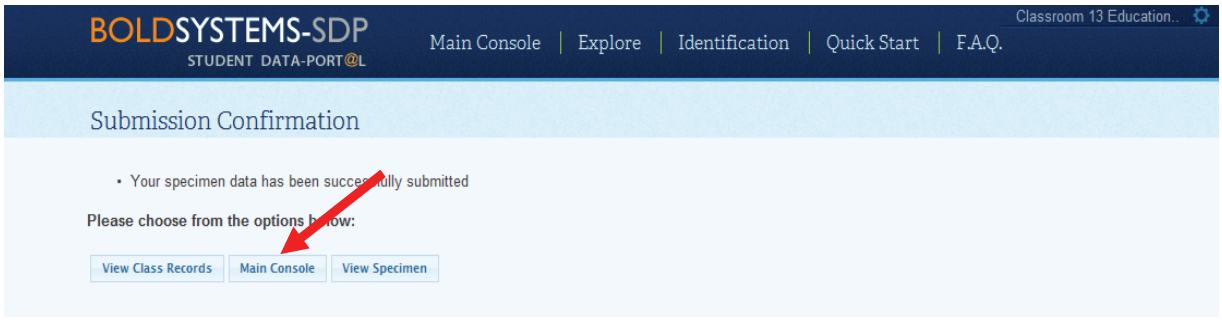
Taxonomic Hierarchy

Kingdom	Animalia – Animal, animaux, animals
Phylum	Arthropoda – Artrópode, arthropodes, arthropods
Subphylum	Crustacea Brünnich, 1772 – crustacés, crustáceo, crustaceans
Class	Malacostraca Latreille, 1802
Subclass	Eumalacostraca Grobben, 1892
Superorder	Eucarida Calman, 1904 – camarão, caranguejo, ermitão, lagosta, siri
Order	Decapoda Latreille, 1802 – crabs, crayfishes, lobsters, prawns, shrimp, crabes, crevettes, écrevisses, homards
Suborder	Pleocyemata Burkenroad, 1963
Infraorder	Caridea Dana, 1852
Superfamily	Pandaloidea Haworth, 1825
Family	Pandalidae Haworth, 1825
Genus	Pandalus Leach, 1814
Species	Pandalus eous Makarov, 1935 – Alaskan pink shrimp

In step D, **Collection Details**, the minimum detail that needs to be entered is the Country/Ocean of sample collection. Use the dropdown menu to enter the information. Unless you caught the fish yourself and are certain of the location where it was caught, you should list the location of the store/restaurant from which you purchased the fish.

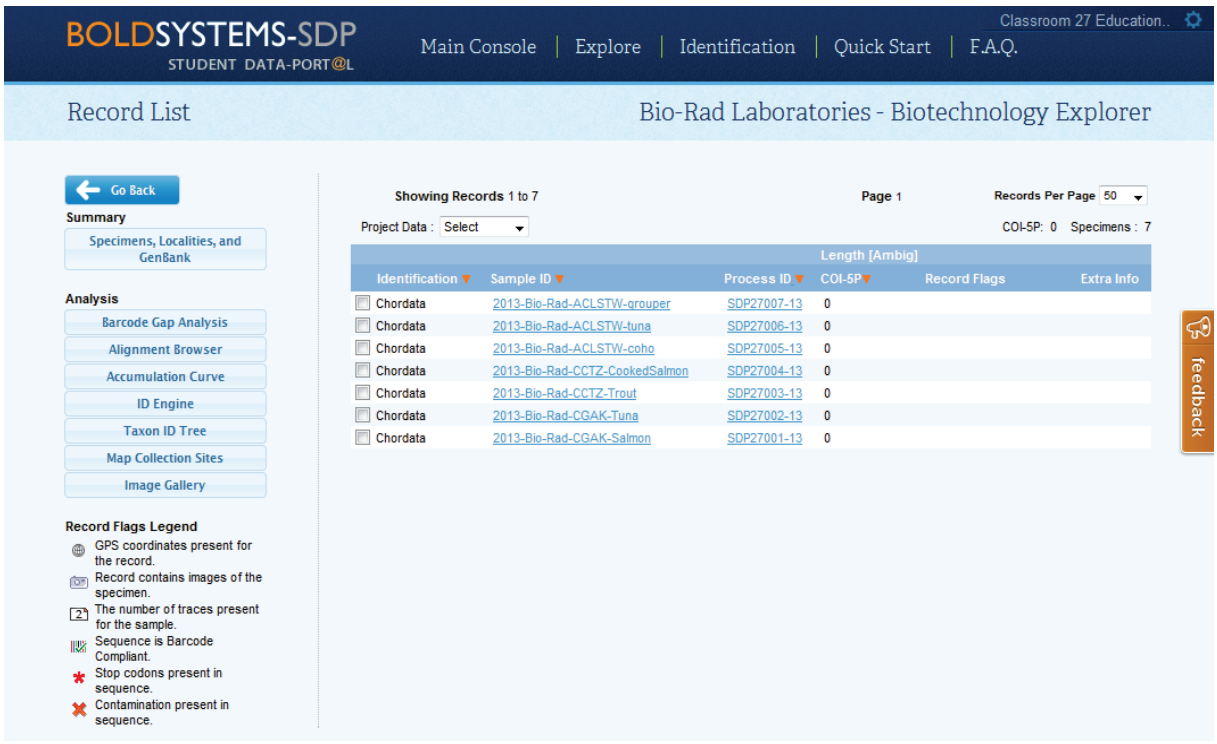


Once all data have been entered, click the Submit box. You will see a Submission Confirmation page.



Repeat the above steps to generate sample IDs for all student samples tested. To get to the main console page in order to create more sample IDs, click the **Main Console** button.

Once all sample IDs have been entered, you can see the complete list by clicking the **View Class Records** button on the Submission Confirmation page or by clicking the **View Data** button on the Main Console Page.



Once you have confirmed that all student samples have sample IDs, click the **Main Console** button to go back to the Main Console page.

Step 3: Uploading forward and reverse trace files into your account on BOLD-SDP

1 day

These steps may be performed by the instructor or by the students. Sample IDs must have been created for all student samples tested in order to upload traces. Also, after each trace is uploaded to BOLD-SDP, the software will analyze each and assign a quality ranking of high, medium, low, or failed. **It can take up to 24 hours for these quality assignments to be made.** Further analysis of data cannot be performed until the quality assignments have been made by the software. Therefore, if sequence trace files are uploaded by the students, no other analyses can be performed during that class period. **If a continuous workflow by students is desired, then all the sequence trace files should be uploaded to the appropriate Sample ID folder at least one day before the class period.**

a. If you are not already logged in, log in to the Main Student Console page of BOLD-SDP by going to **http://www.boldsystems.org/index.php/SDP_Home** and clicking the **Students** button.

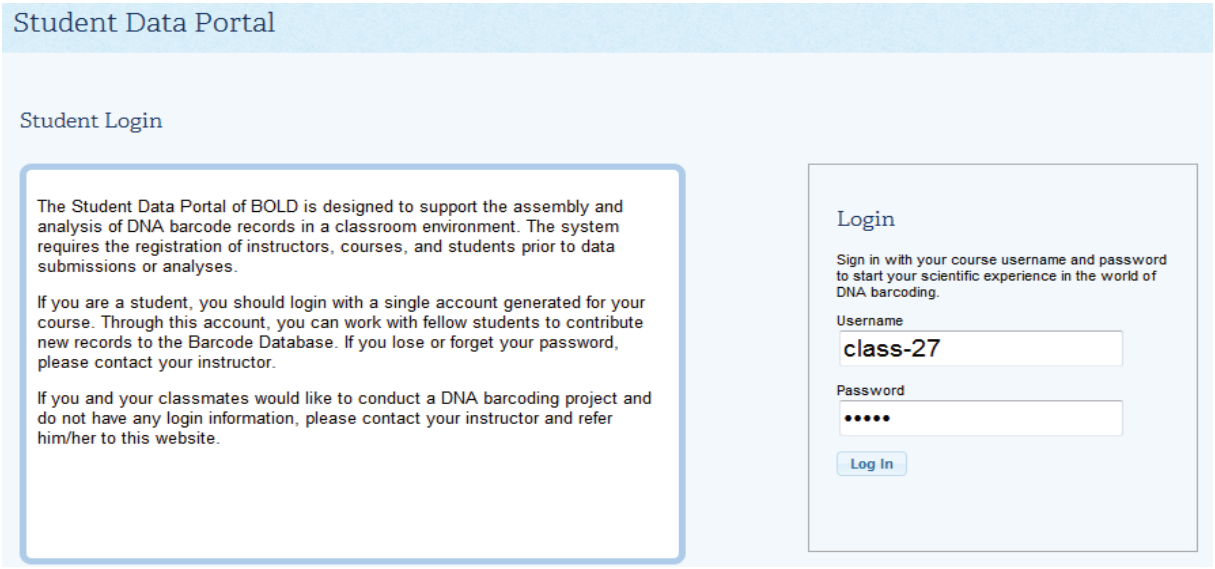
The BOLD Student Data Portal (BOLD-SDP) is a classroom-focused interface to the BOLD database. This platform provides instructors and students the tools necessary to make contributions to DNA barcode library used for identifying species. It also gives students the opportunity to integrate and analyze specimen and sequence data while providing instructors with tools to monitor student progress and evaluate their work. Students can explore the large database of DNA barcode records submitted by scientists around the world, but they will also be able to add their own data to the ever-growing DNA barcoding library. Contact us for support in using this system in a classroom setting (edu@boldsystems.org).

A portal for students to assemble DNA barcode records in a simplified but powerful data collection and analysis environment.

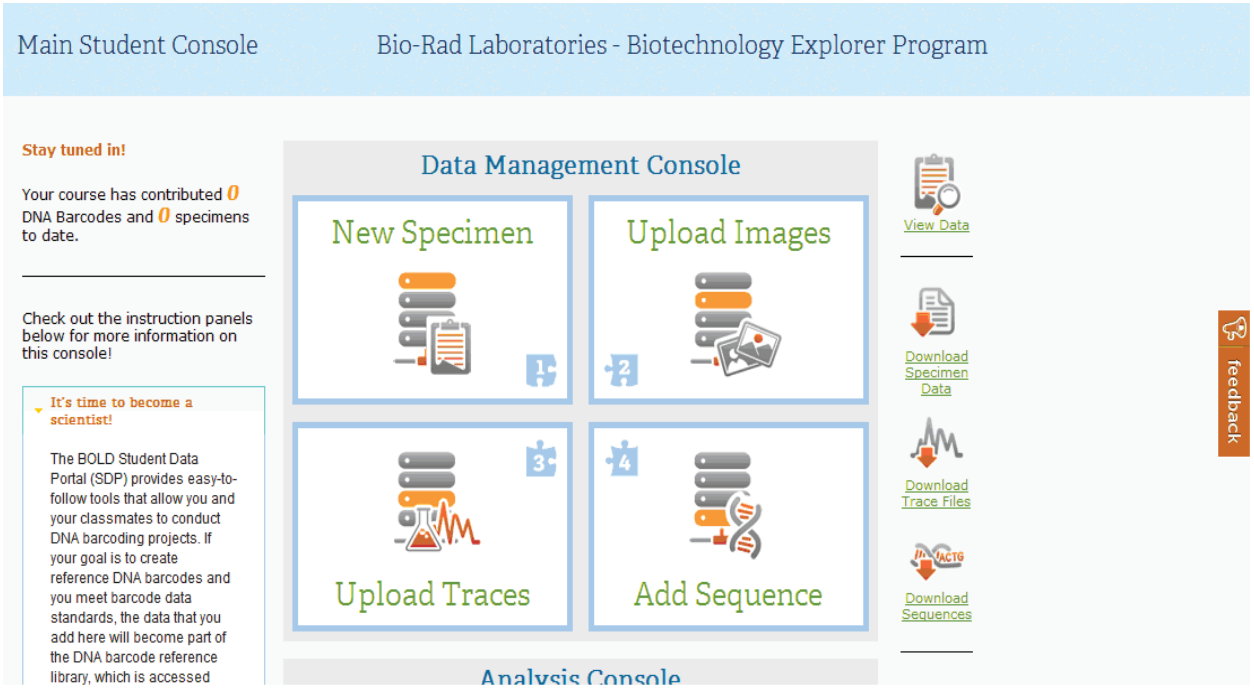
A management interface for instructors to register students and classrooms; review student work and utilize data validation workflows.

Search for and browse data records, images, and geographic distributions for species being barcoded.

b. Enter the course username and password into the appropriate spaces and click the **Log In** button to enter the Main Student Console page. Please note that the password is case sensitive!



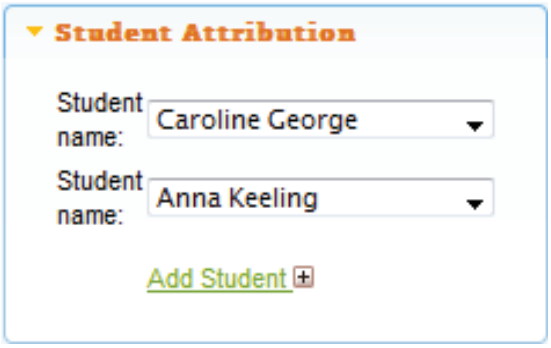
c. You should now see the Data Management Console.



If you have taken photographs of your complete fish, its packaging, or of the sample pieces before you performed the DNA extraction, those photographic images can be uploaded by using the **Upload Images** button. Otherwise, begin by clicking the **Upload Traces** button. Traces refer to the files you received from your sequencing facility. These files contain not only information on the base calls at each location of your PCR product, but also information on the quality of each base call.

d. Click the **Upload Traces** button.

e. Using the Student Attribution box, choose the student(s) who worked on this fish sample. If more than one student worked on this sample, click the **Add Student** button and select his/her name from the dropdown menu. Repeat the **Add Student** process until all students who worked on this sample have been added.



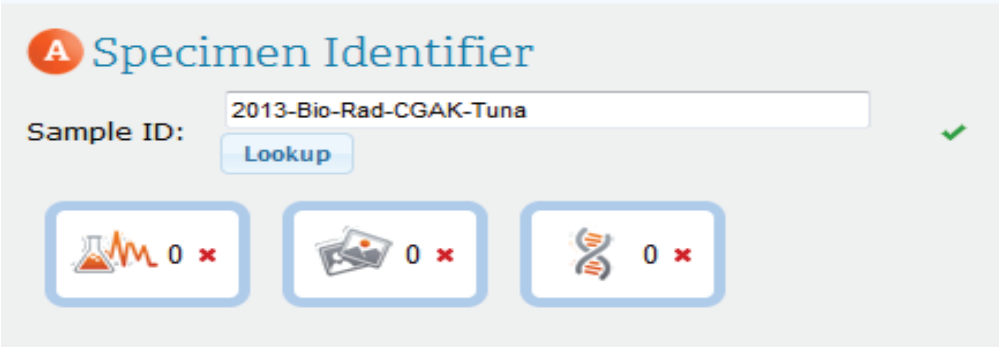
f. In Section A, the Specimen Identifier section of the Upload Traces page, enter the sample ID for your first fish sample. If you do not remember exactly the name of the sample ID for your fish, you can click the **Lookup** button, which will open in a new tab the Record List for your course, and find the sample ID you need. Below is an example of a sample ID list for a class's data.

			Length [Ambig]		
Identification	Sample ID	Process ID	COI-5P	Record Flags	Extra Info
Chordata	2013-Bio-Rad-ACLSTW-grouper	SDP27007-13	0		
Chordata	2013-Bio-Rad-ACLSTW-tuna	SDP27006-13	0		
Chordata	2013-Bio-Rad-ACLSTW-coho	SDP27005-13	0		
Chordata	2013-Bio-Rad-CCTZ-CookedSalmon	SDP27004-13	0		
Chordata	2013-Bio-Rad-CCTZ-Trout	SDP27003-13	0		
Chordata	2013-Bio-Rad-CGAK-Tuna	SDP27002-13	0		
Chordata	2013-Bio-Rad-CGAK-Salmon	SDP27001-13	0		

Note: If you click the sample ID you want to use from the Record List page, it will open another information page but will not fill in the information in the Specimen Identifier Sample ID box.

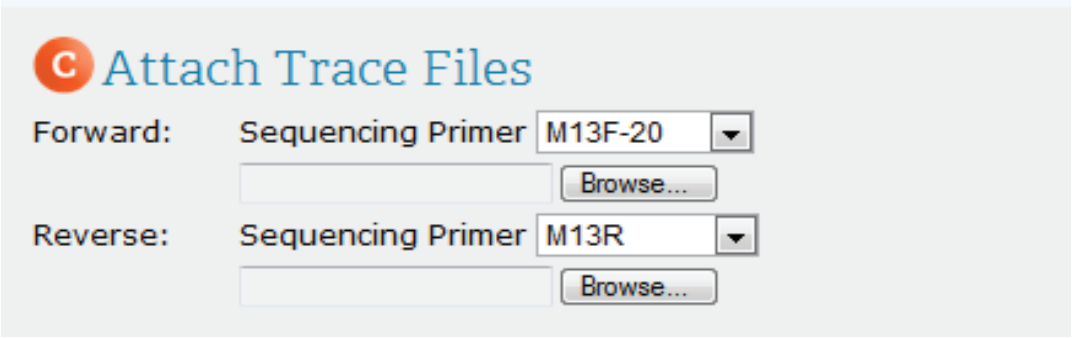
Once you know the name of your sample ID, close or minimize the Record List window.

Type or copy and paste the appropriate sample ID in the Specimen Identifier Sample ID box and press the **Tab** key. If your sample ID was successfully found, you will see three icons appear in the Specimen Identifier box and a green check will appear next to the Sample ID entry. The first of the icons shows how many sequencing traces have been uploaded into BOLD-SDP for this sample. The second icon shows how many photographic images have been uploaded for this sample into BOLD-SDP. The third icon shows how many contiguous sequences have been generated and saved for this sample in BOLD-SDP. For the case shown below, nothing has been uploaded or generated at this point.



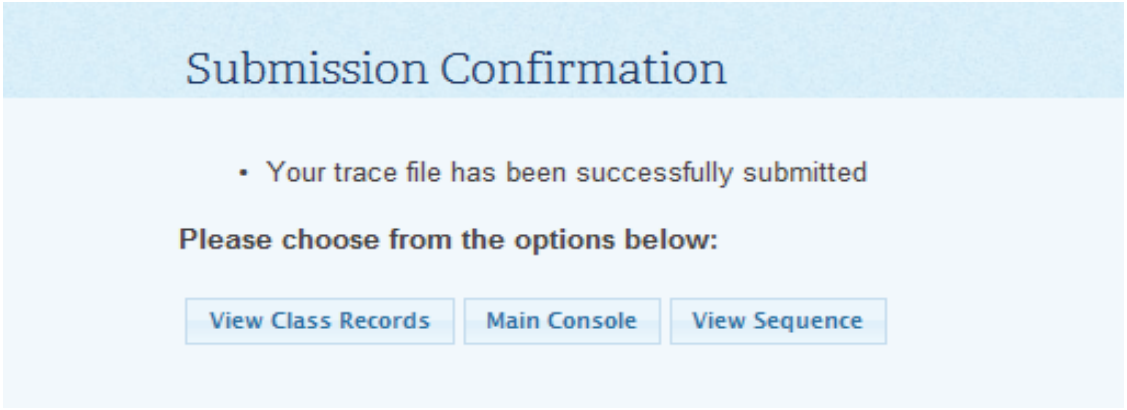
g. In Section B, the PCR Primers section of the Upload Traces page, use the dropdown menus to select the **forward and reverse PCR primers** you used to generate your COI (cytochrome c oxidase subunit I) PCR product. Here, the forward PCR primers used were **C-FishF1t1** and the reverse PCR primers used were **C_FishR1t1**.

h. In Section C, the Attach Trace Files section of the Upload Traces page, use the dropdown menus to select the **forward and reverse sequencing primers** the sequencing facility used to generate your trace files (the sequencing data files). Here the forward sequencing primer used was **M13F-20** and the reverse sequencing primer used was **M13R**.



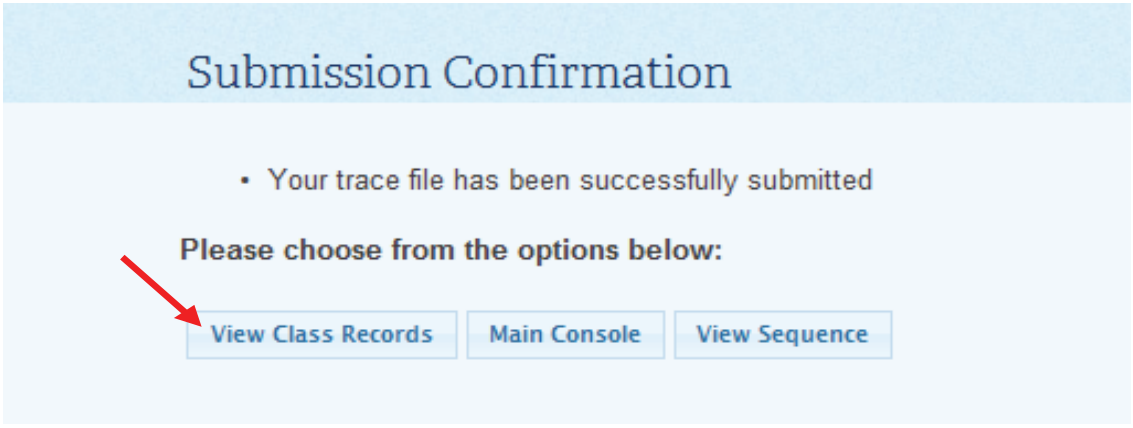
Now, click the **Browse** button under the Forward section and choose your sequencing trace file that corresponds to the forward reaction sent to you from the sequencing facility. This file should end in .ab1. **It is critical that the sequencing trace file name does not contain any characters such as : ; or &. Remove these from the file names before trying to upload them to BOLD-SDP. It is also critical that the name be unique, as the BOLD-SDP portal does not allow duplicate file names. Try to follow the same general naming scheme as the sample ID (year-institution-initials-fish-forward)**. Do the same for the reverse sequencing trace file (year-institution-initials-fish-reverse). Now click the **Submit** button to add your data to the BOLD-SDP database.

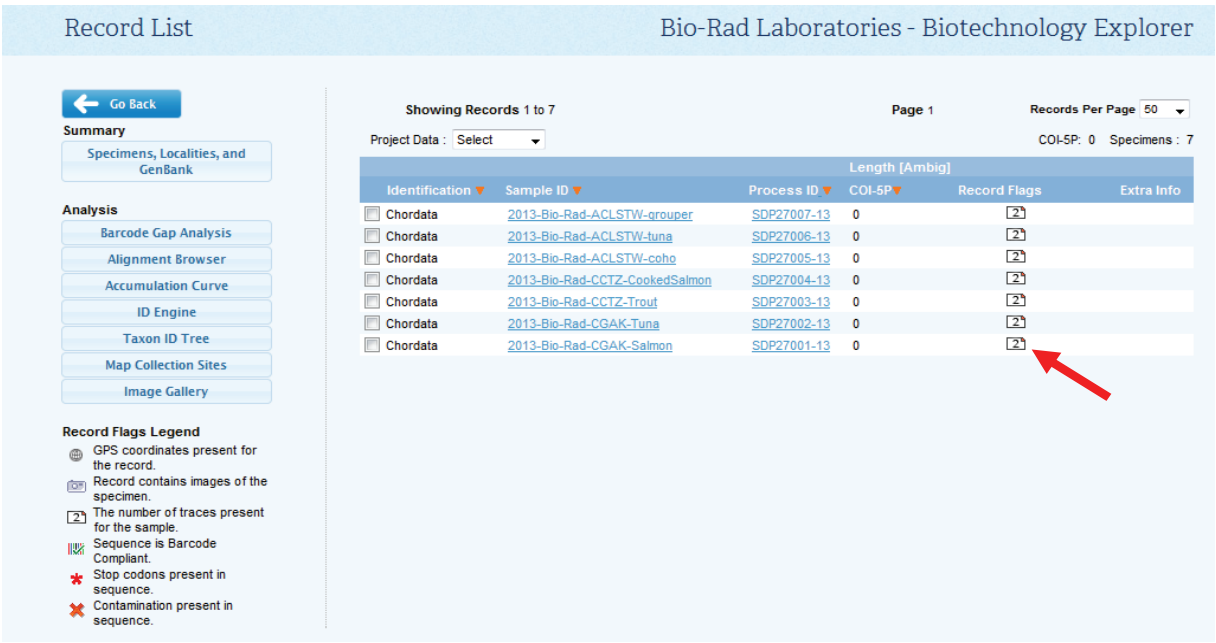
If everything was entered correctly, you will see a submission confirmation page.



i. If you have additional fish samples, click the **Main Console** button. This will open the Data Management page, where you will need to follow steps e–i to upload your traces for your next fish sample. Make sure to use the correct specimen ID for each of the sequencing traces being uploaded.

j. Once all sequencing traces have been uploaded to BOLD-SDP, you can confirm that they are in the database by clicking the **View Class Records** button from the submission confirmation page or **View Data** from the Main Student Console page.





