|  |  |
| --- | --- |
| **Discovery learning: DNA Barcoding & Biodiversity databases** |  |

## Objectives

Upon completion of this module, each student should be able to:

* Explain how databases are used to share biodiversity data broadly
* Extrapolate and analyze biodiversity data from various databases
* Incorporate collected data into a database
* Use two sequence databases (BLAST & BOLD) to identify your unknown sequence from the organism you collected
* Use the GBIF database to obtain more information on your organism
* Submit your identified organism sequence into the BOLD database through the student portal

**Introduction**

***Biodiversity*** refers to the variety of living species on earth. You can think of plants, animals, bacteria, and fungi but this also refers to other levels of diversity like the genetic differences among organisms, and the communities and ecosystems in which they occur. The biodiversity on earth is so rich that many species have yet to be discovered!

Unfortunately, many species are at risk of extinction due to human activities. Therefore it is critical for ongoing monitoring and assessment of biodiversity, so we can identify changes in species distribution patterns and community composition.

Measuring biodiversity comes in a variety forms, for example you can trap an organism and take measurements like length or you can take a picture or an organism and upload to an app like iNaturalist or eBird to document its occurrence. ***DNA barcoding*** is another method to measure biodiversity and is designed to use a single region of DNA to identify a species as simple as scanning a supermarket barcode. Regardless of how you are measuring biodiversity, you can imagine how quickly this data can add up! The effort to standardize and make this data accessible globally has led to ***Biodiversity databases***.

**Activities**

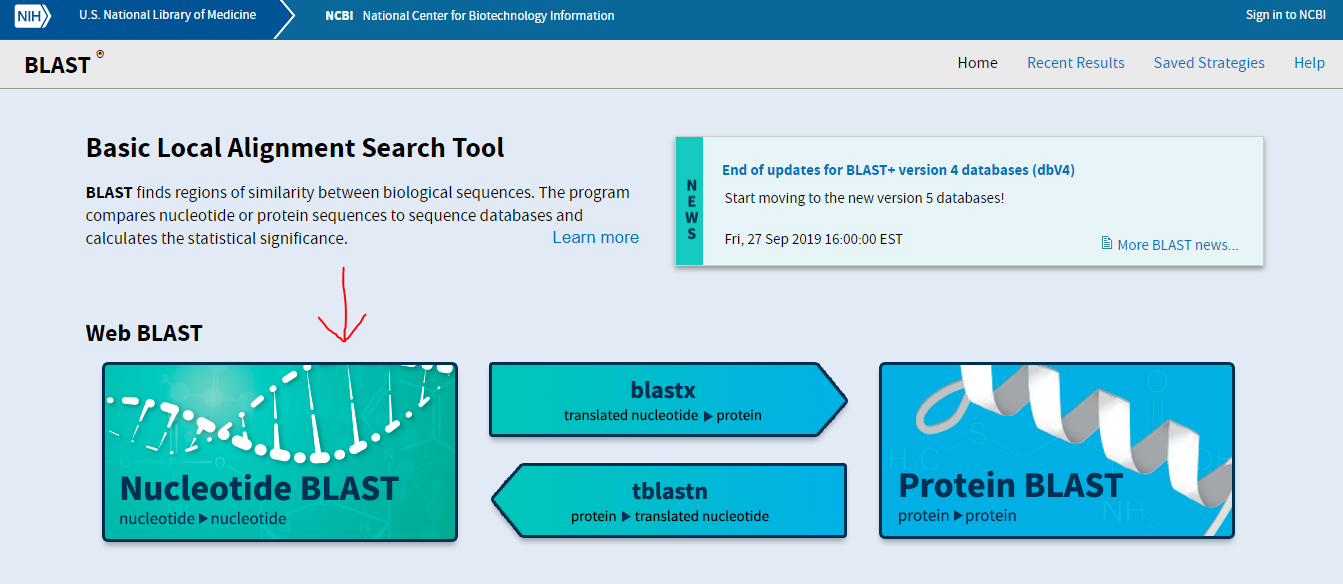
**BLAST Database**

You will use the BLAST database to identify your organism’s identity based on sequence information.