If sequencing trace files were successfully uploaded, there will be an icon showing the number of files for each sample ID.

Since students will need to know their sample ID(s), it might be helpful to print out a screen shot of the Record List page and give this to students.

1 day

Note: If the instructor is uploading all student trace files, repeat this process until all trace files are uploaded to their appropriate specimen ID folders. Then **allow up to 24 hours for the database to calculate quality scores for each trace** before students can begin their analyses.

Step 4: Assessing quality values for student sequencing traces to determine bioinformatics workflow

The BOLD system is designed to carefully screen sequence trace data before fish identification can be made. Because of this, the system does not allow many bioinformatics processes on lower quality sequence data. However, this does not preclude students from learning the steps of sequence analysis, even if their data is of lower quality.

- The first protocol in the Student Instruction Manual involves learning to look at sequence trace files, and this can be done for any sequencing trace, whether high or low quality
- The second protocol in the Student Instruction Manual involves determining the best match — either within the BOLD database and/or using a BLAST search of GenBank — of single sequencing trace files to determine how closely they match if only one sequence trace file is used at a time
- The subsequent protocols involve generating a contig (the best consensus sequence using both the forward and reverse sequencing trace files), trimming off low quality data at the ends, correcting any discrepancies in the contig file, and then determining the best match in the BOLD database to this contig file. If the initial sequencing trace file quality designation was low or fail, these activities cannot be performed with the student data. However, sequencing trace files are available for download at **bio-rad.com/fishbarcoding**, on the download tab, and can be used to complete all of these activities

Students should not be discouraged if their sequencing trace files are of lower quality, but instead should use this as an opportunity to learn about what factors can impact sample purity, sample processing, and sequencing results. There are many places where contamination can occur that are outside the control of the students. Sources of contamination include supermarkets, where fish may have bacterial growth or be cross-contaminated with other fish samples cut with the same knife. Some samples, such as fried fish samples or pickled and canned samples, may have damaged DNA. Pet store “dead fish” samples may have started to decay, destroying the integrity of the DNA within them.

Student Instruction Manual

DNA Sequencing of COI Amplicons Using Dye Terminator Cycle Sequencing

Background — DNA sequencing is a procedure for determining the order in which nucleotides (adenine, guanine, cytosine, and thymine) appear, regardless of whether that DNA consists of small pieces or complete genomes. Over the last several decades, a variety of sequencing methods have been developed for different applications and research goals. A researcher's selection of a particular method is based on a variety of considerations, including speed, cost, accuracy, and the length of the DNA molecule to be sequenced. Dye terminator cycle sequencing — an automated variation of Sanger sequencing — is the method of choice for DNA barcoding. This PCR-based method of automated DNA sequencing is performed at a nominal cost by both commercial and university-based sequencing facilities.

Methodology — for DNA barcoding, two dye terminator sequencing reactions are performed separately for each COI amplicon (PCR product). The forward sequencing reaction will determine the nucleotide sequence of the sense strand of DNA, whereas the reverse sequencing reaction will determine the nucleotide sequence of the antisense strand. Unlike in conventional PCR, only a single oligonucleotide primer is used for each sequencing reaction.

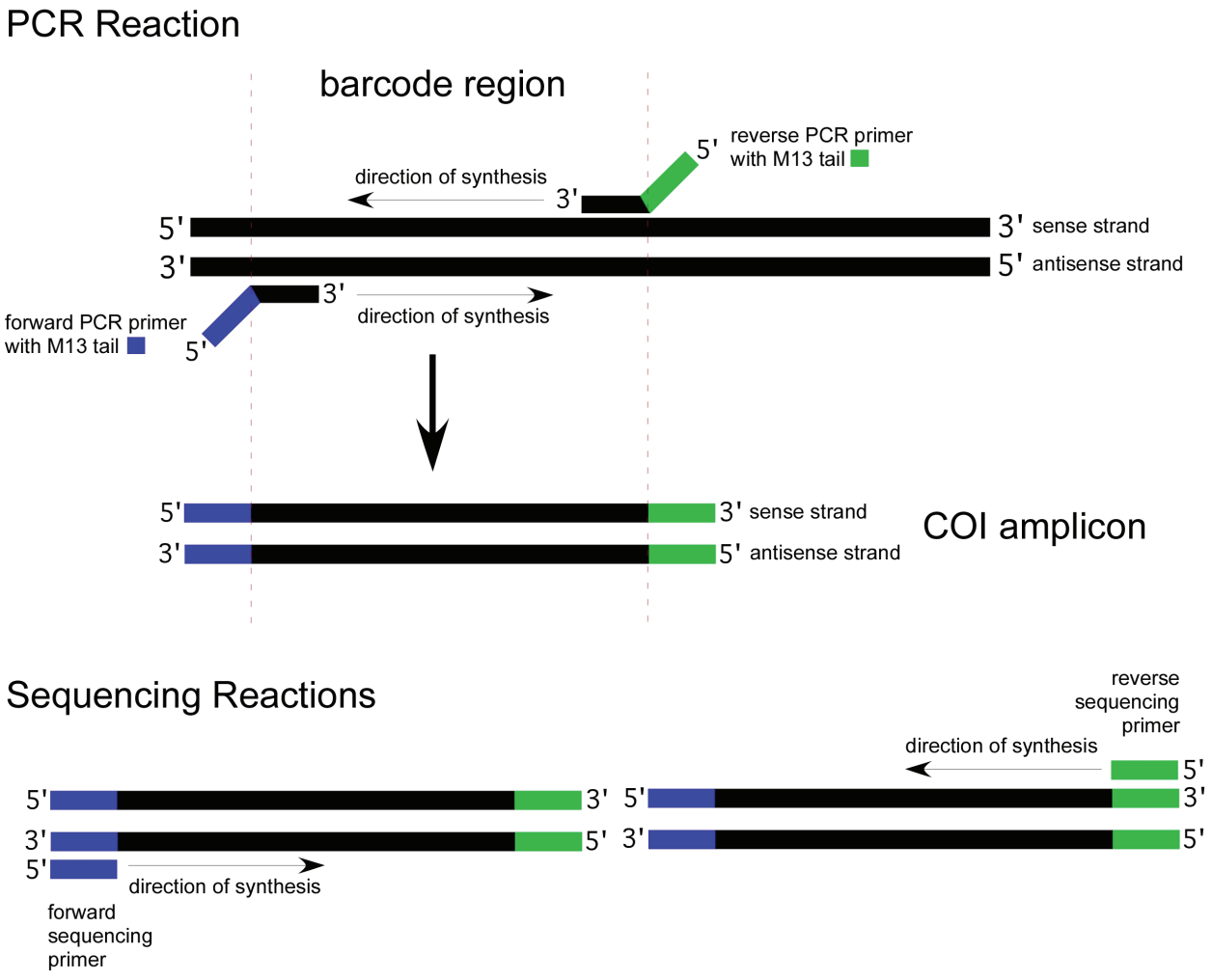


Fig. 1. Both amplification of a portion of DNA and sequencing reactions employ the polymerase chain reaction. PCR amplification of a portion of DNA uses a pair of primers used together in the same reaction with the end result being a final amplicon (double-stranded PCR product) of defined length. Sequencing reactions also use a template of double-stranded DNA; however, sequencing reactions use only one primer per reaction and create single strands of DNA terminated by ddNTPs rather than double-stranded PCR products.

The following components are common to both the forward and reverse sequencing reactions, which are performed in separate tubes:

1. Multiple copies of a double-stranded COI amplicon (the DNA template for each sequencing reaction)
2. A heat-stable DNA polymerase
3. dNTPs (the basic building blocks of DNA)
4. ddNTPs (fluorescently labeled terminator nucleotides that lack an –OH group at position 3 of the ribose ring)

Because the ddNTPs lack an –OH group at position 3 of the sugar, they cannot be involved in further extension of the sequencing reaction PCR product. Therefore once a ddNTP is incorporated, the reaction stops for that one chain. The only component that is different between the forward and reverse sequencing reaction is the single sequencing primer that is used.

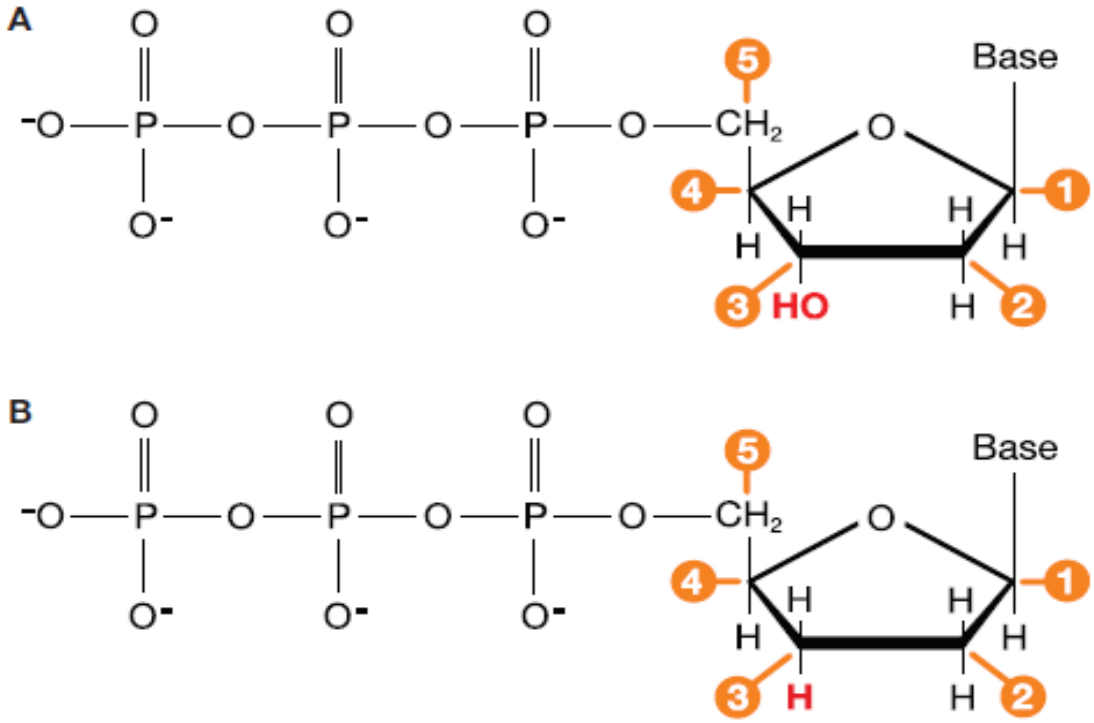


Fig. 2. Structure of dNTPs and ddNTPs. A, dNTPs have a 3'-hydroxyl (3'-OH) group, which is necessary for elongation of DNA. **B**, ddNTPs do not have a 3'-OH; instead, the 3' position has been modified to have a hydrogen (-H) at that position. When a ddNTP is incorporated into a DNA molecule, the synthesis ends at that nucleotide and the DNA chain is terminated. Figure reprinted from Brown K (2011). Biotechnology, A Laboratory Skills Course (Bio-Rad Laboratories), p 186.

Each sequencing reaction progresses through the same major steps of a PCR reaction:

1. During the **denaturation** step, each sequencing reaction mixture is heated to ~95˚C to disrupt the hydrogen bonds that hold the sense and antisense strands of the COI amplicon together.
2. During the **annealing** step, each reaction mixture is lowered to ~50˚C, allowing the sequencing primer to bind to a complementary sequence on one strand of the COI amplicon. The sequencing primer that was added to the forward sequencing reaction binds or anneals to a complementary sequence on the antisense strand according to the base pairing rules. The sequencing primer that was added to the reverse sequencing reaction binds or anneals to a complementary sequence on the sense strand.
3. During the **elongation** step, each reaction mixture is raised to ~72˚C. At this temperature, a heat-stable DNA polymerase finds the 3' end of the sequencing primer and begins joining nucleotides that are complementary to those present in the template strand. For the forward sequencing reaction, the DNA polymerase joins nucleotides that are complementary to those in the antisense strand. For the reverse sequencing reaction, the DNA polymerase joins nucleotides that are complementary to those in the sense strand.

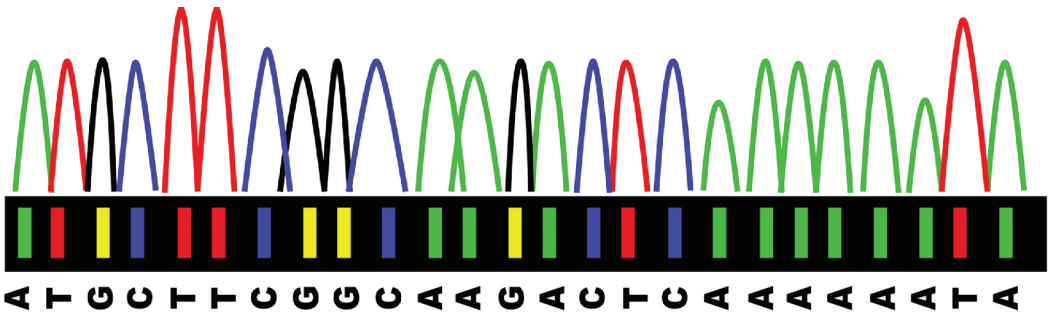
During this step of the sequencing reaction, the DNA polymerase cannot distinguish between dNTPs and ddNTPs present in the reaction mixture. Because a higher proportion of dNTPs are added to each sequencing reaction mixture, they are more likely to be incorporated into the growing DNA chain. However, when a ddNTP lacking a 3' –OH is incorporated, DNA synthesis stops, as no new nucleotides can be added to the growing chain.

The denaturation, annealing, and elongation steps are repeated multiple times, thereby ensuring that at the conclusion of each sequencing reaction, single-stranded DNA fragments of every possible length are generated. Importantly, each fragment terminates with one of the four ddNTPs, which are labeled with a different fluorescent tag.

Upon completion of each sequencing reaction, the fluorescently labeled DNA fragments are separated according to size using capillary electrophoresis, which is electrophoresis performed in a long and extremely narrow tube. As the DNA fragments migrate from smallest to largest through the capillary tube, they pass through a laser, which excites the fluorescent ddNTP at the terminal end of each fragment.

- DNA fragments terminated by ddATP emit green light
- DNA fragments terminated by ddTTP emit red light
- DNA fragments terminated by ddGTP emit yellow light
- DNA fragments terminated by ddCTP emit blue light

The light emitted from fluorescently labeled DNA fragments is detected by the sequencer and represented as a continuous series of colored peaks in an electropherogram, or trace file. The peak from the smallest fluorescently labeled DNA fragment is represented first in the trace file, whereas the peak from the largest fragment is represented last. The information contained in a trace file will be discussed in greater detail below.



Sequence Assembly and Editing in BOLD-SDP

At this point, you have isolated DNA from a fish sample, amplified the *COI* gene from that genomic DNA using PCR, analyzed your COI amplicon using agarose gel electrophoresis, and submitted your COI amplicon for two sequencing reactions — one in the forward direction and one in the reverse direction. You should now have two data files from the sequencing facility for each fish sample COI amplicon you sent for sequencing — one file representing the data from the forward sequencing reaction and one file representing the data from the reverse sequencing reaction. It is now time to analyze that data. The general workflow for analysis of your data is outlined below.

1. Assess the quality of DNA sequencing data.
2. Query either the BOLD database or GenBank for matches for forward and reverse sequencing data.
3. Assemble a single consensus sequence (contig) from your two sequencing reactions.
4. Manually compare any nucleotide calls that are different between the forward and reverse sequencing reactions.
5. Perform a search to determine the identity of your fish sample contig data.

Sequence trace files

As noted, the fluorescently labeled DNA fragments generated during dye terminator cycle sequencing migrate sequentially through a capillary according to their size (smallest to largest) and pass by a laser. Upon exposure to the laser, fragments terminated by a ddATP emit green light, fragments terminated by ddTTP emit red light, fragments terminated by ddCTP emit blue light, and fragments terminated by ddGTP emit yellow light. The light signals are detected by the DNA sequencer, processed by a software program, and represented as a series of colored peaks in a trace file (yellow light signals emitted from DNA fragments terminated by ddGTP are represented as black peaks in the trace file to make them more readable on a white background).

The software also uses an algorithm to assign base calls (nucleotides) to each peak in the trace file and to compute a confidence or quality (Q) score for each base call. The quality score represents the level of confidence that a base call was made correctly. To compute quality scores, the algorithm examines several parameters associated with the shape and resolution of the peak as well as the signal-to-noise ratio at each position in the trace file. The resulting scores are logarithmically linked to error probabilities according to the following equation:

$$Q = -10 \log_{10} P$$

where P represents the probability of an incorrect base call.

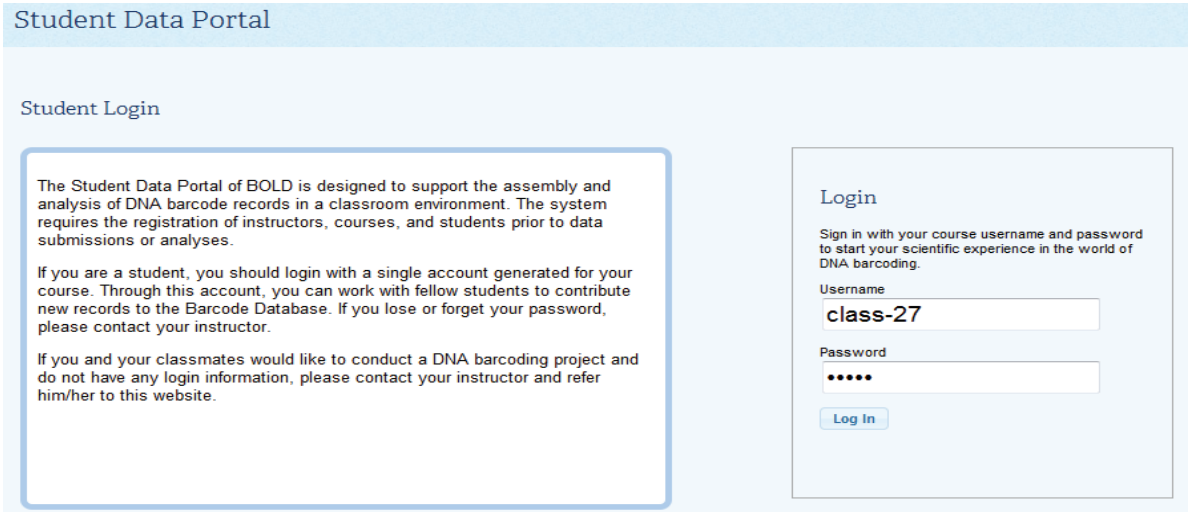
Based on this equation, a quality score of 20 indicates that the probability of an incorrect base call is 1 in 100, whereas a quality score of 40 indicates that the probability of an incorrect base call is 1 in 10,000. Generally speaking, for submission of a vouchered sample sequence, quality scores below 50 are considered unacceptable. This stringency is much higher than for normal sequence assessment, due to the BOLD database's standard of including data of only the highest integrity.

Step 1. Assessing the quality of DNA sequencing data

One of the first steps in bioinformatics is to analyze the quality of the sequencing data. The ability to look at sequencing data and assess whether or not further analyses are worthwhile is a critical skill. For example, did both of your sequencing reactions result in approximately the same length of sequence data? Did both sequencing reactions work well to give you high quality data? It is important to know the quality of a sequencing reaction because, for example, if there are a lot of ambiguous base calls, how can you differentiate between poor quality data or actual base differences between different fish?

a. Go to the Main Student Console page of BOLD-SDP by going to: **http://www.boldsystems.org/index.php/SDP_Home** and clicking the **Students** button to access the Student Login page.

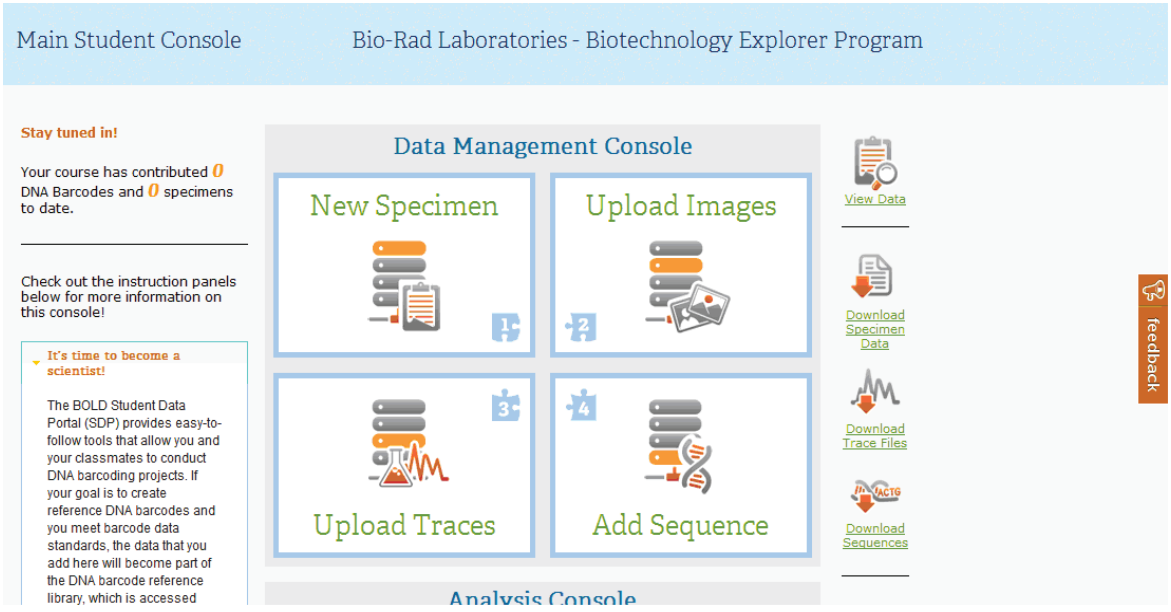
b. Your instructor will provide a course username and password. Enter these into the appropriate spaces and click **Log In** to enter the Main Student Console page. Please note that the password is case sensitive!



Course Username: _____

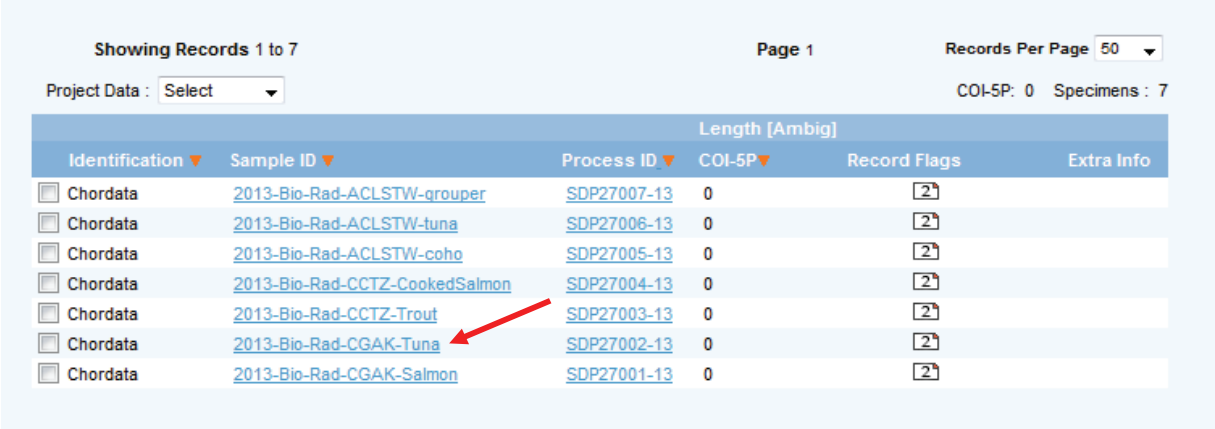
Course Password: _____

c. You should now see the Data Management Console.



d. In the right sidebar, click the **View Data** button.

e. On the Record List page for your class, find the line with the sample ID that corresponds to your first fish sample. If your instructor created these sample IDs, get the sample ID names for your fish from him/her. Record the sample IDs that correspond to your two fish samples.

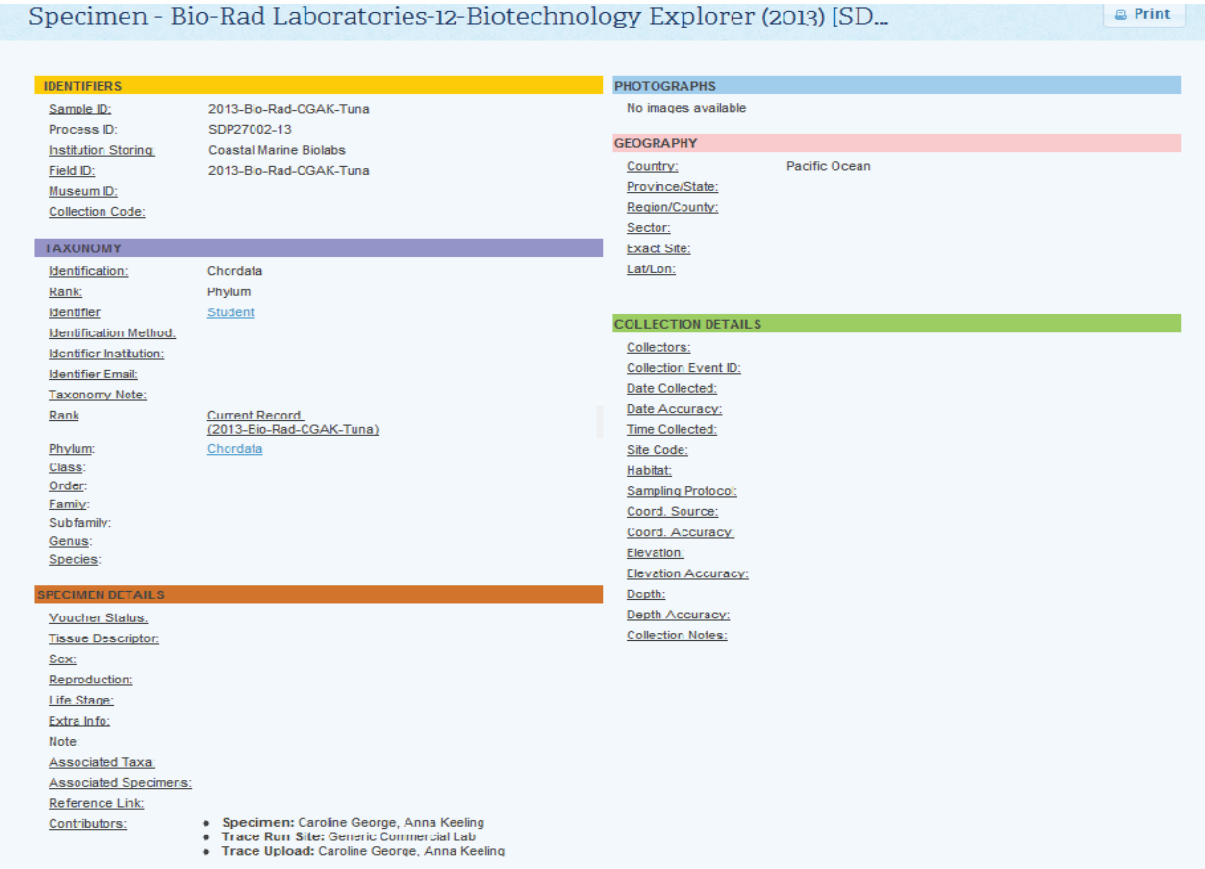


Record Sample ID for Fish 1: _____

Record Sample ID for Fish 2: _____

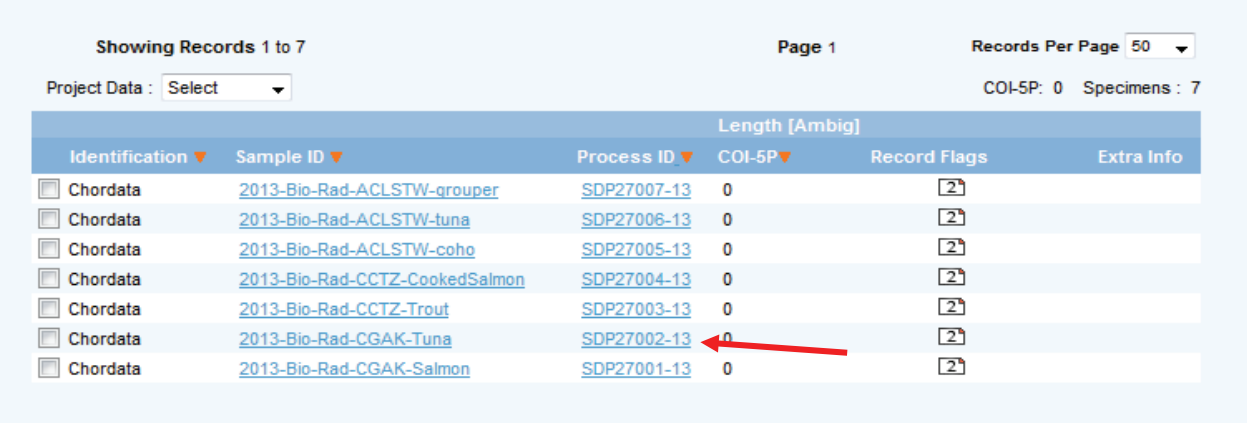
Then, click the **Sample ID** link for your first fish sample.

This will open a new page that includes all the specimen data that either you or your instructor has entered for this sample. Recording this sort of data is critical for submission of new barcode data for fish species because it is a trail of evidence detailing where the fish was found and other information about the sample. This is similar to the evidence trail needed for crime scenes. How much would you trust a DNA sample if the crime scene investigator was not sure exactly what room and location the sample was found in and on what day? It could be from the wrong crime scene!

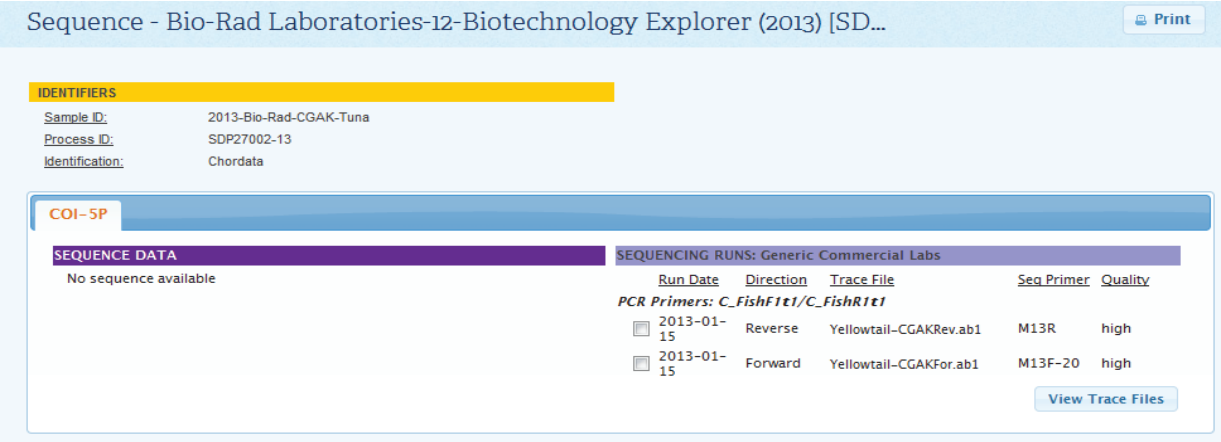


Once you are finished reviewing the data on the Specimen page, close or minimize the page.

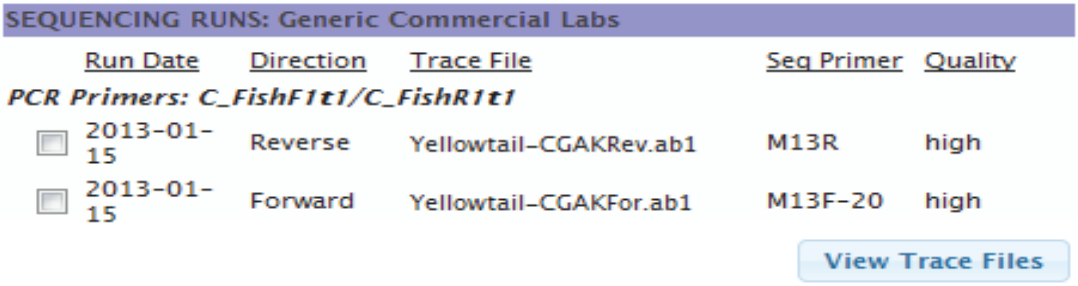
f. Now click the **Process ID** link for your specimen on the Record List page.



A sequence page should open in a new window.



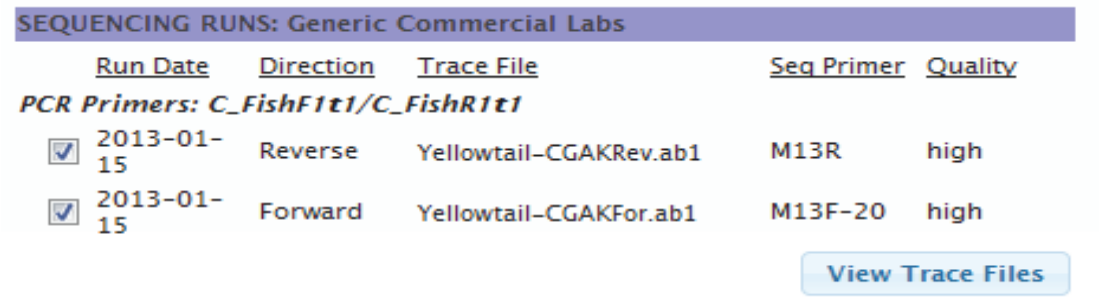
The PCR primer names, sequencing primer names, and trace file names should appear in the Sequencing Runs pane.



BOLD-SDP also displays a quality designation (high, medium, low, or fail) for each of the trace files. These designations are based on the average quality scores of the base calls in each trace file and are designations based on BOLD standards of data quality for uploading.

Note: Trace files with low or failed designations indicate that there might be a problem with the initial fish sample, isolation of the fish DNA, the PCR reaction, or the sequencing reaction. Trace files with low or failed designations generally cannot be used to assemble contigs using the BOLD Sequence Editor. You may nevertheless use the BOLD Identification System (BOLD-IDS) to determine the **possible** identity of the specimen from which your sequence data were generated. Those steps will be outlined in step 2.

g. To examine the sequencing trace files, **select both check boxes** that appear next to their filenames and then click **View Trace Files**.



h. The forward and reverse trace files are displayed in the top and bottom panes of the Trace Viewer page, respectively.

The Trace File Viewer displays quality values for individual base calls in the trace files using a histogram. The quality value for each base call can be determined by comparing the height of its blue shaded bar to the vertical scale on the right-hand side of the trace file window. Continuous stretches of low quality base calls on the 5' and 3' ends of each trace file are displayed in gray.

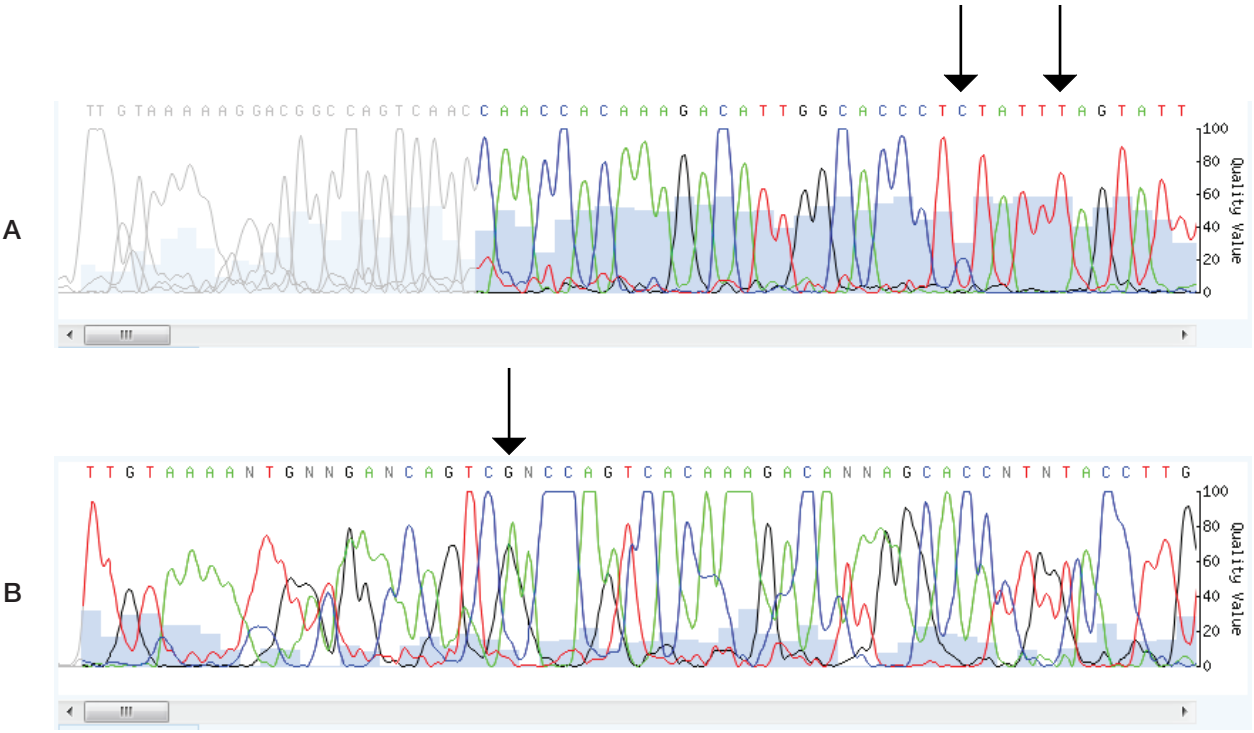
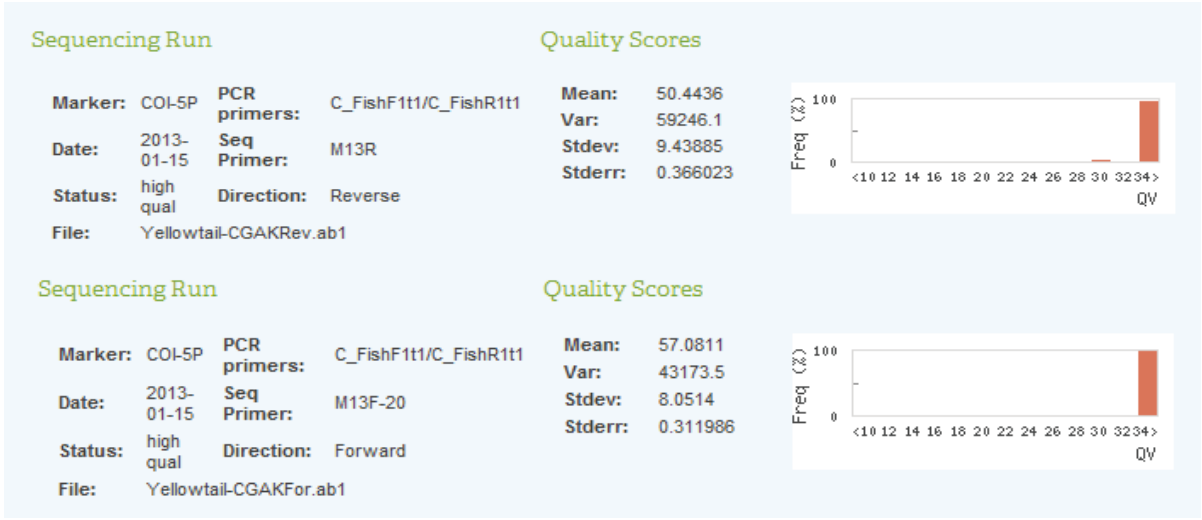


Fig. 3. A comparison of quality values in sequencing trace files. A, the C base call under the first arrow has a lower quality value. The light blue shaded bar shows a quality value of ~30, which means there is a 1 in 1,000 chance that this was the incorrect base call. Notice the peak for this base call is much smaller than the other surrounding peaks. The T base call under the second arrow has a much higher quality value. The light blue shaded bar shows a quality value of ~60, which means there is only a 1 in 1,000,000 chance that this was the incorrect base call. The peak height and shape and the presence of other peaks at the same location (signal-to-noise ratio) impact the quality value score for each base call. **B**, for the second sequencing trace file example, all the quality scores are low (mostly below 20, meaning there is a greater than 1 in 100 chance the call is not correct). While there are some nice tall peaks at each location, there are also lots of underlying peaks. At the base call under the arrow, how confident would you be to call this a G versus an A? According to the quality value score (~10), there is a 1 in 10 chance this was called incorrectly! This is why only high quality data are used for determining barcodes. This DNA sample probably had contamination from another fish sample, which led to two PCR products being generated and then being sequenced at the same time.

BOLD-SDP also computes quality statistics and displays them in tabular and graphical format above each trace file window (see Figure 4). Of these statistical values, the mean and standard deviation are the most informative. The mean refers to the average quality score for the base calls in a given trace file. The standard deviation (Stdev) is a measure of how close the quality scores for the base calls are to the mean. A low standard deviation value indicates that the quality scores are clustered near the mean, whereas a high standard deviation value indicates that the quality scores are dispersed over a large range of values. Lower standard deviation values therefore indicate a greater level of consistency in the quality of base calls, which imparts a higher degree of confidence in the overall accuracy of the trace file. The bar charts that appear above each trace file window show the percentage of base calls that correspond to different quality scores. The data displayed in these charts provide an indication of the range of quality scores for the base calls.

A



B

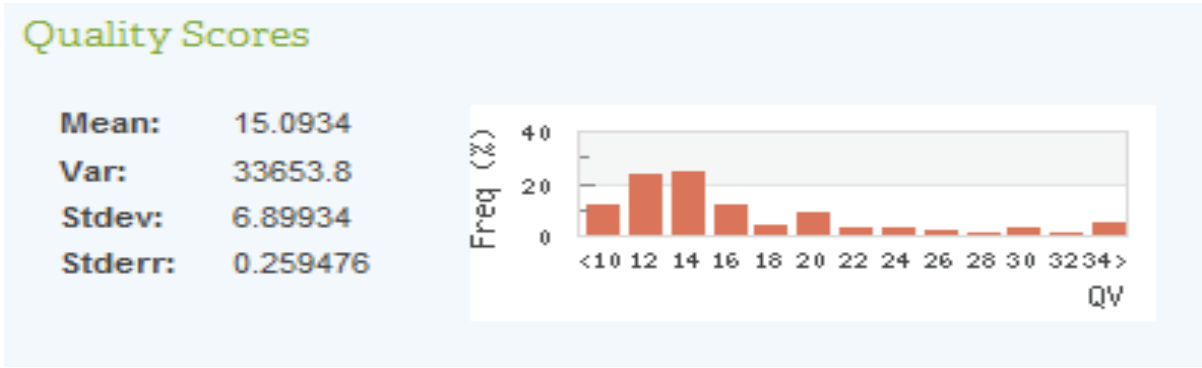


Fig. 4. Histograms and statistics for examples of forward and reverse sequencing trace quality scores. **A**, high quality sequencing traces; **B**, a failed (low quality) sequencing trace. Notice that the mean quality score for the high quality sequencing traces is 50–57, which means that on average, there is less than a 1 in 250,000 chance that the base calls are incorrect. For the low quality sequencing trace, the average quality score is 15 and this would mean that on average there is a 1 in 32 chance that any particular base call is incorrect. The standard deviations are comparable for both the high quality and low quality trace files, which means that overall, the quality is either mostly good (for the high quality traces) or mostly bad (for the low quality trace).

Record the mean quality scores and standard deviations for your first fish sample, then open up the trace and follow the instructions in steps f–h above for your second fish sample.

	Mean Quality Score	Standard Deviation
Fish 1: Forward sequence	_____	_____
Fish 1: Reverse sequence	_____	_____
Fish 2: Forward sequence	_____	_____
Fish 2: Reverse sequence	_____	_____

The scroll bar at the bottom of each trace file window allows you to examine the sequences along their entire length. Even in the absence of quality values for individual base calls in the trace files, the quality of the base calls can be inferred from the resolution of their corresponding peaks. Notice that the peaks in the beginning (5' end) of the forward trace file are broad, overlapping, and poorly resolved. This gray region of the trace file corresponds to low quality base calls, which correlate with high error probabilities.

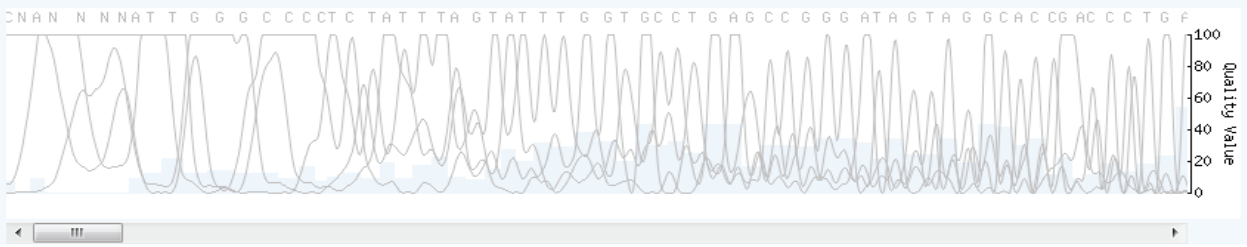


Fig. 5. Poor quality sequence at the beginning of the forward sequencing reaction trace. These base calls are colored gray and most bioinformatics programs automatically trim them off, meaning that they will not be used in further analyses.

As you scroll to the right, the non-gray peaks appear sharp, well resolved, and non-overlapping. This region of the trace files corresponds to high quality base calls, which correlate with low error probabilities.

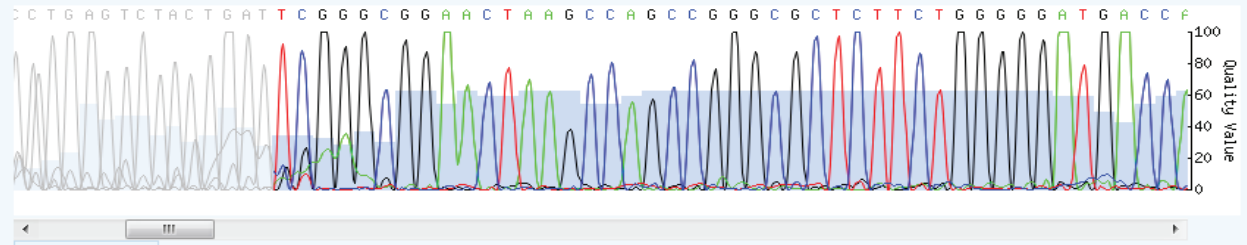


Fig. 6. The beginning of the good quality sequence. All of these base calls have high quality values, the shapes of the peaks are very uniform, and there is little to no overlap of base calls at each peak region.