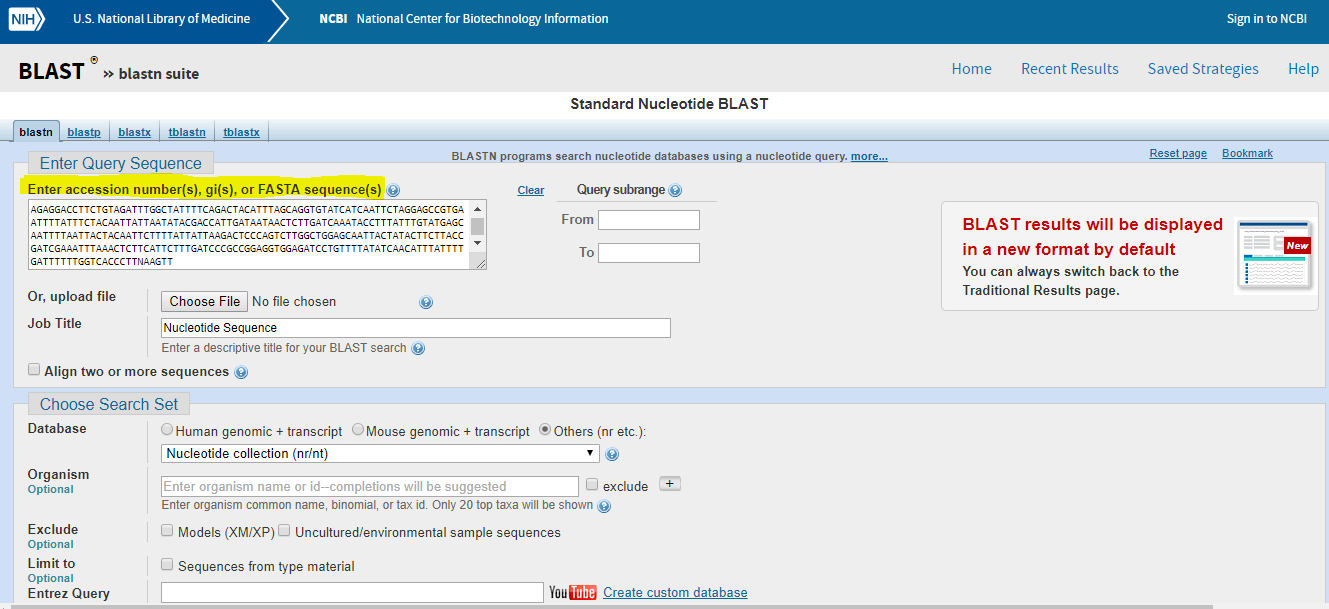
You have been assigned a sequence and you will use this worksheet to submit your answers:

(<https://docs.google.com/spreadsheets/d/1lJzQRG0GNdOTSCkGwUCSsXW2nKszPoJrjsQK5Bi9gog/edit?usp=sharing>)

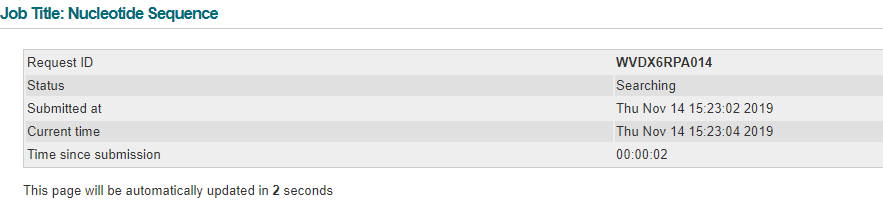
Navigate to the BLAST website (<https://blast.ncbi.nlm.nih.gov/Blast.cgi#alnHdr_320521487>). Click “Nucleotide BLAST” as pictured below.



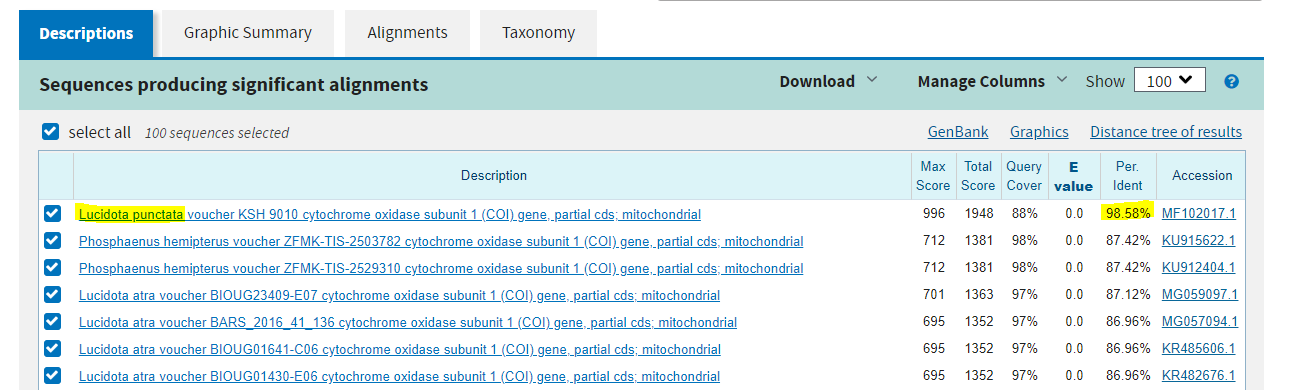
Next, where it says enter accession number(s), insert your DNA sequence.