As you scroll even further to the right (toward the 3' end of the trace files), the peaks become lower in amplitude and begin to broaden and overlap. This gray region of the trace file corresponds to low quality base calls.

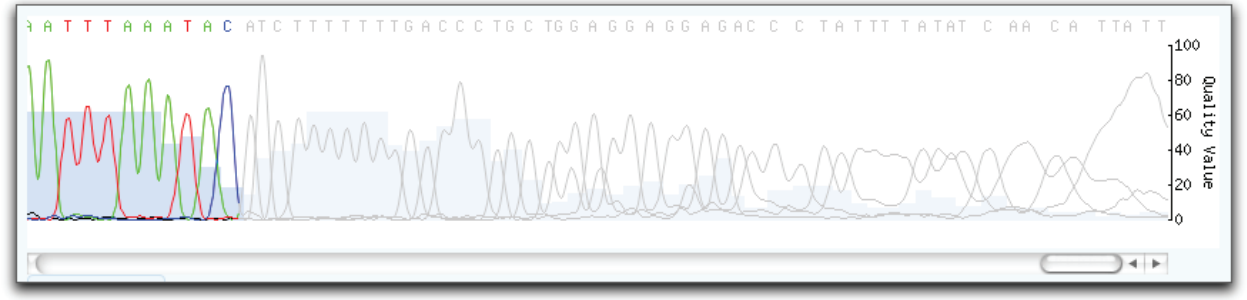


Fig. 7. The 3' end of the forward sequencing reaction trace where the quality values again drop for each base call.

The low quality base calls that appear at the beginning and end of a trace file arise from technical limitations of dye terminator sequencing. These limitations are caused by and complicated by a variety of different interacting factors associated with capillary electrophoresis and the underlying chemistry of this particular sequencing method. Regardless of their causes, low quality base calls must be eliminated from the sequence in order to preserve its overall accuracy.

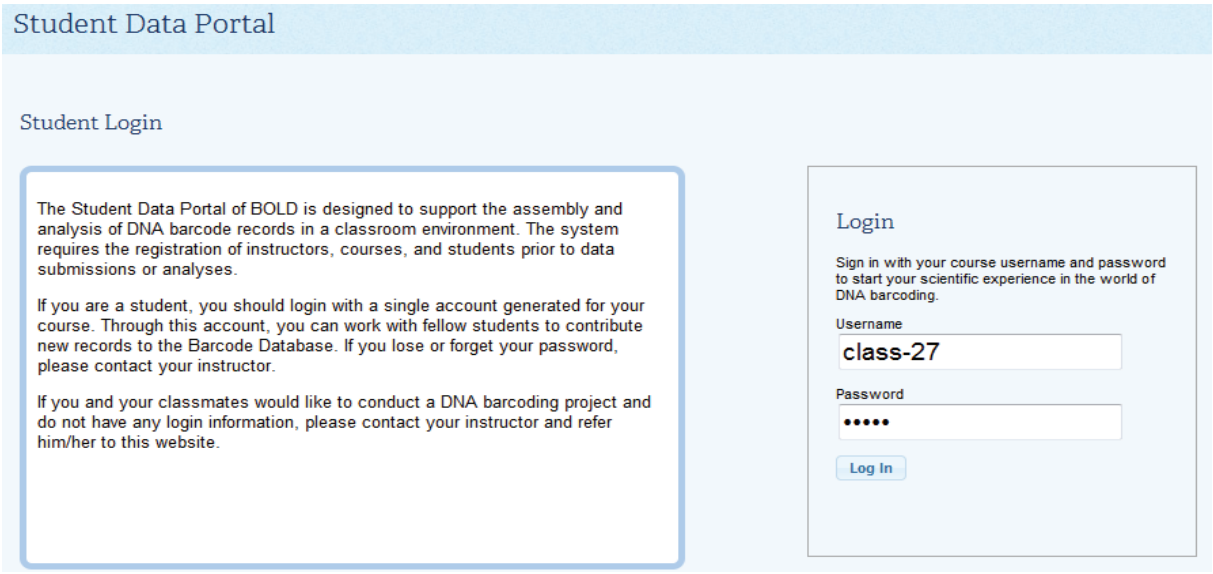
Step 2. Query either the BOLD database or GenBank for matches to forward and reverse sequencing data

You have now looked at your traces and have recorded the mean quality scores for each trace for each fish sample and have also learned about how to assess the quality of sequencing traces. Now it is time to compare your forward and reverse fish sequencing trace results to sequences that have been deposited into two different databases. The first is BOLD, which contains sequences for vouchered samples that require an extremely high level of quality to be considered for submission. The second is GenBank, which is one of the world's largest repositories of DNA, RNA, and protein sequencing data. The BOLD database can be searched by using BOLD-SDP and if a sequence match is not found, then GenBank can be searched using BLAST (basic local alignment search tool) programs, which find short regions where pairs of sequences match.

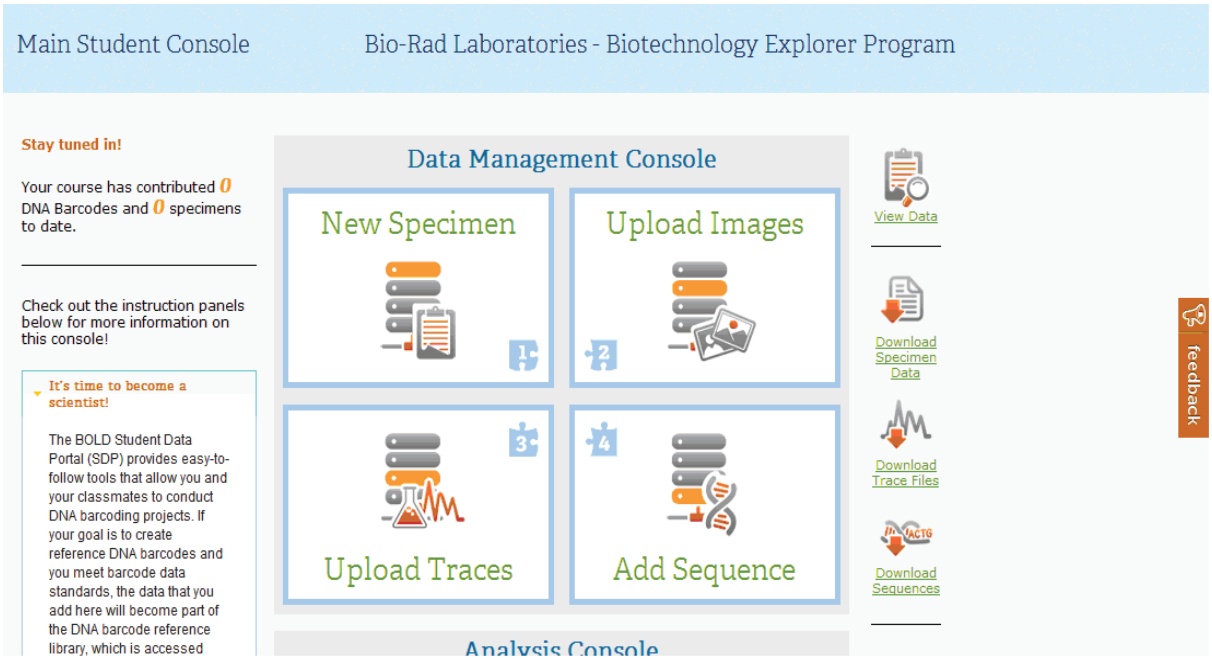
The **blastn** program in GenBank is used to compare a nucleotide sequence to a database of nucleotide sequences. Here blastn will be used to compare your forward and reverse sequences to the GenBank database of all nucleotide sequences. Once the searches are complete, blastn counts all the nucleotides in the matching region and awards two points for every pair of bases that match. If one sequence has an insertion, a deletion, or a gap (more than one base missing) relative to the other, blastn takes points away from the score. The net result is that a blastn score is equal to two times the length of the matching region. The completed search will return a blastn score and E value for each match of your query sequences relative to sequences in the GenBank database. The results also include an alignment of your sequences to each match in the database so that you can compare them. The meaning of the blastn scoring will be explained in more detail below.

a. Go to the Main Student Console page of BOLD-SDP by going to: **http://www.boldsystems.org/index.php/SDP_Home** and clicking the **Students** button to access the Student Login page.

b. Enter the course username and password in the appropriate spaces and click **Log In** to enter the Main Student Console page. Please note that the password is case sensitive!

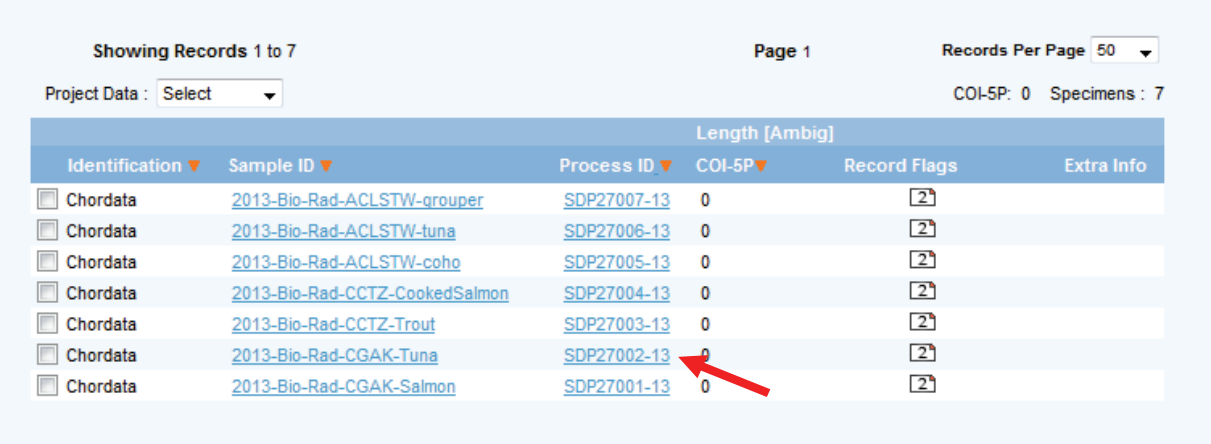


c. You should now see the Data Management Console.

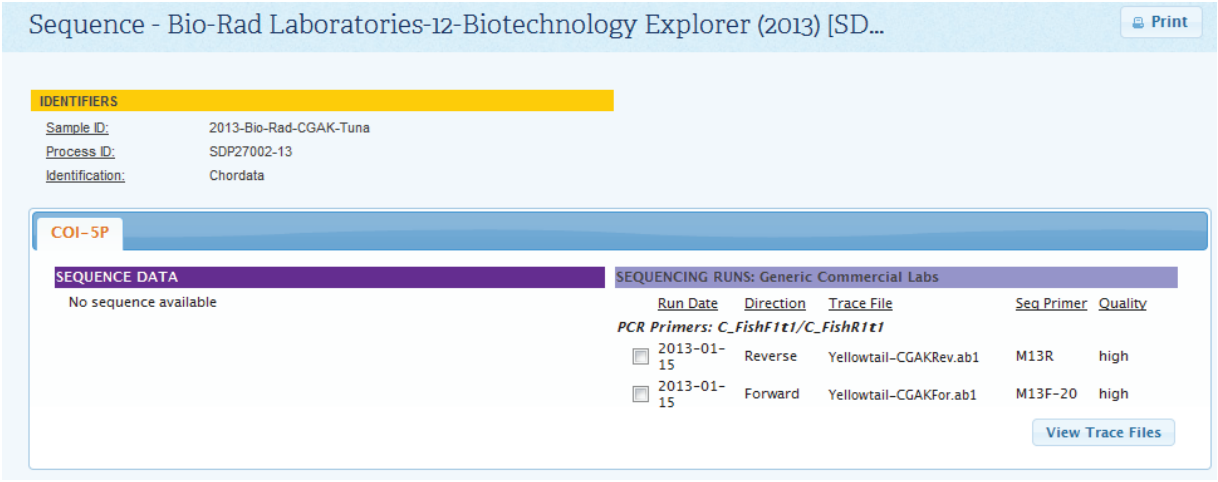


d. In the right sidebar, click the **View Data** button.

Now click the **Process ID** link for your specimen on the Record List page.



A sequence page should open in a new window.



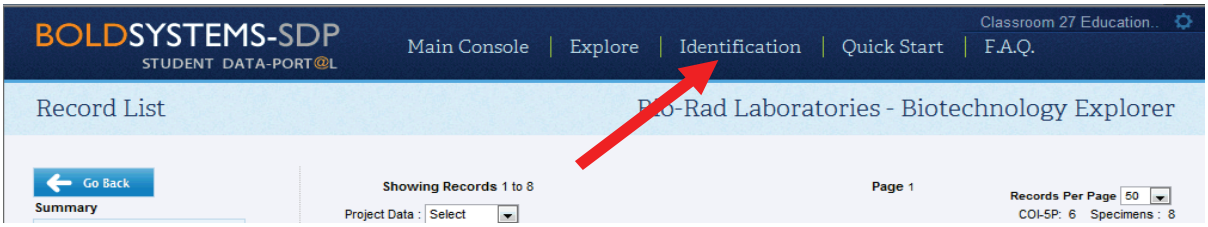
e. Open the sequencing trace files by **selecting both check boxes** that appear next to their filenames and then click the **View Trace Files** button.

f. The forward and reverse trace files are displayed in the top and bottom panes of the Trace Viewer page, respectively.

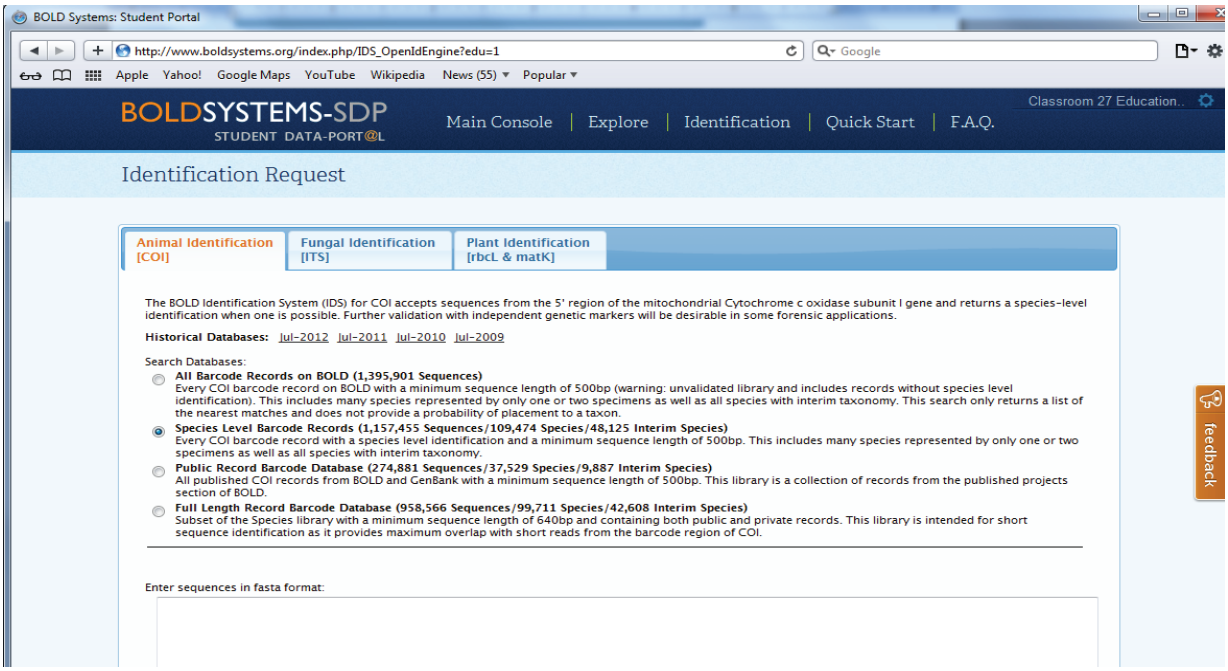
g. Click the **View Sequence** button for your upper sequence. This will bring up a text window that contains the sequence generated by one of your sequencing reactions. Wherever there is an “n”, the software could not determine what the base call should be for this location.

h. Highlight the text in this window and then copy it by typing **Control+C** (PC) or **Command+C** (Mac).

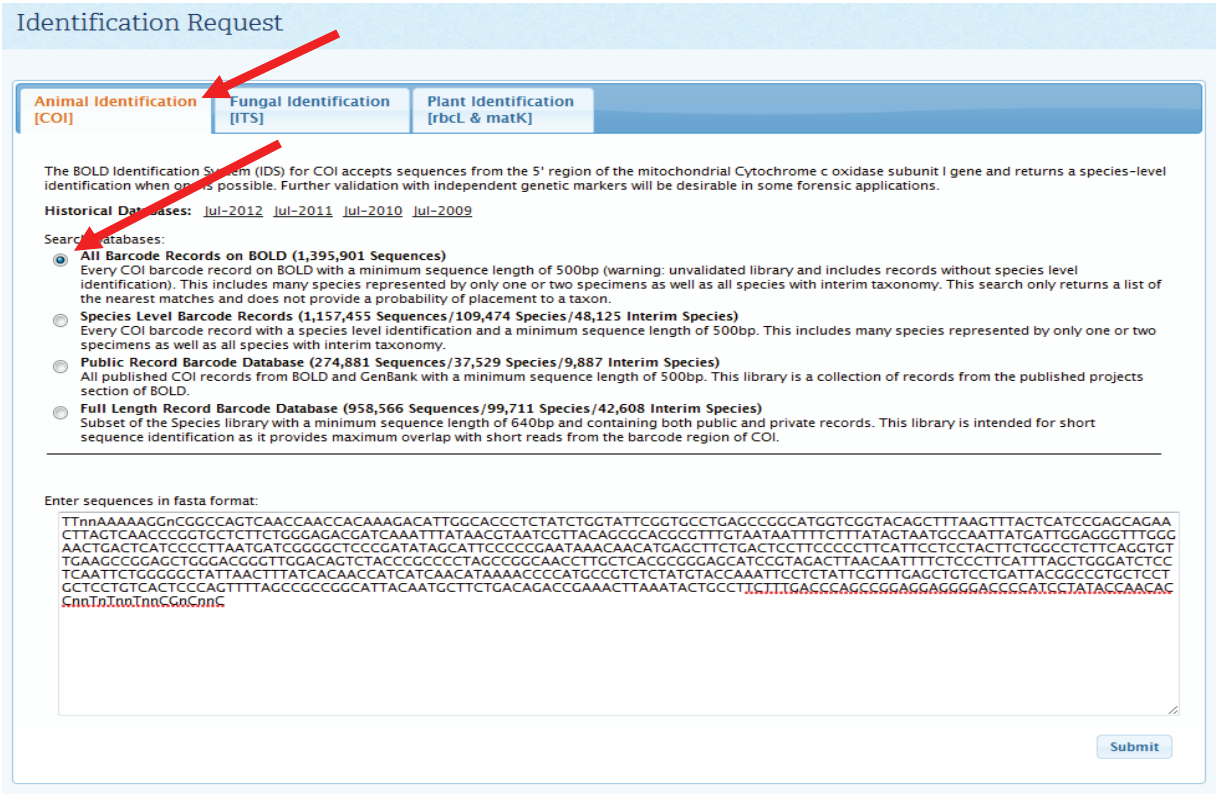
i. In the BOLD-SDP Record List window, click the **Identification** tab.



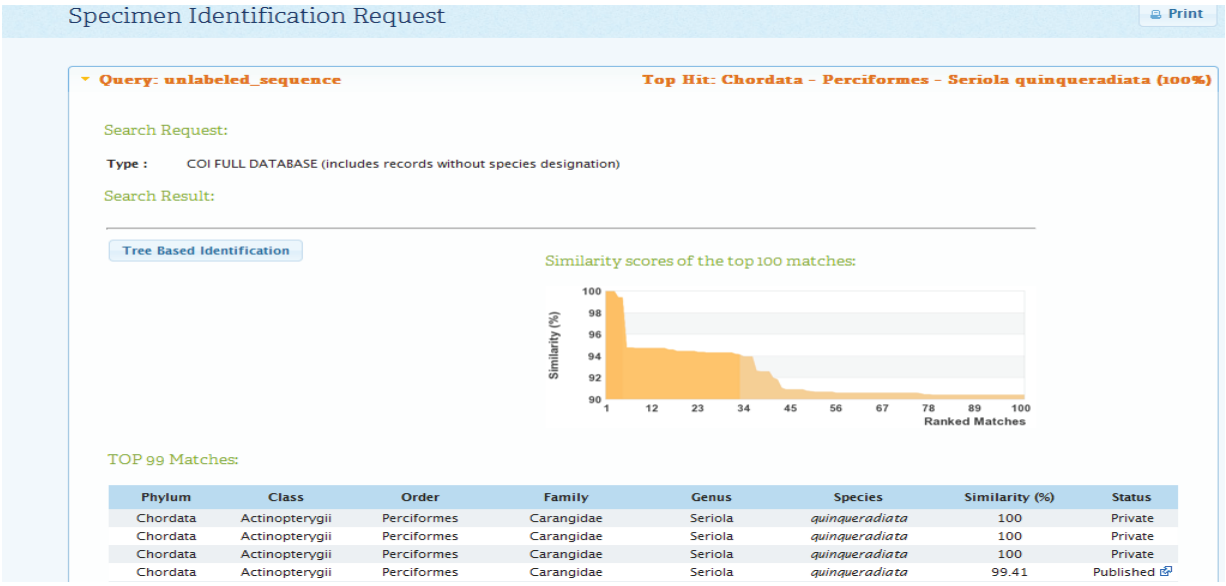
j. This will open up a new window for searching the BOLD database.



k. Make sure the Animal Identification (COI) tab is chosen. The default search is for the Species Level Barcode Records. Change this selection to **All Barcode Records on BOLD** and then paste your sequence (**Control +V** on a PC or **Command +V** on a Mac) into the box labeled **Enter sequences in fasta format** and then click **Submit**. A sequence in FASTA format contains a single line of description followed by lines of sequence data. While lacking some of the formatting associated with FASTA format sequences, your pasted nucleotide sequence alone is sufficient for submission.



l. A window containing the search results will open if a match can be found. Because you submitted sequence data from only a single sequencing reaction, BOLD-IDS may be unable to return a conclusive species level match for your identification request.



If a match is found, record the phylum, class, order, family, genus, species (or any level of taxonomical match that was found) and % similarity for the top match to your sequence.

Sequencing trace file name: _____

Phylum: _____

Class: _____

Order: _____

Family: _____

Genus: _____

Species: _____

% Similarity: _____

m. Repeat steps g–n for your reverse sequencing trace file.

Sequencing trace file name: _____

Phylum: _____

Class: _____

Order: _____

Family: _____

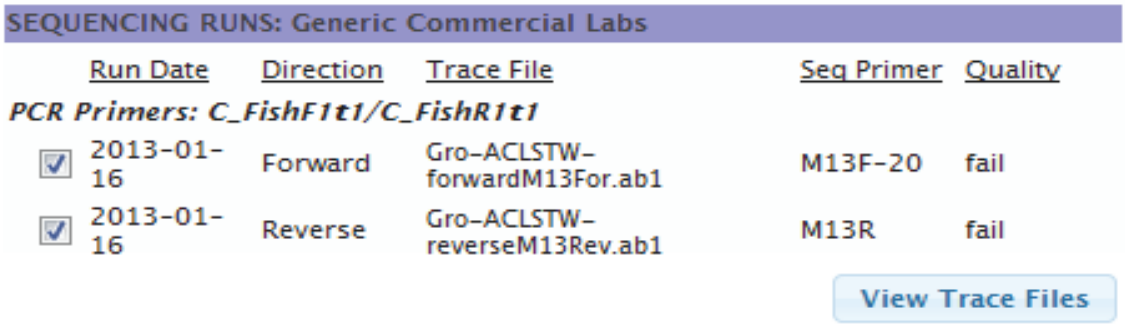
Genus: _____

Species: _____

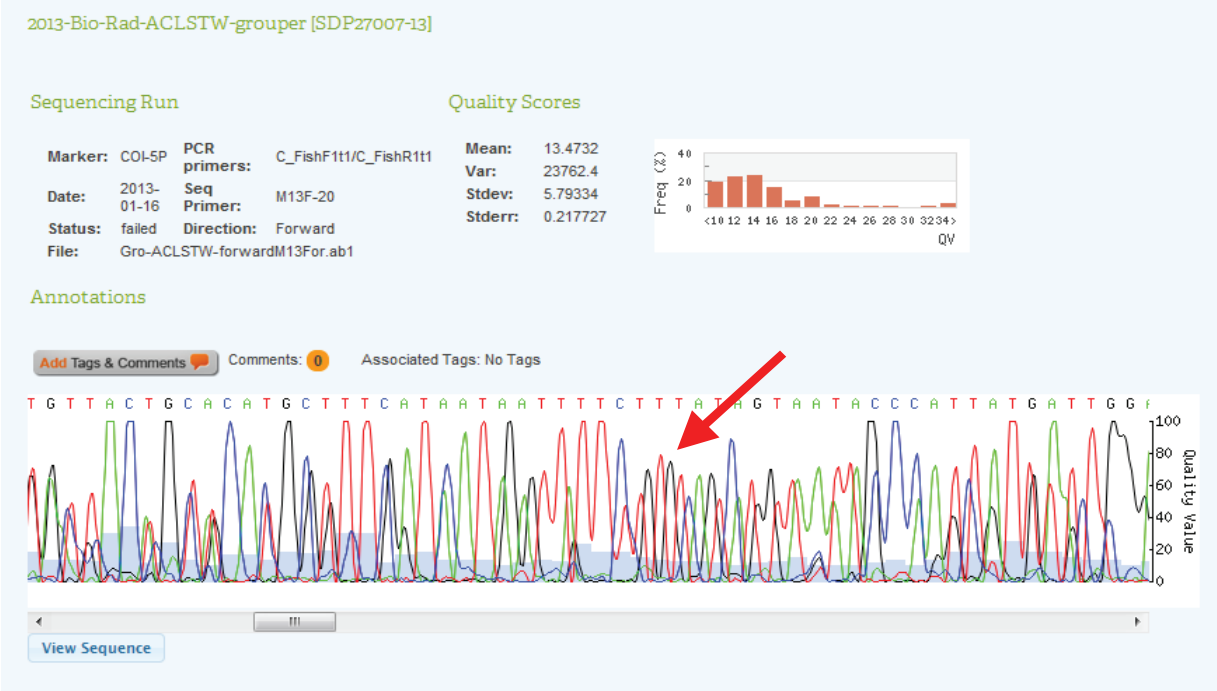
% Similarity: _____

If you got two matches to the genus or species level, were they the same species and genus? Did they have the same % similarity?

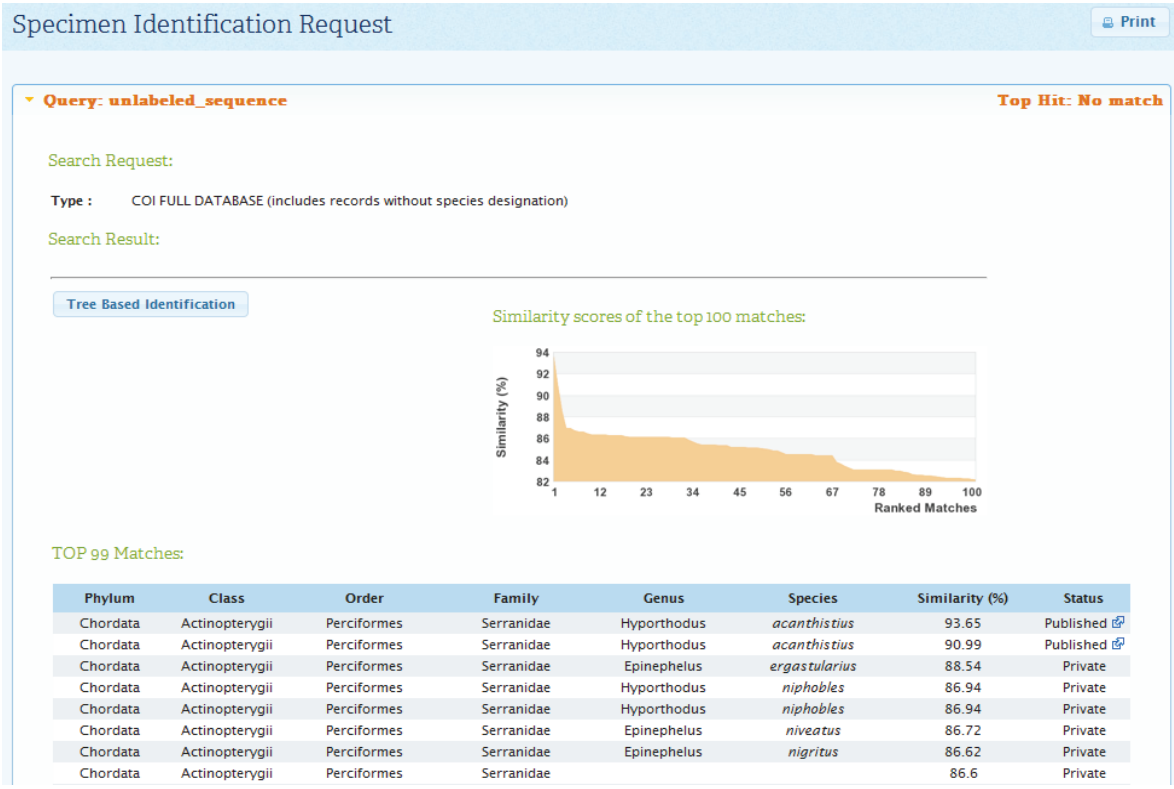
If you have lower quality data, it might not be possible to find a match using the BOLD database. The example listed above was for high quality sequences. The one shown below is for low quality sequences.



In this case, both sequences have failed quality data. The sequencing trace looks like an overlap of two different sequences, and this means that at some point, contamination occurred. This contamination might have occurred when the fish was prepared for sale, during isolation of a piece, during DNA isolation, during PCR, or during sequencing. Where the contamination occurred cannot be determined, just that there was contamination.

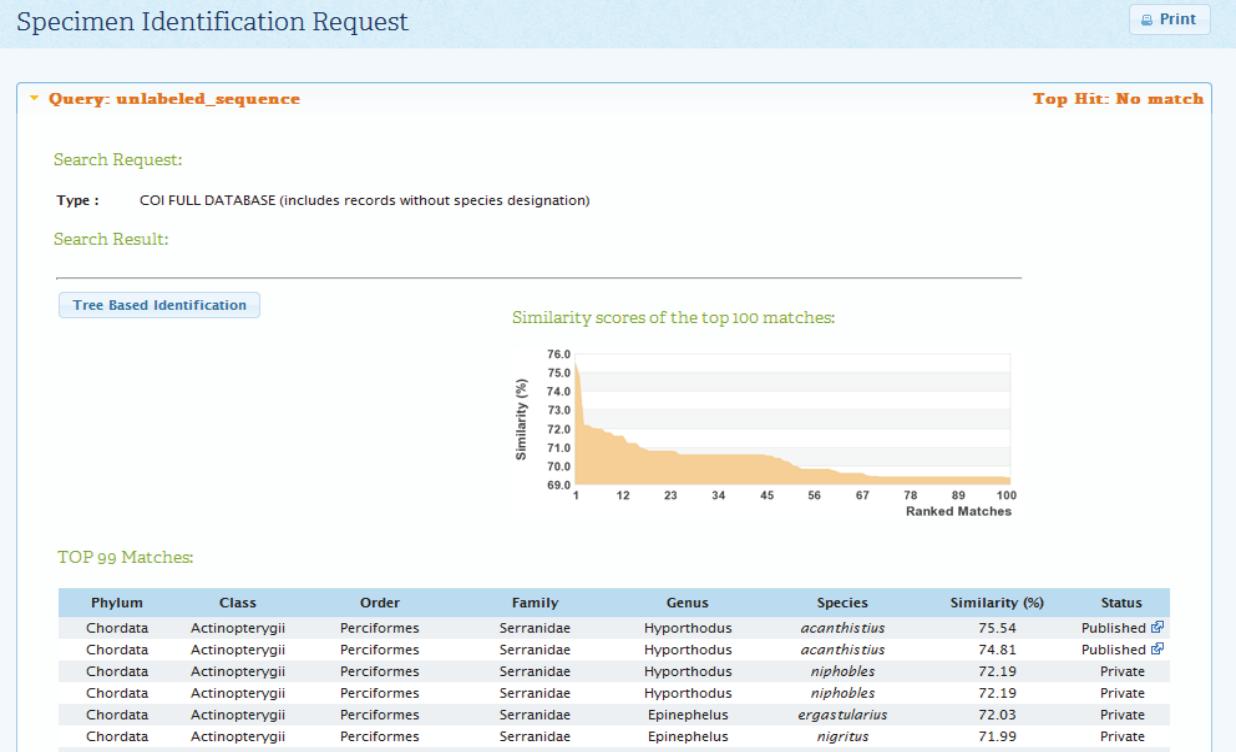


If the sequence for this trace is submitted for identification in the BOLD system, no very strong match is found. In fact, the top matches include different genera.



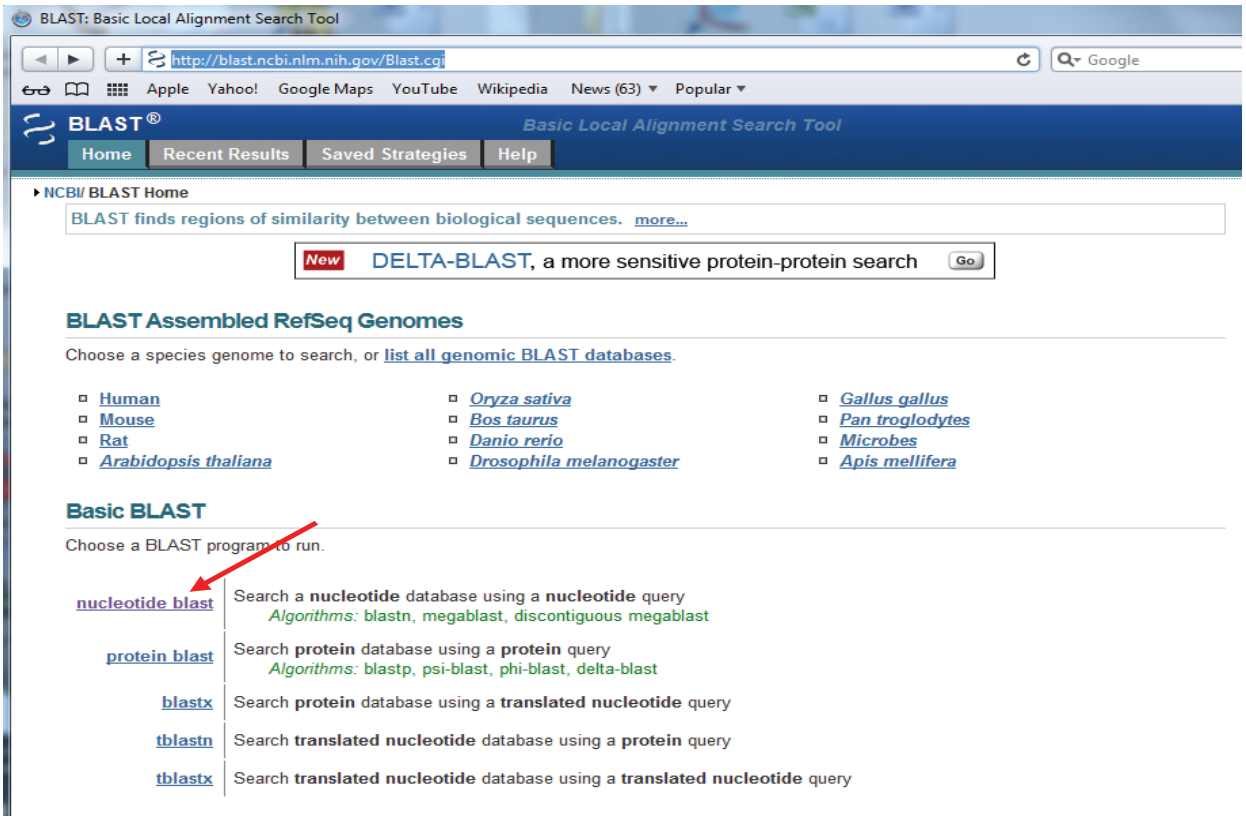
Phylum	Class	Order	Family	Genus	Species	Similarity (%)	Status
Chordata	Actinopterygii	Perciformes	Serranidae	Hyporthodus	*acanthistius*	93.65	Published
Chordata	Actinopterygii	Perciformes	Serranidae	Hyporthodus	*acanthistius*	90.99	Published
Chordata	Actinopterygii	Perciformes	Serranidae	Epinephelus	*ergastularius*	88.54	Private
Chordata	Actinopterygii	Perciformes	Serranidae	Hyporthodus	*niphobles*	86.94	Private
Chordata	Actinopterygii	Perciformes	Serranidae	Hyporthodus	*niphobles*	86.94	Private
Chordata	Actinopterygii	Perciformes	Serranidae	Epinephelus	*niveatus*	86.72	Private
Chordata	Actinopterygii	Perciformes	Serranidae	Epinephelus	*nigritus*	86.62	Private
Chordata	Actinopterygii	Perciformes	Serranidae			86.6	Private

The second sequence trace identification is even more ambiguous with even less similarity.

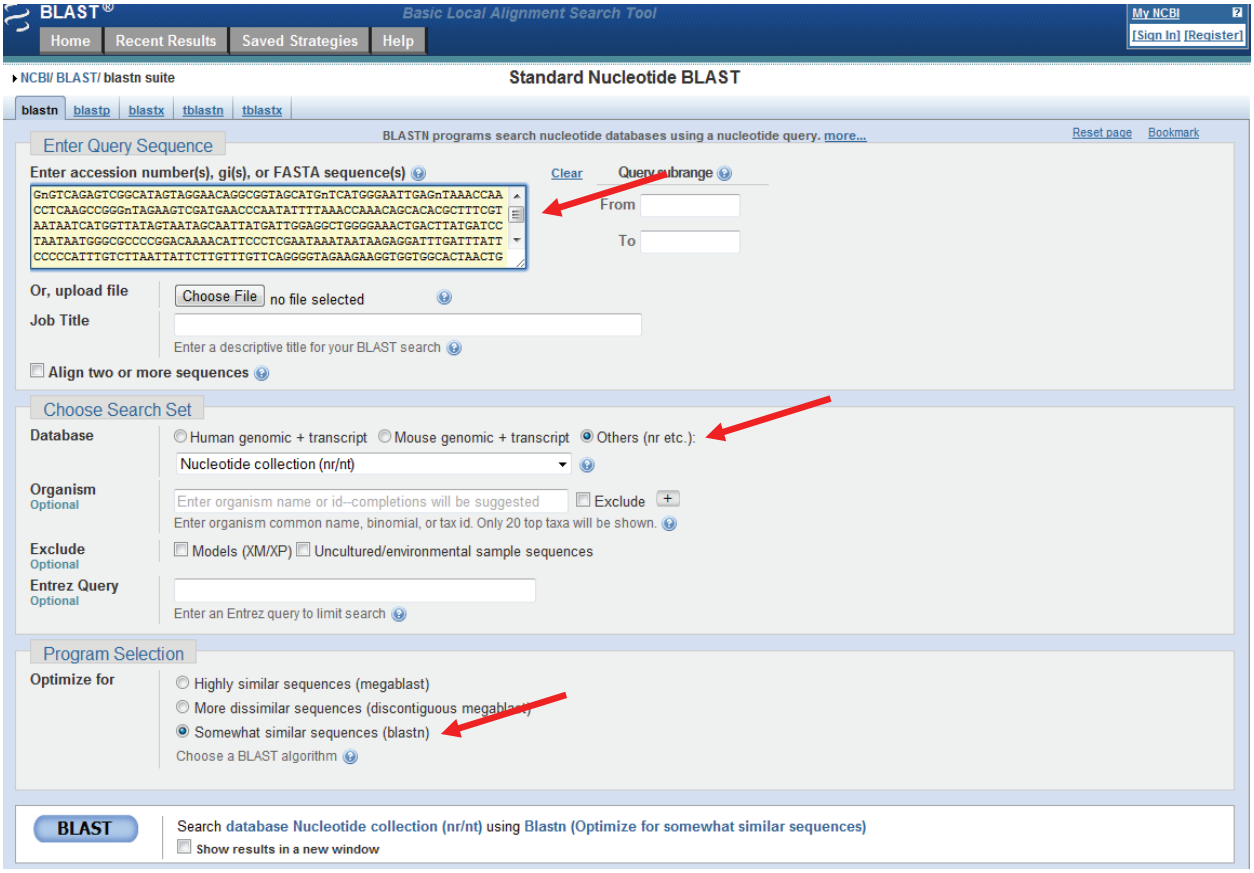


The reason for such low matches can be found if a BLAST search is performed using GenBank.

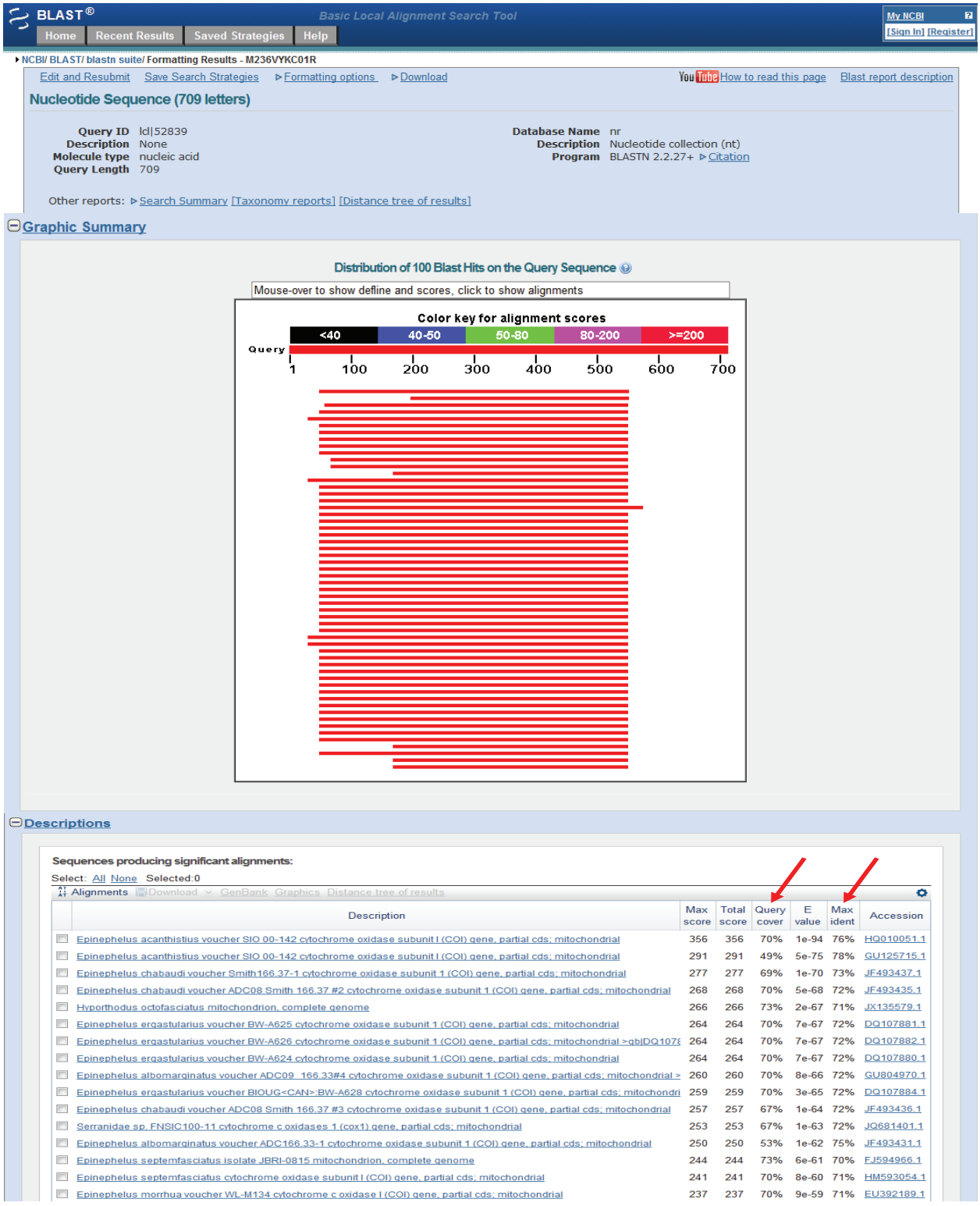
Open a new internet browser window and go to **http://blast.ncbi.nlm.nih.gov/Blast.cgi**. Click **nucleotide blast** to open a window to search the nucleotide database (as opposed to a protein database).



Under **Enter Query Sequence**, paste your sequence into the **Enter accession number(s)** box. Under **Choose Search Set**, make sure **Others** is selected for Database and that **Nucleotide collection (nr/nt)** is selected in the dropdown box. All other boxes under Choose Search Set can be left blank. Under **Program Selection**, choose **Somewhat similar sequences (blastn)**. Then click the blue **BLAST** button to begin the search.



Your results will appear after the database has been searched.



An explanation of BLAST search results follows below.

Understanding blastn results

The results from a blastn search include many different kinds of information and statistics. These bits of information include the size of the database, length of each query sequence, statistics that describe the number and percent of matching bases, a BLAST score, and the E value.

At the top of the blastn results page is a graphical representation of the results. A thick red bar represents the full length of the query sequence. In this case, the grouper contig is 709 bases long.

Below the thick red query bar are thinner bars of various colors that represent sequences from the database (subject sequences) that align with the query sequence. The accession and description of the subject sequence is given in the box above the graph when the bar is moused over. The colors of these bars show the degrees of alignment using the color key above the red query bar. The colors are based on the blastn max scores. The subject sequence with the highest max score aligns at the top.