Next, simply scroll down and click blast!https://lh6.googleusercontent.com/s-zIj52T-WELUez0PzjWCvyUR6cBt_vuXC1Jbn0v_ar8uWILEysA3BogiAOkSRvGKq5OarwKcKESEuvSmgHkB0jaNtEwHfckzGSTsehzBK3fiTrdN4PXNEdcQrDtJh_5qefBdAuE

You will see the function running below, and it will take a few moments to load.

After the Blast loads, scroll down to view significant alignments.

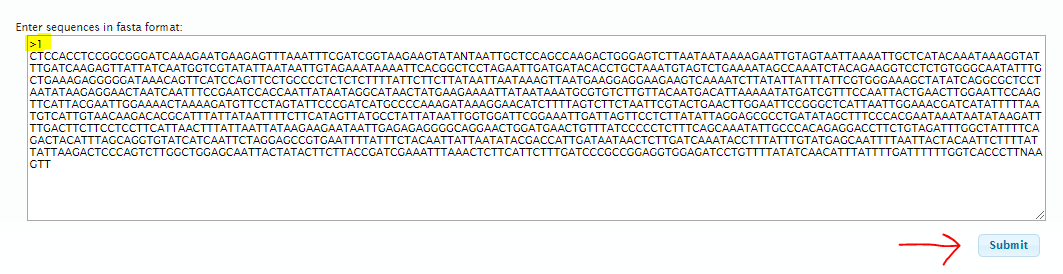


A Percent Identification of 99% or higher is preferable. **Record the suggested identification only if it has a percentage of 85 or above and add the Accession number.**

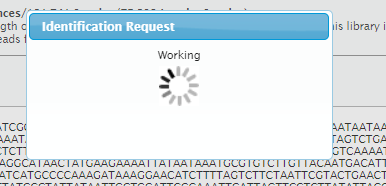
When you are finished exploring and identifying your sequence, scroll up to “Edit Search” to restart if needed.

**BOLD Database**

Navigate to the BOLD website (<http://v3.boldsystems.org/index.php/IDS_OpenIdEngine>). Simply copy and paste your sequence where it says “Enter sequences in fasta format.” All this means is to run, the program needs a greater than (>) and a number (e.g. 1) ahead of the sequence. This is simply a formatting preference. See the figure below for an example.

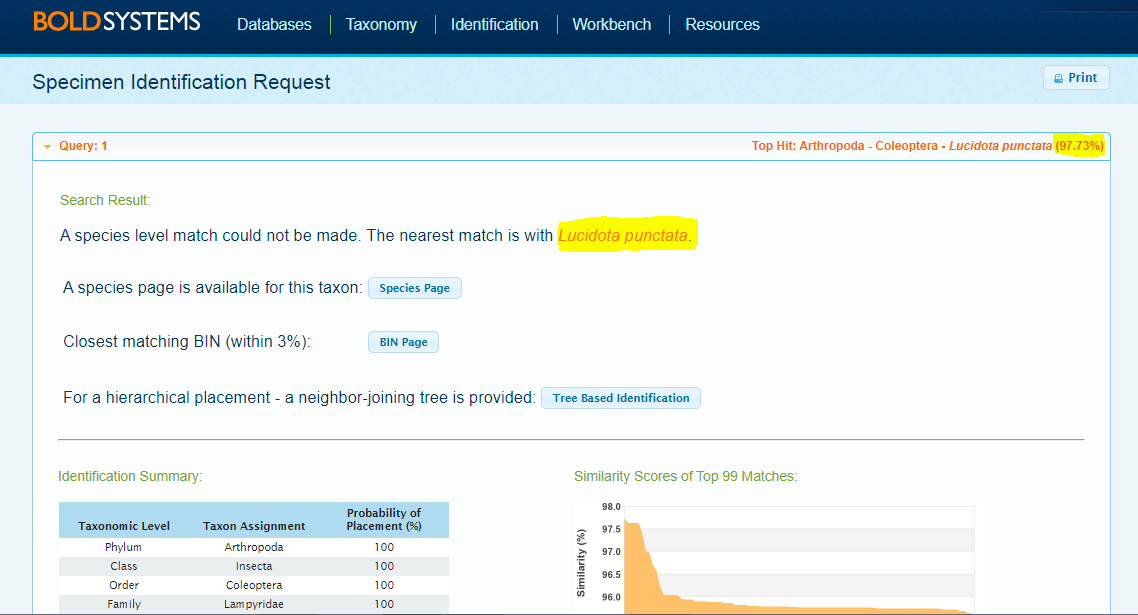


For simplicity, enter >1 followed by your sequence. Then, press submit to run.