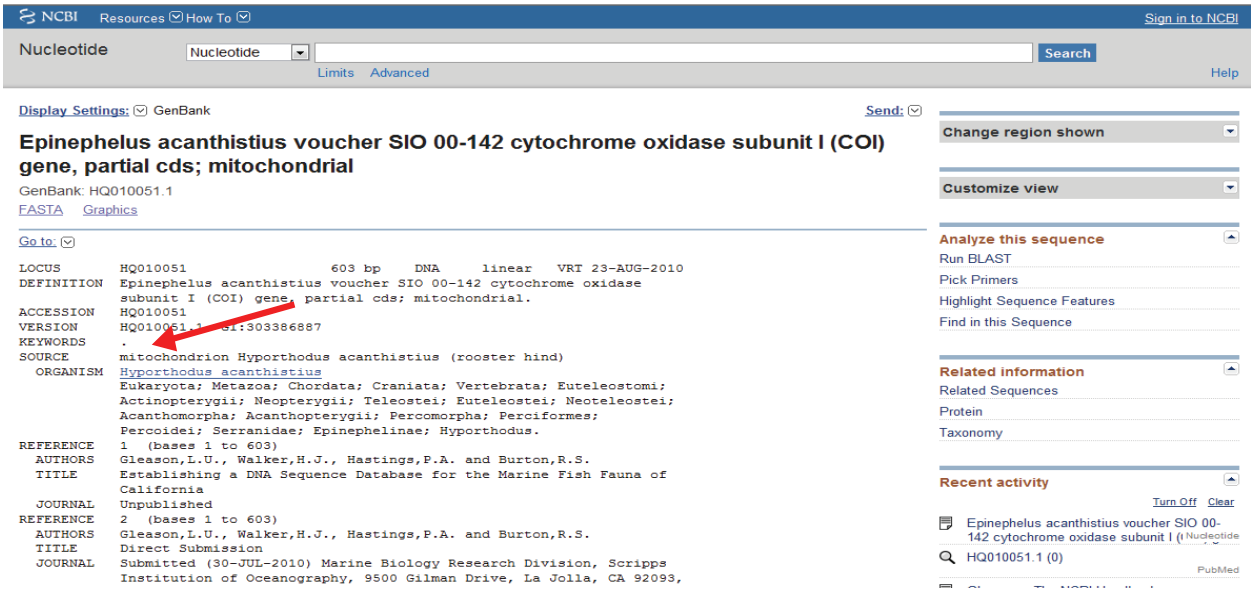
If the subject sequences do not continuously match the query, the colored bars are connected by thin gray lines representing regions where there is no homology to the query.

Next in the blastn results page is a table that summarizes the statistics. Each row contains a matching sequence with the best-matching sequence at the top of the table.

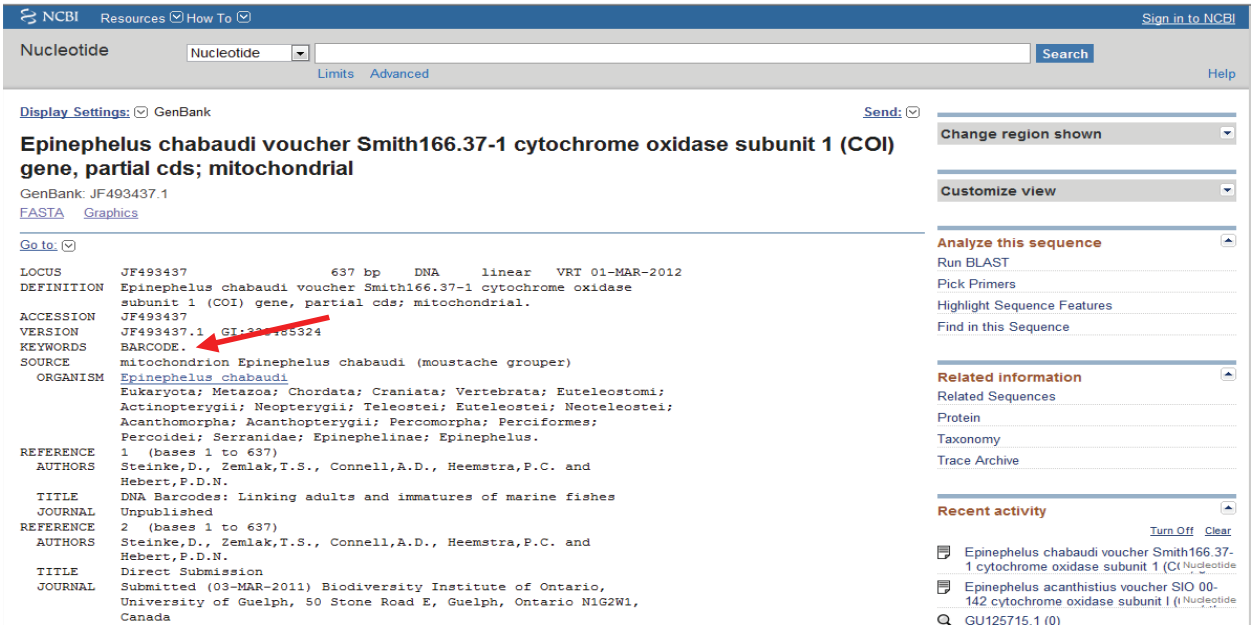
Accession — an accession number is the unique identifier given to a DNA sequence when it is submitted to a database. (It can also refer to a submitted protein sequence.) The submitted data can be for an entire genome, a chromosome within a genome, an entire mitochondrial genome (such as JX135579.1 from the sequences listed above for *Hyporthodus*) or it can be for shorter sequences such as the sequences for vouchered samples listed (such as sample JF493437.1 for *Epinephelus chabaudi*).

As previously mentioned, though GenBank is one of the world's largest repositories of DNA, RNA, and protein sequencing data, this database lacks the same stringency employed by the BOLD database. BOLD requires additional data to be included with reference sequences published in its database to ensure accuracy of the species identified within it, while GenBank does not. For instance, within GenBank, some sequences may be labeled as voucher specimens, yet they do not adhere to the requirements for vouchered sample data as set forth by BOLD. The best way to determine whether a sequence match in GenBank has truly adhered to the stringent requirements of the BOLD database is to examine the Keywords line in the accession entry. Entries that contain the word BARCODE in the Keywords line adhere to BOLD data standards, while entries lacking this designation may not adhere to BOLD data standards.

This sequence for *Epinephelus acanthistius* does not contain the BARCODE designation, though the name of the entry includes the word voucher in its description. This sequence may not have the additional data required by BOLD to truly serve as a reference barcode for species identification.



This sequence for *Epinephelus chabaudi* does contain the BARCODE designation, and this indicates that the entry conforms to BOLD data standards.



Description — the description refers to the source of the matching sequence. In the case of the *Hyporthodus octofasciatus*, or eightbar grouper, sequence match, the complete mitochondrial genome of this species of grouper has been sampled. Other sequences are from COI barcodes and come from different species such as *Epinephelus acanthistius*, rooster hind (HQ010051.1), *Epinephelus chabaudi*, moustache grouper (JF493437.1), *Epinephelus ergastularius*, sevenbar grouper (DQ107881.1), and *Epinephelus albomarginatus*, white-edged grouper (GU804970.1).

Max score — each of the colored bars in the BLAST alignment graph (at the top of the BLAST search results page) have been assigned a score based on the extent of the match. The max score comes from the block of aligned sequence that had the highest score. Because the blastn score is about twice the number of matching nucleotides, it is possible to infer that the maximum score of 356 for the top sequence represents either approximately 700 matching bases or a longer region that contains gaps.

Total score — the total score is obtained by adding the scores from the region of the query sequence that matches any region on the sequence in the database. In this example, since the *COI* gene is a mitochondrial gene, there should not be long gaps of nonmatching sequence followed by large stretches of matching sequence, so this score will be comparable to the max score. Total score is more important when trying to match different genomic DNA sequences that include both exon sequences expected to have higher similarity and intronic sequences that are not expected to have much similarity between different species. Since there are no introns in mitochondrial genes, the total score should be the same as the max score for your DNA barcode sequences.

Query coverage — the query coverage corresponds to the fraction of the entire query sequence that is matched by parts of the subject sequence. In this case for the top match, 70% of the query matches the subject sequence (the sequence in GenBank). The query that was submitted was 709 bases long and 482 of these bases were found to align with the subject sequence in the database.

E value — the E value can have such a large range that it is reported as a power of 10 (expressed as an exponent; for example e-2 means 10^{-2}). For each subject sequence (match in the database), the E value represents the number of equally good matches to the query sequence that would be expected in a database of the same size containing random sequences. When E values are below 1, they can be translated to the probability that two sequences will match to the same extent.

This would mean that with an E value of 0.01, there is a 1% chance of finding an equally good match in a database of random sequences. While low E values are good, high E values suggest that it is possible to find an equally good match by random chance. In the top row of this example, the E value is 1e-94. This means that there is an essentially 0% chance of finding this match in a database of random sequences. In other words, a match is statistically very unlikely to occur by random chance.

Two additional factors have a strong influence on E values: the length of the sequence, because it is easier to find a perfect match to a shorter sequence than it is to a longer sequence, and the size of the database, because it is easier to find a match in a larger database than it is in a smaller one.

Max identity — this column shows the block of a sequence that has the highest percentage of matching bases. In this example, the maximum identity of any matching block is 76% with the *Epinephelus acanthistius* voucher COI sample (HQ010051.1). This sequence has 381 of 503 aligned bases matching and three gaps as well. This can be seen if you scroll down the GenBank search page until you reach this first match.

Links — the final column in the blastn alignment table contains links to other databases that are identified in a key above the table on the BLAST results page. In this example, there are no links to other databases.

View the alignments either by scrolling down the page or by clicking the link in the Max score column or a subject sequence in the alignment graph. The sequence alignments are organized by subject sequence, with all the regions that match one subject sequence grouped together. The sets of alignments are presented in order of maximum score, with the set containing the longest and best alignment shown first.

Record the top three species that align with your sequence and record each max score, query coverage, E value, and max identity.

Species	Max Score	Query Coverage	E Value	Max Identity

The take-home message from the example search is that the best match found was the same as that found in BOLD, but it should be noted that this search covered only 70% of the submitted sequence and of that 70%, the maximum identity was only 76%. So, for a sequence containing 709 base pairs (bp), only 496 bp were used to make a match and of those, only 381 did actually match the best case match. Based on this match, would you feel great confidence in concluding that the fish was the specific species of grouper known as the rooster hind? How confident would you be in concluding this was a member of the grouper genera?

Step 3. Assembling COI barcode contigs from trace files

Important Note: Steps 3–4 must be completed in their entirety.

Although low quality or ambiguous base calls are normally found at the 5' and 3' ends of a trace file, they may also appear elsewhere in the sequence. If only a single trace file was generated for a given COI amplicon, it would be difficult or impossible to confidently determine the identity of a base call that is assigned a low quality score value (that is, a value <20). However, a second trace file contains duplicate data that can help determine its identity with a greater level of statistical certainty.

Bringing two trace files into register and displaying them in the same window enables a researcher to identify regions of agreement or disagreement in base calls. In cases where a low quality base call is found in one trace file, the researcher can find the position of the base call in the other trace file and compare the differences in quality scores. If a higher quality score value (>20) is assigned to the base call in the second trace file, then that base call is regarded as the correct nucleotide and accepted.

The algorithm that operates within the BOLD-SDP Sequence Editor (and the Trace File Viewer described above) automatically reverses the sequence of base calls and peaks in the reverse trace file (which corresponds to the sequence of the antisense strand) so that they read in the opposite direction. It then converts each base call to its complementary nucleotide and recolors the corresponding peaks accordingly. This conversion therefore displays the reverse complement of the sequence read on the reverse trace file. For example, a base call of T that appears at the first position of the trace file above a red peak is replaced with a base call of A and moved to the last position above a green peak of the same shape and height as the original red peak. The quality value assigned to the original base call is also shifted to the last position. Next, the program aligns this sequence of complementary base calls and appropriately recolored peaks with the unaltered sequence of base calls and peaks of the forward trace file (which corresponds to the sequence of the sense strand). The largely overlapping DNA sequences are displayed in a project window of the BOLD-SDP Sequence Editor, as shown below.

The forward trace file appears in the top pane of the Online Sequence Editor window along with its sequence of base calls. The reverse complement of the reverse trace file appears in the lower pane along with its corresponding base calls. For both trace files, quality scores are represented graphically in the form of a histogram, where higher bars indicate higher quality scores and vice versa. The vertical scale on the right side of each trace file histogram displays the numerical quality values.

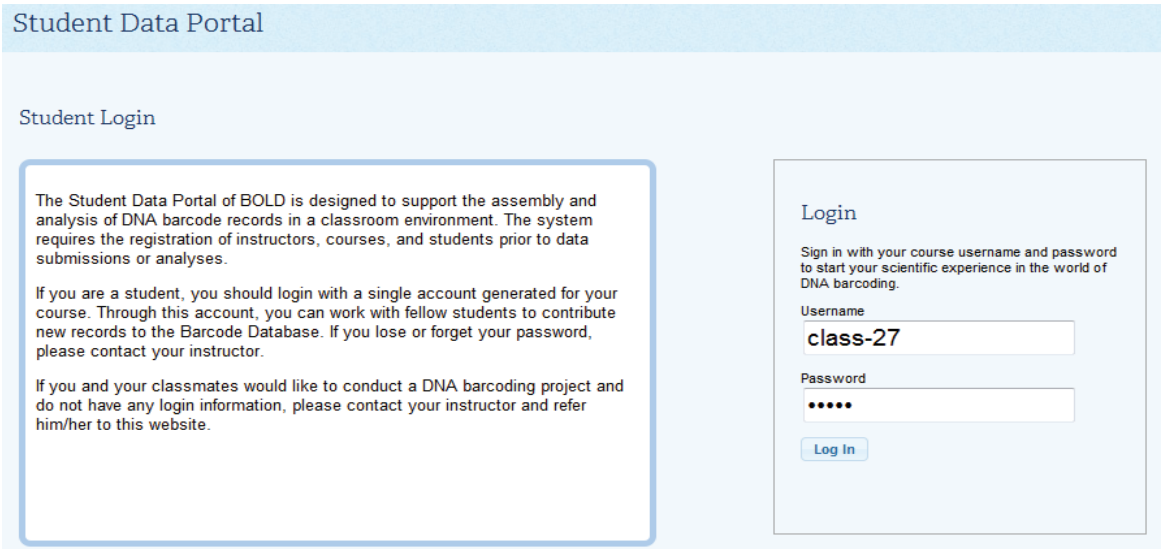
The nucleotide sequence at the bottom of the project window represents a contig – a continuous nucleotide sequence assembled from two overlapping DNA sequences (in this case, from the forward trace file and the reverse complement of the reverse trace file). The BOLD-SDP Sequence Editor compares the quality scores of base calls at every position of the trace files and accepts the base call with the higher quality score for inclusion in the contig. The bars that appear above each nucleotide in the contig are graphical representations of quality values, which are color-coded according to the legend in the upper right-hand corner of the window.

It is important to realize that the algorithm utilized by BOLD-SDP to make these comparisons is not perfect, so you must scan through the contig to ensure that no errors were made. This task is simplified by examining the quality scores that appear over each nucleotide in the contig. These scores represent the algorithm's confidence that the correct base call was chosen for inclusion in the contig. Low quality scores flagged with orange or red bars require human inspection.

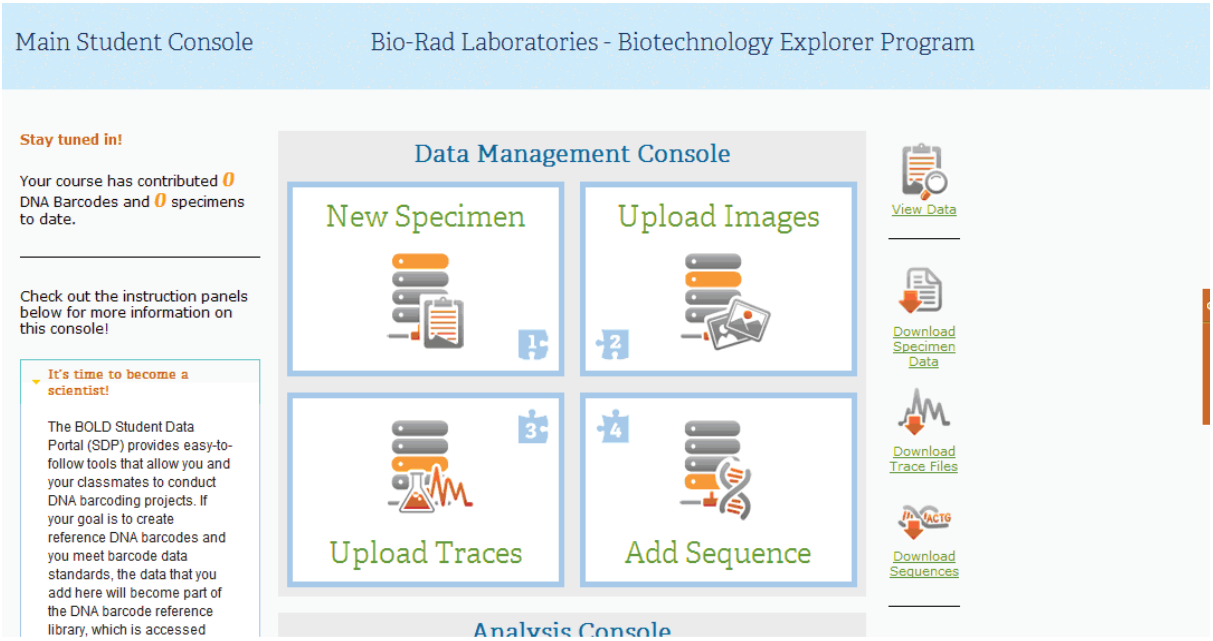
To assemble contigs for your forward and reverse trace files, please follow the steps outlined below:

a. Go to the Main Student Console page of BOLD-SDP by going to: **http://www.boldsystems.org/index.php/SDP_Home** and clicking the **Students** button to access the Student Login page.

b. Enter the course username and password in the appropriate spaces and click **Log In** to enter the Main Student Console page. Please note that the password is case sensitive!



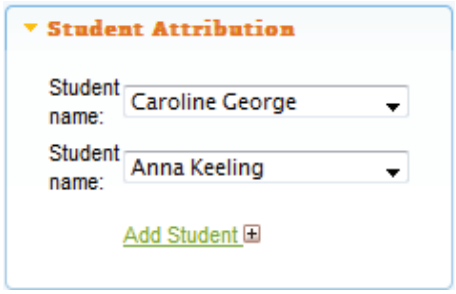
c. You should now see the Data Management Console.



d. On the Main Student Console page, click the **Add Sequence** icon.

e. The Add Sequence Page will open. For each student group, you will need to enter multiple pieces of information.

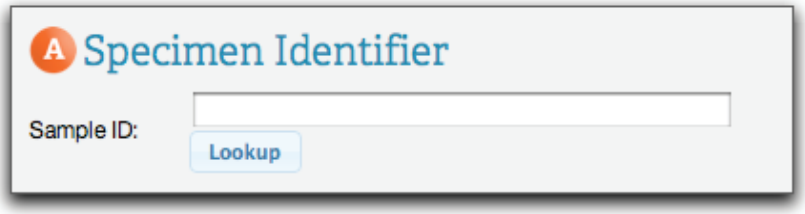
First, using the **Student Attribution box**, choose the student(s) who worked on this fish sample. If more than one student worked on this sample, click the **Add Student** button and select his/her name from the dropdown menu. Repeat the **Add Student** until all students who worked on this sample have been added.



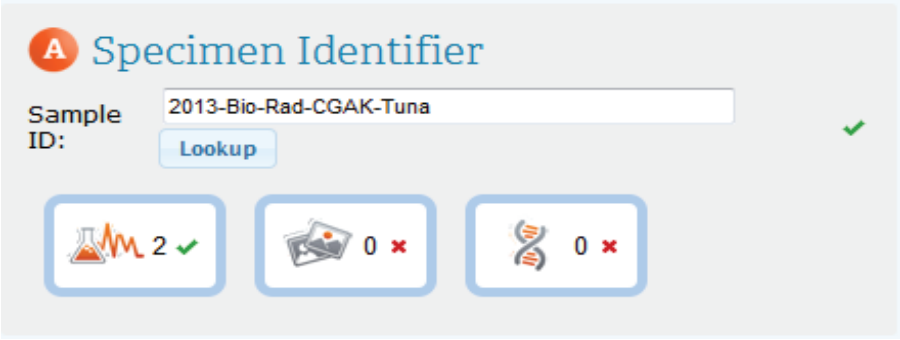
f. In Section A of the Add Sequence page, type or copy and paste a sample ID that corresponds to the trace files that you wish to assemble and edit and then press the **Tab** key. If you forgot the sample ID, click **Lookup** to find its ID in your class record list. Record the sample IDs for your two fish samples below.

Sample ID for Fish 1 :_____

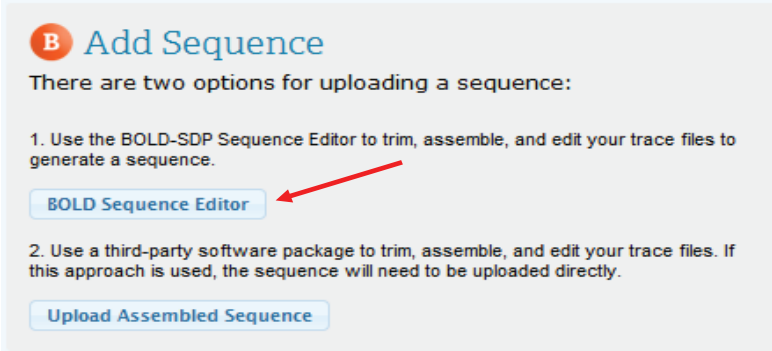
Sample ID for Fish 2: _____



If you have typed in the sample ID correctly, you should see icons such as the ones below appear once the system is ready to process your sequence trace files. There should also be a green check mark next to your sample ID entry. In the example below, the first icon means that there are two sequencing trace files for this sample ID. The second icon means that there are no photographs entered for this particular sample ID. The third icon means that there are no contig sequences generated for these two trace files at this point.

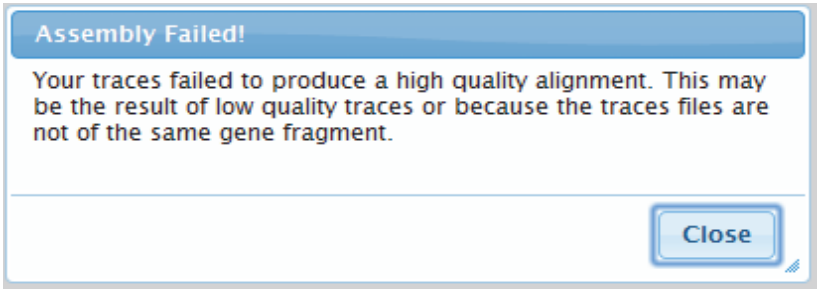


g. In Section B of the Add Sequence page, click the **BOLD Sequence Editor** button to load the trace files associated with the specimen/sample into the BOLD-SDP Sequence Editor.



If the BOLD-SDP software can align your sequences to produce a contig, a new window, the Online Sequence Editor, will appear.

If a contig cannot be generated, then you will get the following error message:

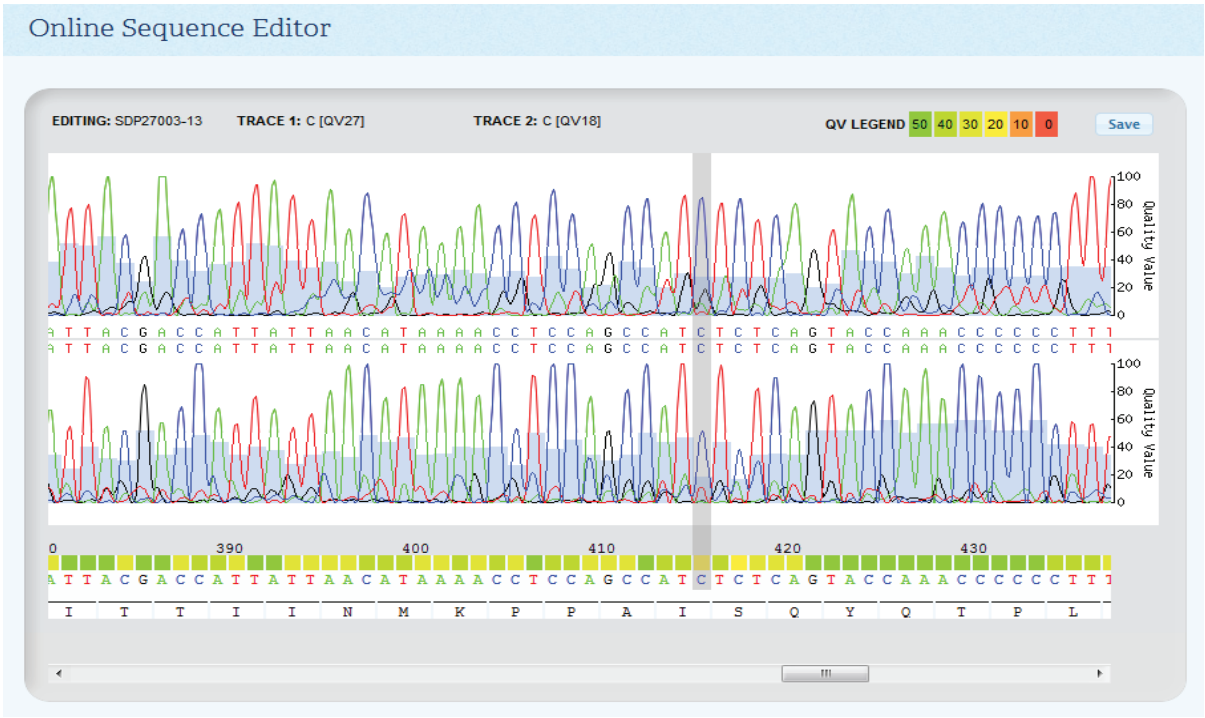


There are several possible reasons why a contig cannot be generated. One is that the uploaded sequence trace files did not come from the same sample. Make sure that your two sequencing traces are in the correct specimen ID folder. Another reason is lower quality data. Assembly programs are algorithms that have defined alignment parameters, such as how many base pair differences between the two sequences can be allowed before alignment is deemed a failure, or how many base pairs in a row must align in order to be confident that the contig sequence represents the best data. If a contig cannot be formed, you can speak with your instructor about possibly working on classmates' data along with them.

The BOLD-SDP Sequence Editor simplifies the editing process by automatically eliminating continuous stretches of low quality base calls from the contig. This process is known as trimming. It is important to realize that although these base calls are not included in the contig, they are still displayed in the forward and reverse trace files and colored gray.

The scroll bar at the bottom of the assembly project window allows you to examine the trace files and contig along their entire lengths (moving from 5' to 3'). In the BOLD-SDP Sequence Editor, start by scanning the entire length of the assembly to identify low quality bases, which are flagged with orange or red bars above the consensus sequence. Moving the mouse pointer over a base call in the trace files or consensus sequence will highlight the alignment position and display the base calls and associated quality scores/values at the top of the editor. Clicking a base will expose the editing tool, which enables a base call to be revised or made ambiguous. Do this by selecting one of the six options in the dropdown menu (**A**, **T**, **C**, **G**, **N**, or **-**).

Note: You cannot delete or insert a base call using this software, only change a base call to an N (ambiguous) or – (which serves the same purpose). This is because the program screens base calls and the codons they represent against known COI sequences.



h. The first step in the editing process is to carefully inspect the color of the bars over each nucleotide in the contig (the consensus sequence the BOLD-SDP program generated by determining the best sequence data from the forward and reverse sequencing reactions), starting from the 5' end (left side).

Contig sequence with colored bars above nucleotide calls

i. The bars are graphical representations of quality values, which are color-coded according to the legend in the upper right-hand corner of the window. It is important to watch for quality scores <20 (which are indicated by orange and red bars). If you discover an orange or red bar in the contig, **highlight the nucleotide** beneath it with your mouse. Notice that the corresponding base call in each trace file is also highlighted. Next, carefully inspect the quality scores for the corresponding base calls in both trace files. If the quality score for the base call is >20 in at least one trace file, the nucleotide in the contig can be regarded as correct.

   It can be highly subjective deciding the base call for bases at the 5' and 3' ends of the sequence where there might not be overlap between the two sequencing files. It can also be highly subjective to make base calls where both sequencing reactions have yielded low quality sequences. If evidence points to the base call being wrong, click on the **Edit Base** box dropdown menu and choose the base call you feel is more appropriate.

Note: Changes you make will apply only to the Contig sequence. You cannot change the raw data in your original trace files. Should you make a change that you later do not feel confident about, you can perform the Sequence function again and the sequence trace files that appear will be your original raw data; you will need to redo all base calls you want to make for the final contig.

Step 4. Inspecting contigs for the presence of stop codons

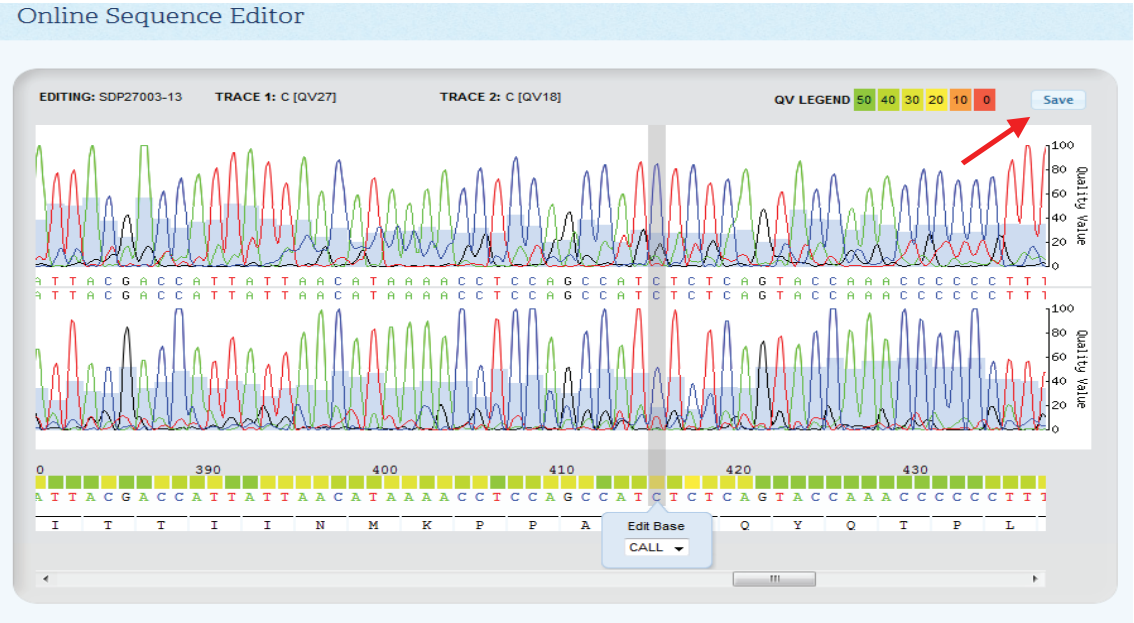
COI is a mitochondrial gene that directs the production of a protein subunit vital for cellular respiration. All mitochondrial protein-coding genes terminate in a stop codon — a triplet nucleotide that may take one of several forms depending on the taxon. During the process of transcription, the stop codon of a protein-coding gene is transcribed into messenger RNA. At the conclusion of translation, the stop codon binds a release factor, which signals the ribosome to dissociate and release the newly synthesized amino acid chain.

The 650 bp region of the *COI* gene that you amplified by PCR is located upstream of the stop codon found in the mitochondrial DNA template. Accordingly, stop codons should be absent in your edited contig. The presence of a stop codon indicates one of three likely possibilities: 1) a nucleotide was erroneously omitted in the contig, 2) an extra nucleotide was erroneously included, or 3) a base call was incorrectly made.

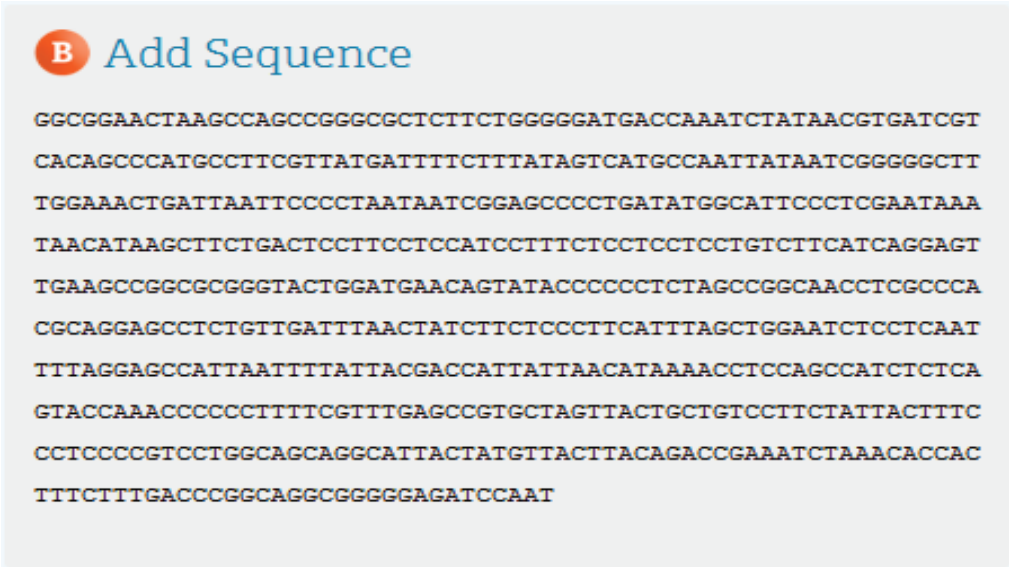
The BOLD-SDP Sequence Editor enables you to examine your sequence for the presence or absence of stop codons. Because the COI barcode region that you amplified is also downstream of the start (ATG) codon found in the mitochondrial DNA template gene, the Auto Translator algorithm built into the BOLD-SDP Sequence Editor must first organize your contig into three reading frames. For reading frame 1, nucleotides are grouped into codons beginning with the first nucleotide in the contig. For reading frame 2, nucleotides are grouped into codons beginning with the second nucleotide in the contig (the first nucleotide is ignored). For reading frame 3, nucleotides are grouped into codons beginning with the third nucleotide in the contig (the first and second nucleotides are ignored). The translator then uses a translation matrix similar to a genetic code table used in classroom settings to determine the amino acid sequence of each reading frame. It then compares the three amino acid sequences to a database of known COI amino acid sequences to determine which reading frame is correct. The correct amino acid sequence is displayed at the bottom of the sequence editor project window.

a. Examine your contig sequence for stop codons represented by an *. Look carefully at the three base calls that translate to a stop codon. If any of them are ambiguous, you can change them to the correct base call. It is possible to change all three base calls for the stop codon to an N and this should change the translation from a stop codon to an X. The X is not a stop codon, but it also does not stand for a specific amino acid. It just represents an unknown call.

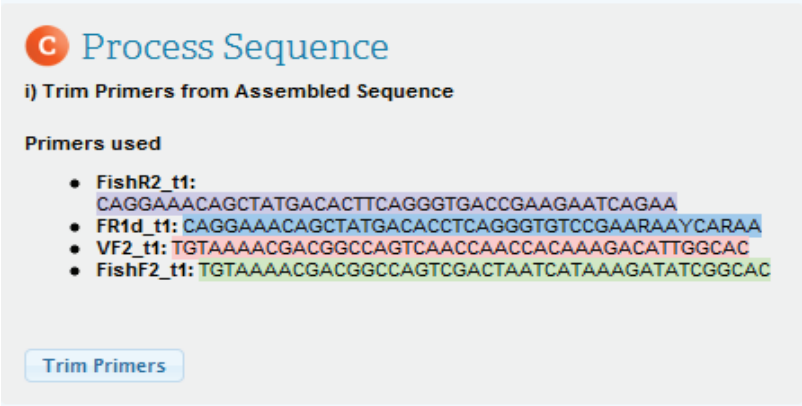
b. If no stop codons were detected in the amino acid translation of the contig, then click the **Save** button. A window will appear that lets you know that your edited sequence has been saved and you should proceed with the sequence uploader to submit your sequence to BOLD. Click **OK**. BOLD-SDP will take you back to the Add Sequence work page.



Your contig sequence should now be shown in Box B, Add Sequence.

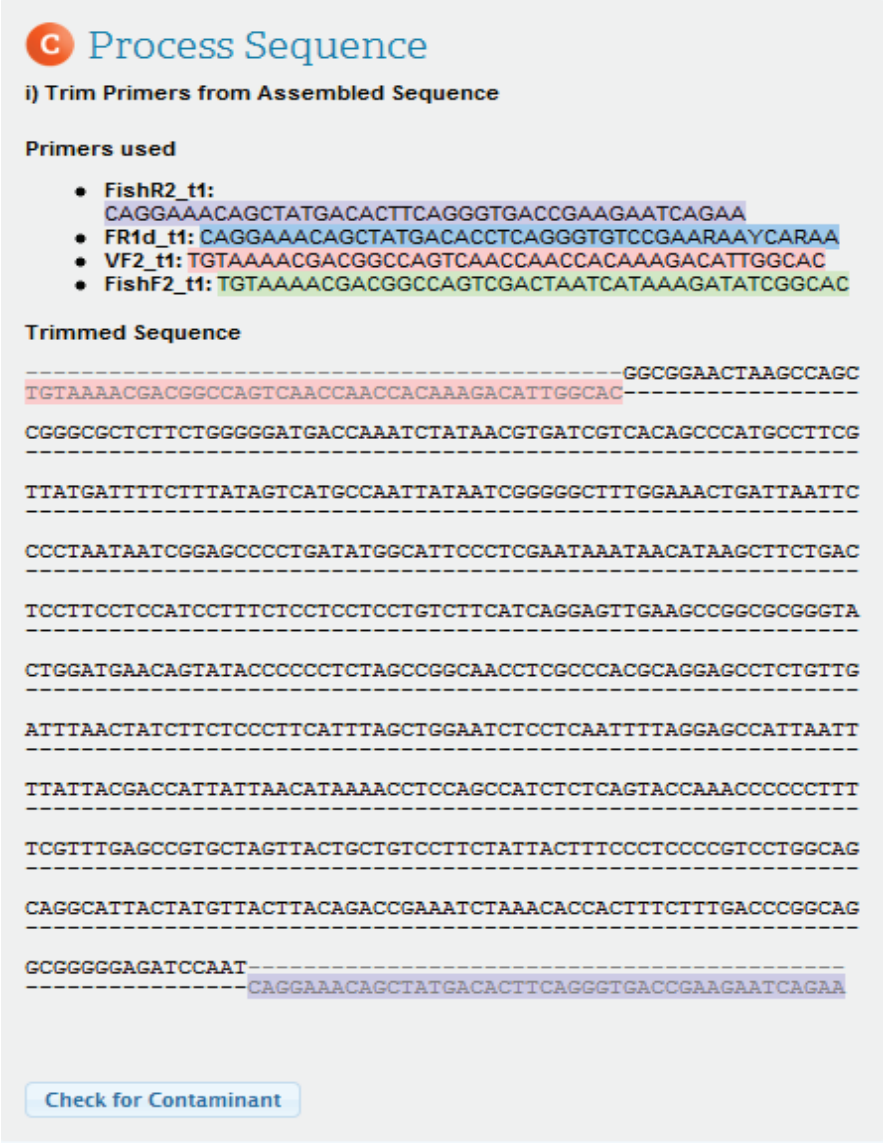


c. Next, you will trim off the sequences at the 5' and 3' ends of your contig that correspond to the PCR primers. If the PCR primers were entered correctly when your sequencing trace files were uploaded, you should see a list of four primer sequences: FishR2_t1, FR1d_t1, VF2_t1, and FishF2_t1. These were the four primers that were mixed with PCR master mix for your PCR reaction. Automatically trim the primer sequences from your edited sequence by clicking the **Trim Primers** button in Section C, Process Sequence.

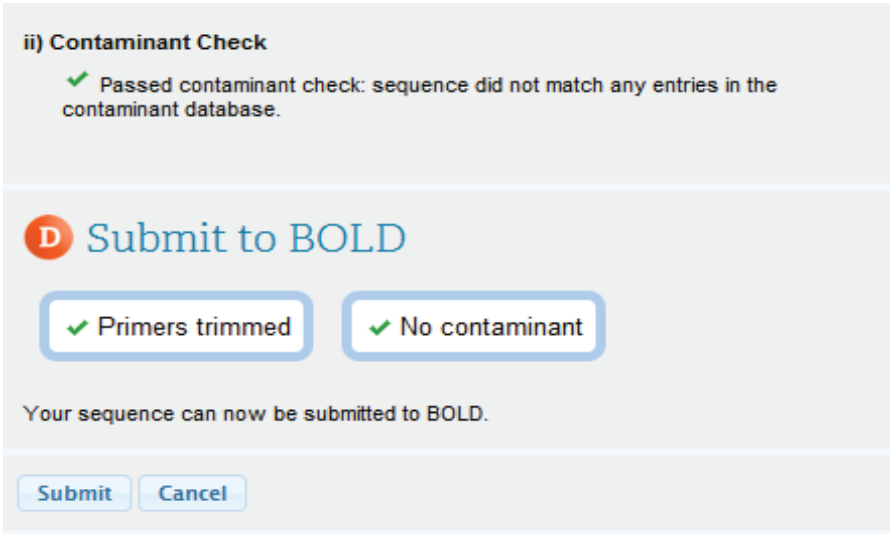


d. Once BOLD-SDP performs the trimming function, you can see which of the primers contained in the mixed primers you used for PCR actually matched best for your fish sample.

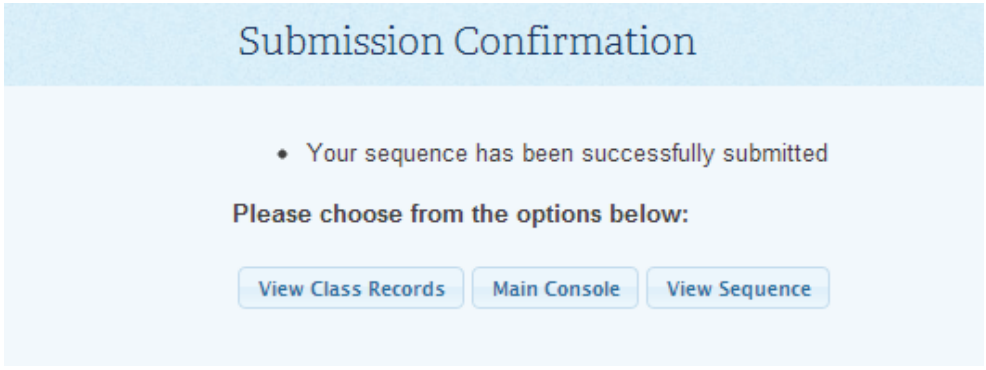
e. Now it is time to check whether any contaminants were present in your sample. BOLD has a list of standard contaminant sequences from bacteria that may have been PCR amplified instead of your fish DNA. This has been a common issue in barcoding samples and in fact, GenBank has several sequences attributed to fish that are actually marine bacteria. Click the **Check for Contaminant** button to inspect your sequence for the presence of common lab contaminants, including human contaminants.



f. Once the contaminant check is passed click the **Submit** button to link the edited and validated barcode sequence to your specimen/sample.



Once the sequence has successfully been uploaded into BOLD, you will receive a confirmation page.

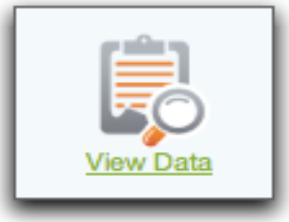


Step 5. Verify that edited COI sequence was incorporated in sample record

The final step in the editing process is to verify that the edited sequence was integrated into the barcode record of the appropriate sample. To perform this function:

a. Navigate to the Main Student Console page of BOLD-SDP.

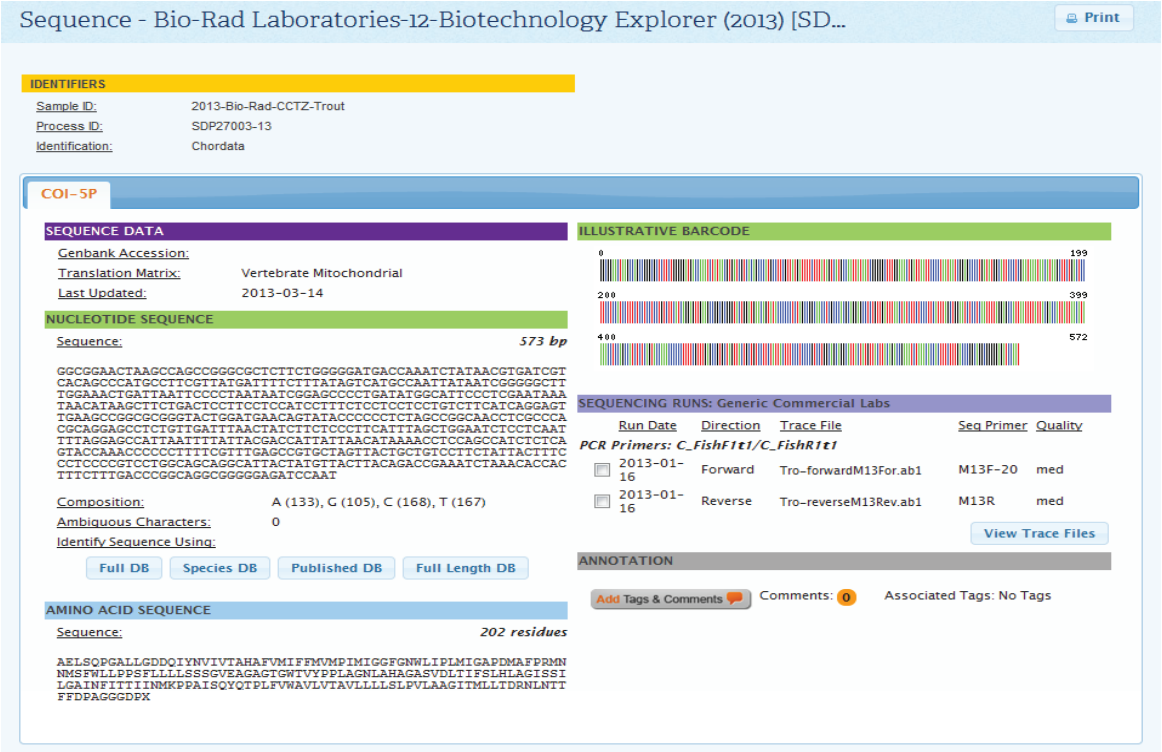
b. In the right sidebar of the Main Student Console page, click the **View Data** icon.



c. On the Record List page, locate the row for the sample that you linked to the recently uploaded sequence.

				Length [Ambig]		
	Identification	Sample ID	Process ID	COI-5P	Record Flags	Extra Info
	Chordata	2013-Bio-Rad-ACLSTW-grouper	SDP27007-13	0		
	Chordata	2013-Bio-Rad-ACLSTW-tuna	SDP27006-13	0		
	Chordata	2013-Bio-Rad-ACLSTW-coho	SDP27005-13	0		
	Chordata	2013-Bio-Rad-CCTZ-CookedSalmon	SDP27004-13	0		
	Chordata	2013-Bio-Rad-CCTZ-Trout	SDP27003-13	573 [0n]		
	Chordata	2013-Bio-Rad-CGAK-Tuna	SDP27002-13	0		
	Chordata	2013-Bio-Rad-CGAK-Salmon	SDP27001-13	0		

d. Click the **Process ID** link for the sample to open its sequence page in a new window.



The edited COI nucleotide sequence can be found in the Nucleotide Sequence pane along with associated data, including sequence length (in base pairs), sequence composition (that is, the number of A's, C's, T's, and G's in the sequence), and the number of ambiguous characters or nucleotides (N's).

NUCLEOTIDE SEQUENCE

Sequence: *529 bp*

```
AGAACTGAGCCAGCCGGGCGCTCTTCTAGGGGATGATCAGATTTACAACGTAATCGTCAC
AGCCCATGCCTTCGTTATGATTTTCTTTATAGTCATGCCGATTATGATCGGAGGCTTTGG
AAACTGATTAATTCCCCTAATGATCGGAGCCCCTGATATGGCATTCCCTCGAATAAATAA
CATAAGCTTCTGACTCCTTCCGCCATCCTTTCTCCTCCTCCTATCTTCCTCTGGAGTTGA
AGCCGGGGCTGGCACCGGGTGAACAGTTTATCCCCCTCTGGCCGGCAACCTCGCCCACGC
AGGAGCCTCAGTTGATCTGACTATCTTCTCCCTTCATTTAGCCGGGATCTCCTCAATTTT
AGGAGCCATTAATTTTATTACGACCATTATTAACATAAAGCCCCCAGCTATCTCTCAGTA
CCAAACCCCACTTTTTGTTTGAGCTGTGCTAGTCACTGCTGTTCTTCTACTACTCTCTCT
CCCCGTTCTGGCAGCAGGCATTACTATGTTACTTACAGACCGAAATCTA
```

Composition:	A (117), G (98), C (154), T (160)
Ambiguous Characters:	0

The amino acid translation and total number of amino acid residues encoded by your COI nucleotide sequence are located in the lower left pane of the Sequence page in the Amino Acid Sequence pane.

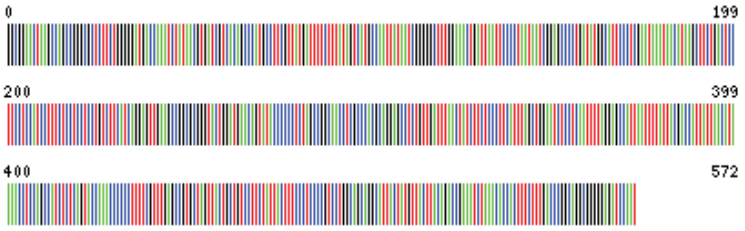
AMINO ACID SEQUENCE

Sequence: *202 residues*

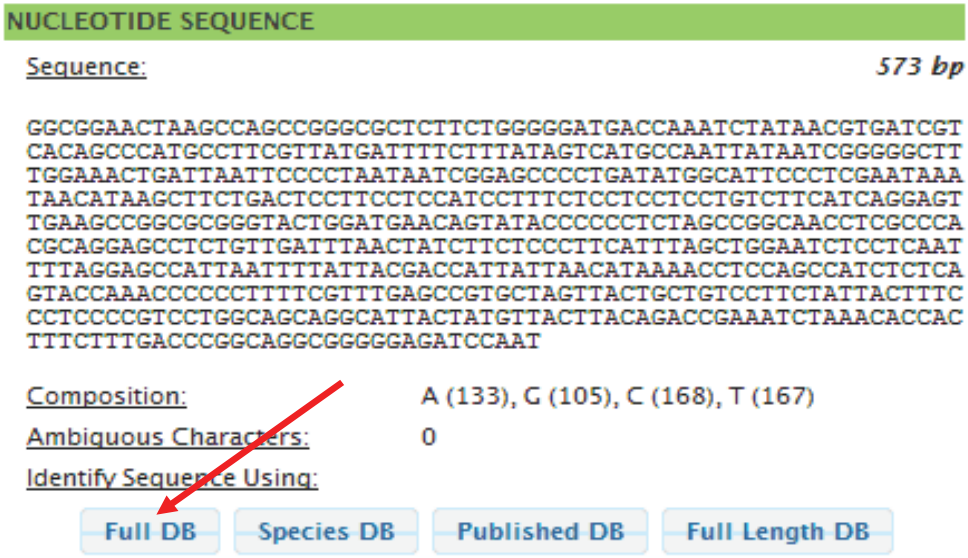
```
AELSQPGALLGDDQIYNVIVTAHAFVMIFFMVMPIMIGGFGNWLIPLMIGAPDMAFPRMN
NMSFWLLPPSFLLLLSSSGVEAGAGTGWTVYPPLAGNLAHAGASVDLTIFSLHLAGISSI
LGAINFITTIINMKPPAISQYQTPLFVWAVLVTAVLLLLSLPVLAAGITMLLTDRNLNTT
FFDPAGGGDPX
```

The illustrative barcode in the upper right-hand corner of the Sequence page represents each nucleotide in your barcode sequence as a different colored line. A is represented with green lines, T with red lines, C with blue lines, and G with black lines.

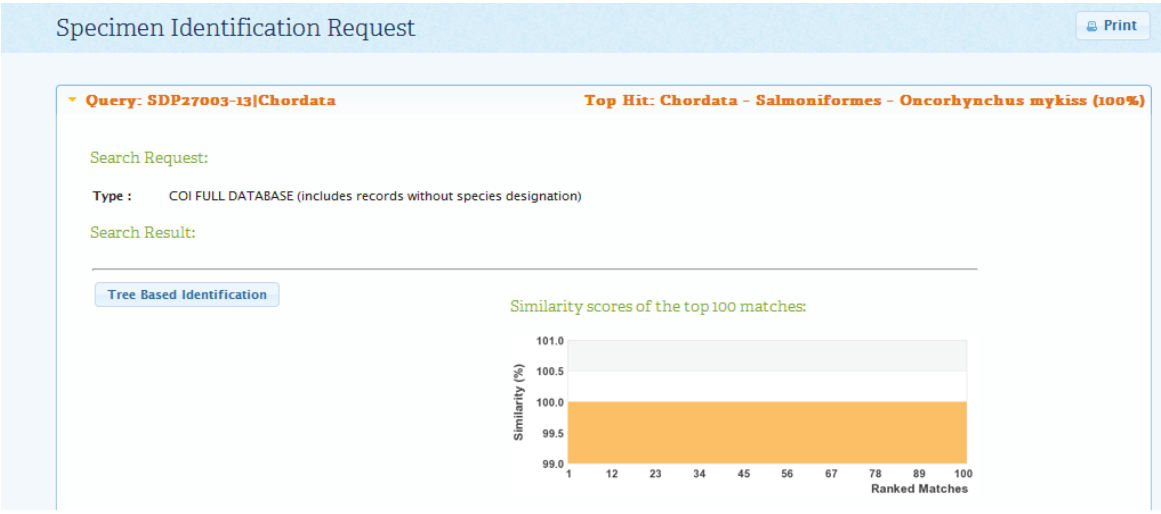
ILLUSTRATIVE BARCODE



e. To compare the barcode sequence in your record with other barcode sequences in the BOLD species database, click the **Full DB** button at the bottom of the Nucleotide Sequence pane.



A Specimen Identification Request window will open that contains different forms of information. The Search Result pane near the top of the page contains a summary statement of the search performed by the BOLD Identification System (BOLD-IDS), which is supported by the data displayed in other sections of the page.



The list of the top matches follows in a table.

TOP 99 Matches:

Phylum	Class	Order	Family	Genus	Species	Similarity (%)	Status
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Early-Release
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published

If a match is found, record the phylum, class, order, family, genus, species (or any level of taxonomical match that was found), and % similarity for the top match to your sequence.

Sequencing trace file name: ______________________________

Phylum: ______________________________

Class: ______________________________

Order: ______________________________

Family: ______________________________

Genus: ______________________________

Species: ______________________________

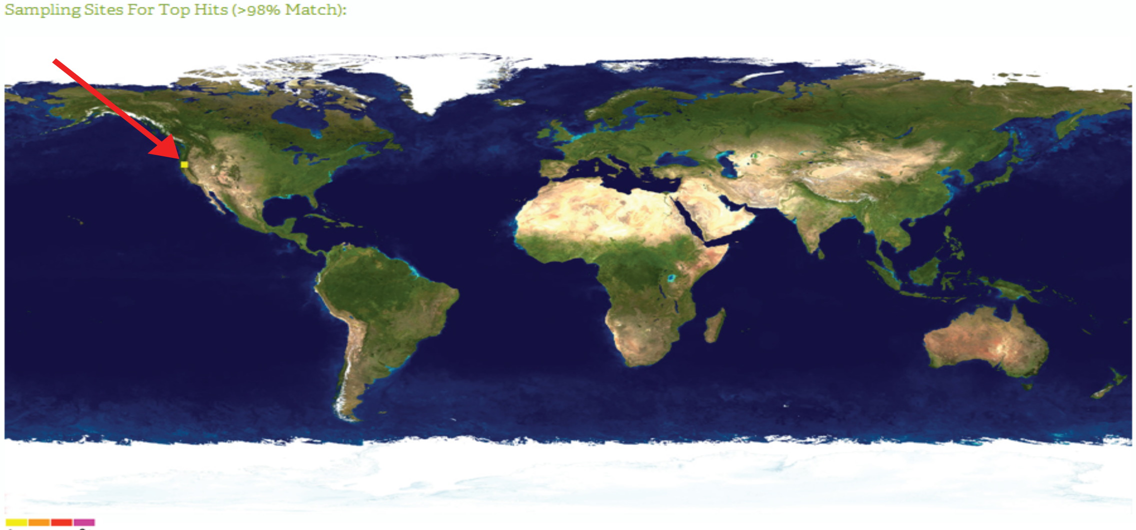
% Similarity: ______________________________

How does this result compare to the results for your single sequence matches done previously? Would you expect the match to be better using your contig sequence or using single sequences? Why?

If you have <100% similarity with species in the database, what does this mean? What sorts of differences would you expect within species (in other words, between different individuals of the same species), between two different species of fish, between two different genera of fish, between two different families of fish?

The world map at the bottom of the page shows the collection site of specimens with COI sequences that are >98% similar to the COI barcode sequence of your specimen/sample.

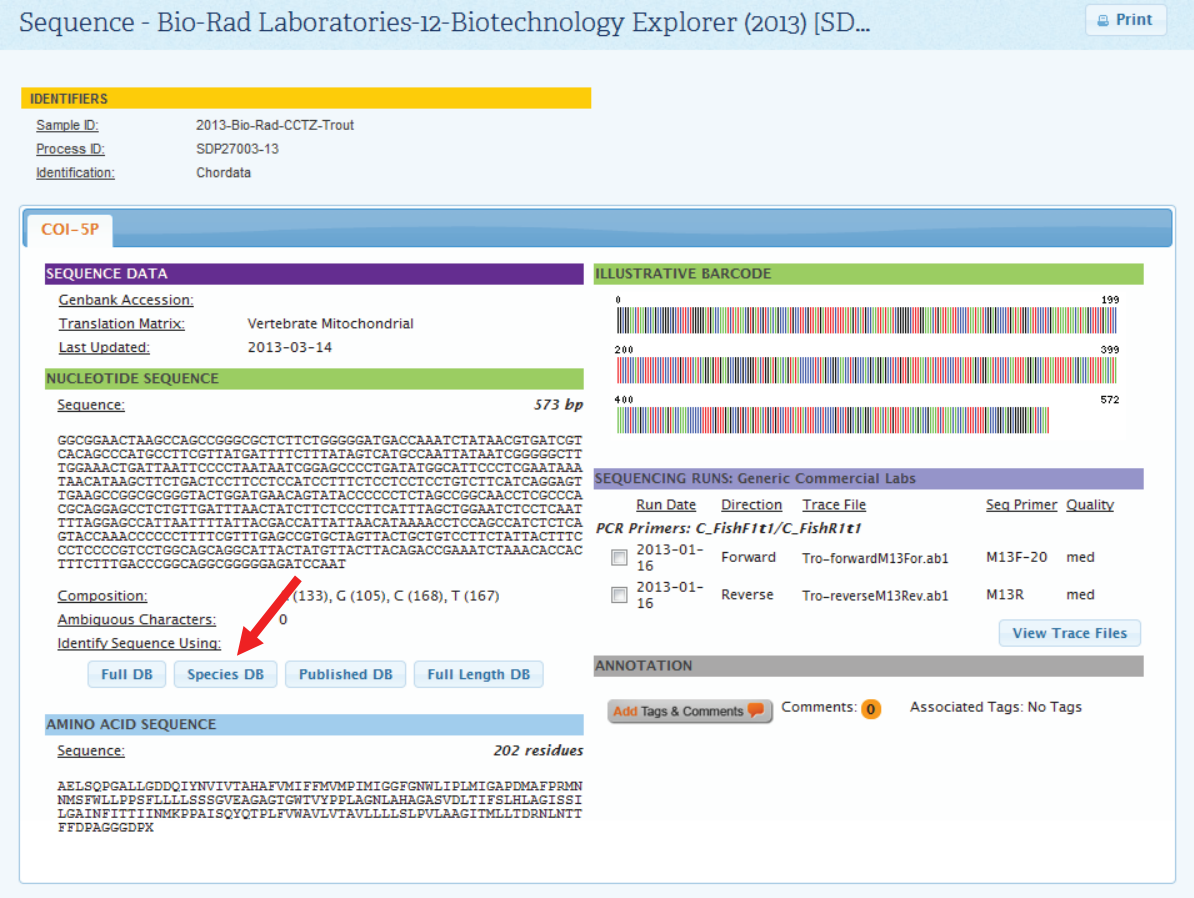
Sampling Sites For Top Hits (>98% Match):



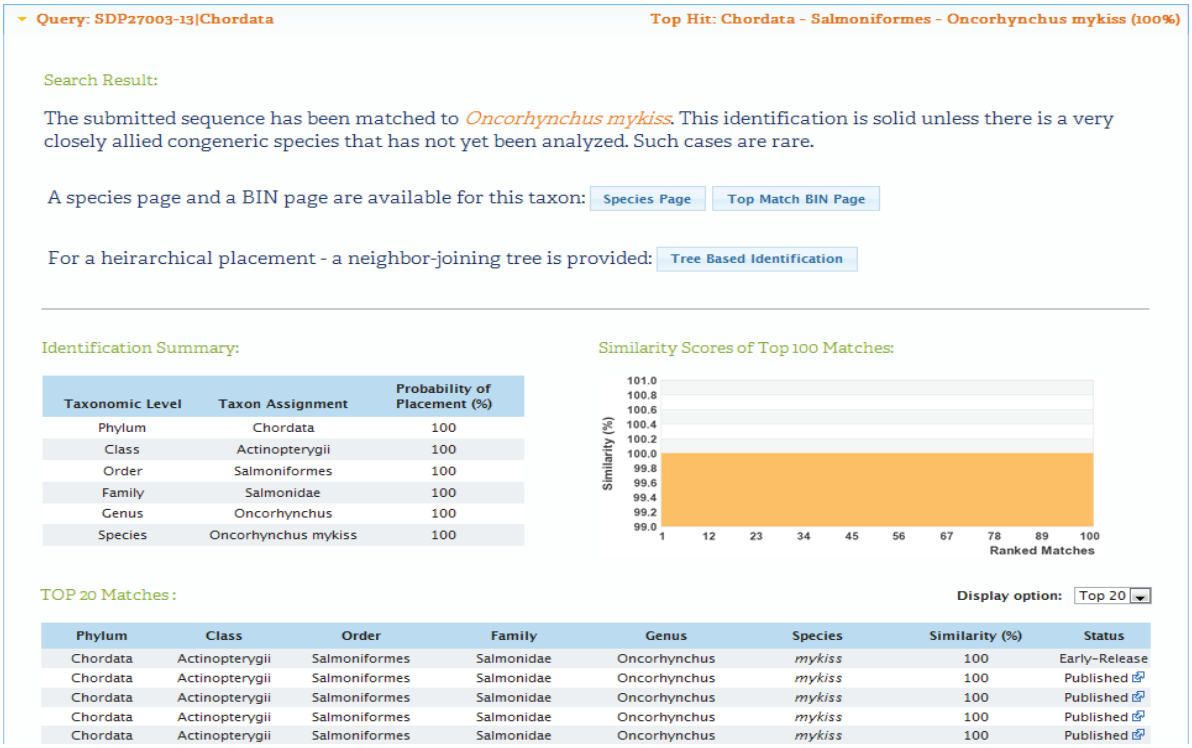
Did the identification match what you thought your species was? If it did not, what was the % similarity?

f. In addition to comparing your sample to samples in the full database, you can also compare your sample to the species database. The full database has all vouchered samples that have been submitted, but the species database is even more stringent and contains only samples for which both genus and species were confirmed. The full database (All Barcode Records database) is the unvalidated library, while the records in the species database (Species Level Barcode database) are all completely validated. So in terms of levels of stringency for submission, GenBank would be the least stringent, then the All Barcode Records database on BOLD, and finally the Species Level Barcode Records database on BOLD.

   From your Sequence page, click the **Species DB** button.



g. This page will return even more specific data if a match can be found. When searching the species database, the % match at the phylum, class, order, family, genus, and species levels will all be listed.



Query: SDP27003-13|Chordata — Top Hit: Chordata - Salmoniformes - Oncorhynchus mykiss (100%)

Search Result:

The submitted sequence has been matched to *Oncorhynchus mykiss*. This identification is solid unless there is a very closely allied congeneric species that has not yet been analyzed. Such cases are rare.

A species page and a BIN page are available for this taxon: Species Page | Top Match BIN Page

For a heirarchical placement - a neighbor-joining tree is provided: Tree Based Identification

Identification Summary:

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
Phylum	Chordata	100
Class	Actinopterygii	100
Order	Salmoniformes	100
Family	Salmonidae	100
Genus	Oncorhynchus	100
Species	Oncorhynchus mykiss	100

Similarity Scores of Top 100 Matches:

TOP 20 Matches: Display option: Top 20

Phylum	Class	Order	Family	Genus	Species	Similarity (%)	Status
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Early-Release
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published
Chordata	Actinopterygii	Salmoniformes	Salmonidae	Oncorhynchus	*mykiss*	100	Published

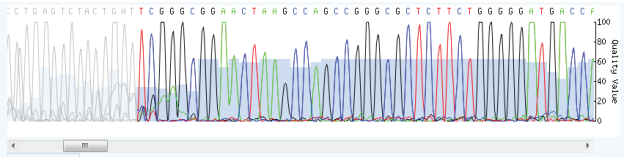
Compare your matches from single sequences using BOLD and/or GenBank versus using the All Barcode Records database and the Species Level Barcode Records database. Do you see any differences? Do they all point to the same fish genus and species? Did you find any strange matches from any of the databases? Why might those have occurred?

You have now completed the bioinformatics analysis of your fish samples. You isolated DNA, used PCR to amplify the *COI* gene, analyzed the PCR products using gel electrophoresis, and had your samples purified and sequenced. This is the same workflow performed by researchers participating in the International Barcode of Life project. The difference is the strict control and vouchering of the samples they use and the requirements for the sequencing data before it can be submitted. It is hard to know you have the correct sequence if you are not sure what you started with!

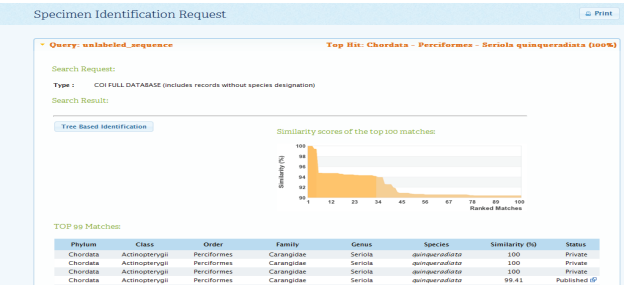
Did you find that the fish you tested was what you expected it to be? If not, do you have enough confidence in your data to determine whether there was a problem in the workflow or if there may have been market substitution? The more familiar you become with this process, the more confident you will be in your results. Maybe you did have a sample that was called red snapper but was really tilapia — you would not be the first. But hopefully with more barcoding being done on fish samples, you will be the last!

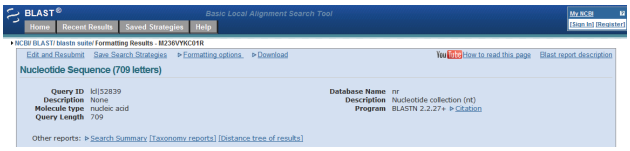
Student Workflow

Assess quality of DNA sequencing data (look at chromatograms and examine quality scores)



Query BOLD database and/or GenBank for matches to forward and reverse sequencing data

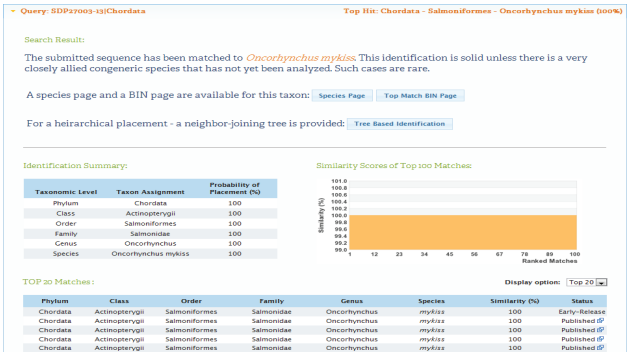




Assemble a single sequence (contig) from forward and reverse sequencing reaction data

Manually compare any base calls that have differences between the forward and reverse sequencing reaction data, check for stop codons, and save contig

Perform a search to determine the identity of your fish sample contig data



Quick Guide: DNA Barcoding Bioinformatics using BOLD-SDP

Student login

Course username: _____

Course password: _____

Sample ID for fish 1: _____

Sample ID for fish 2: _____

	Mean Quality Score	Standard Deviation
Fish 1: Forward sequence data	_____	_____
Fish 1: Reverse sequence data	_____	_____
Fish 2: Forward sequence data	_____	_____
Fish 2: Reverse sequence data	_____	_____

BOLD-SDP barcode record search against trace file data

Fish 1:

Forward sequencing trace file name: _____

Phylum: _____

Class: _____

Order: _____

Family: _____

Genus: _____

Species: _____

% Similarity: _____

Reverse sequencing trace file name: _____

Phylum: _____

Class: _____

Order: _____

Family: _____

Genus: _____

Species: _____

% Similarity: _____

Fish 2:

Forward sequencing trace file name: _____

Phylum: _____

Class: _____

Order: _____

Family: _____

Genus: _____

Species: _____

% Similarity: _____

Reverse sequencing trace file name: _____

Phylum: _____

Class: _____

Order: _____

Family: _____

Genus: _____

Species: _____

% Similarity: _____

Blastn database search against trace file data

Top 3 species that align with your sequences for fish 1:

Species	Max Score	Query Coverage	E Value	Max Identity

Top 3 species that align with your sequences for fish 2:

Species	Max Score	Query Coverage	E Value	Max Identity

BOLD-SDP database search against contig sequence

Database matches to assembled contig for fish 1:

Sample ID: _____

Phylum: _____

Class: _____

Order: _____

Family: _____

Genus: _____

Species: _____

% Similarity: _____

Database matches to assembled contig for fish 2:

Sample ID: _____

Phylum: _____

Class: _____

Order: _____

Family: _____

Genus: _____

Species: _____

% Similarity: _____

BIO-RAD

Bio-Rad Laboratories, Inc.

Life Science Group

Web site *www.bio-rad.com* ***USA*** *800 424 6723* ***Australia*** *61 2 9914 2800* ***Austria*** *01 877 89 01* ***Belgium*** *09 385 55 11* ***Brazil*** *55 11 5044 5699*
Canada *905 364 3435* ***China*** *86 21 6169 8500* ***Czech Republic*** *420 241 430 532* ***Denmark*** *44 52 10 00* ***Finland*** *09 804 22 00*
France *01 47 95 69 65* ***Germany*** *089 31 884 0* ***Greece*** *30 210 9532 220* ***Hong Kong*** *852 2789 3300* ***Hungary*** *36 1 459 6100* ***India*** *91 124 4029300*
Israel *03 963 6050* ***Italy*** *39 02 216091* ***Japan*** *03 6361 7000* ***Korea*** *82 2 3473 4460* ***Mexico*** *52 555 488 7670* ***The Netherlands*** *0318 540666*
New Zealand *64 9 415 2280* ***Norway*** *23 38 41 30* ***Poland*** *48 22 331 99 99* ***Portugal*** *351 21 472 7700* ***Russia*** *7 495 721 14 04*
Singapore *65 6415 3188* ***South Africa*** *27 861 246 723* ***Spain*** *34 91 590 5200* ***Sweden*** *08 555 12700* ***Switzerland*** *026 674 55 05*
Taiwan *886 2 2578 7189* ***Thailand*** *800 88 22 88* ***United Kingdom*** *020 8328 2000*

Bulletin 6398 Rev A US/EG *Sig 1212*