Once BOLD searches its database, it will redirect you to a results page.

The results page will tell you if a species level match was made, the nearest match, or if no match was made at all. In the top right, the percent identification will be displayed. **Again, record the suggested identification only if it has a percentage of 85 or above.**



You should now have identified your organism to species or genus level! A tentative match for those who are 90% or below.

If needed, simply press the back arrow on your browser to restart.

**GBIF Database**

Now go to the GBIF database <https://www.gbif.org/> and click **Occurrences.** Next type the name of your organism into the search box and search



This will take you to a page with different tabs, explore each one and record information on the Second tab of Google spreadsheet.

****

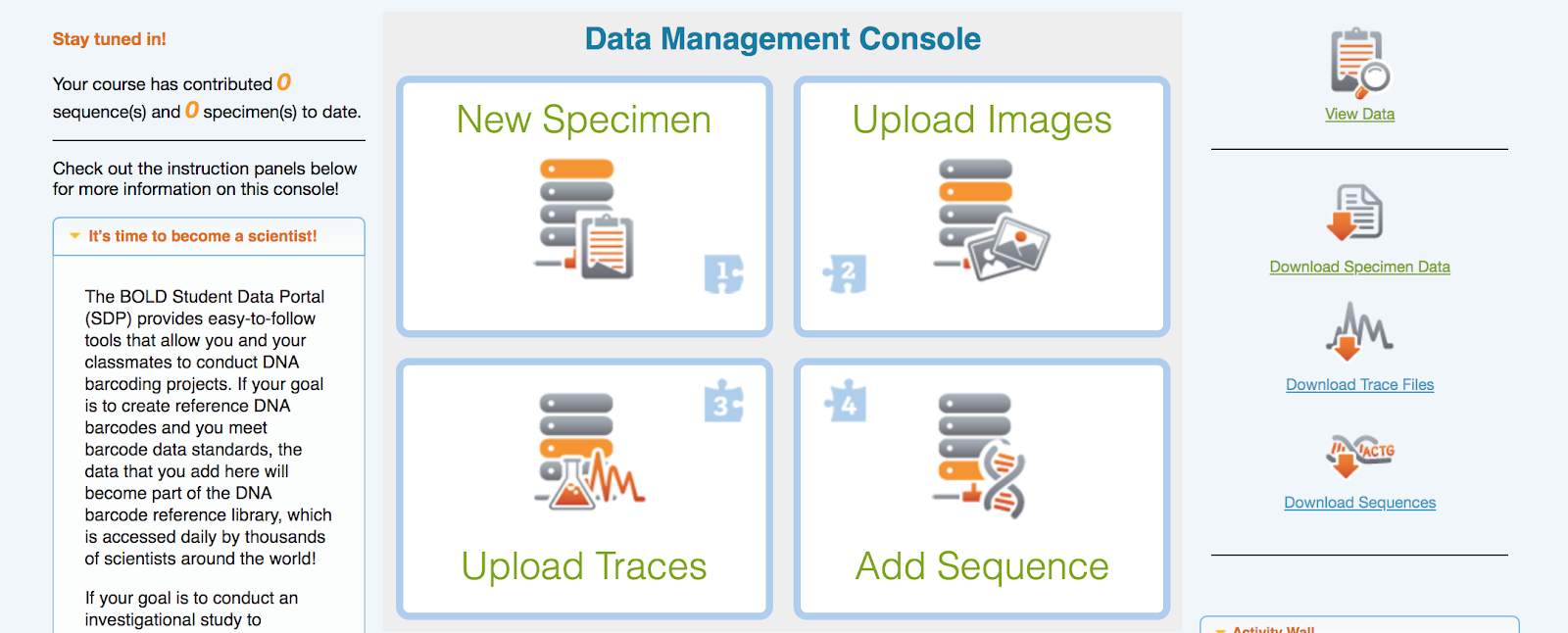
**Be sure to record specific information on the spreadsheet tab labelled GBIF!**

**BOLD Student Portal**

Now go to the BOLD student portal <http://v3.boldsystems.org/index.php/SDP_Home> and **click Students**

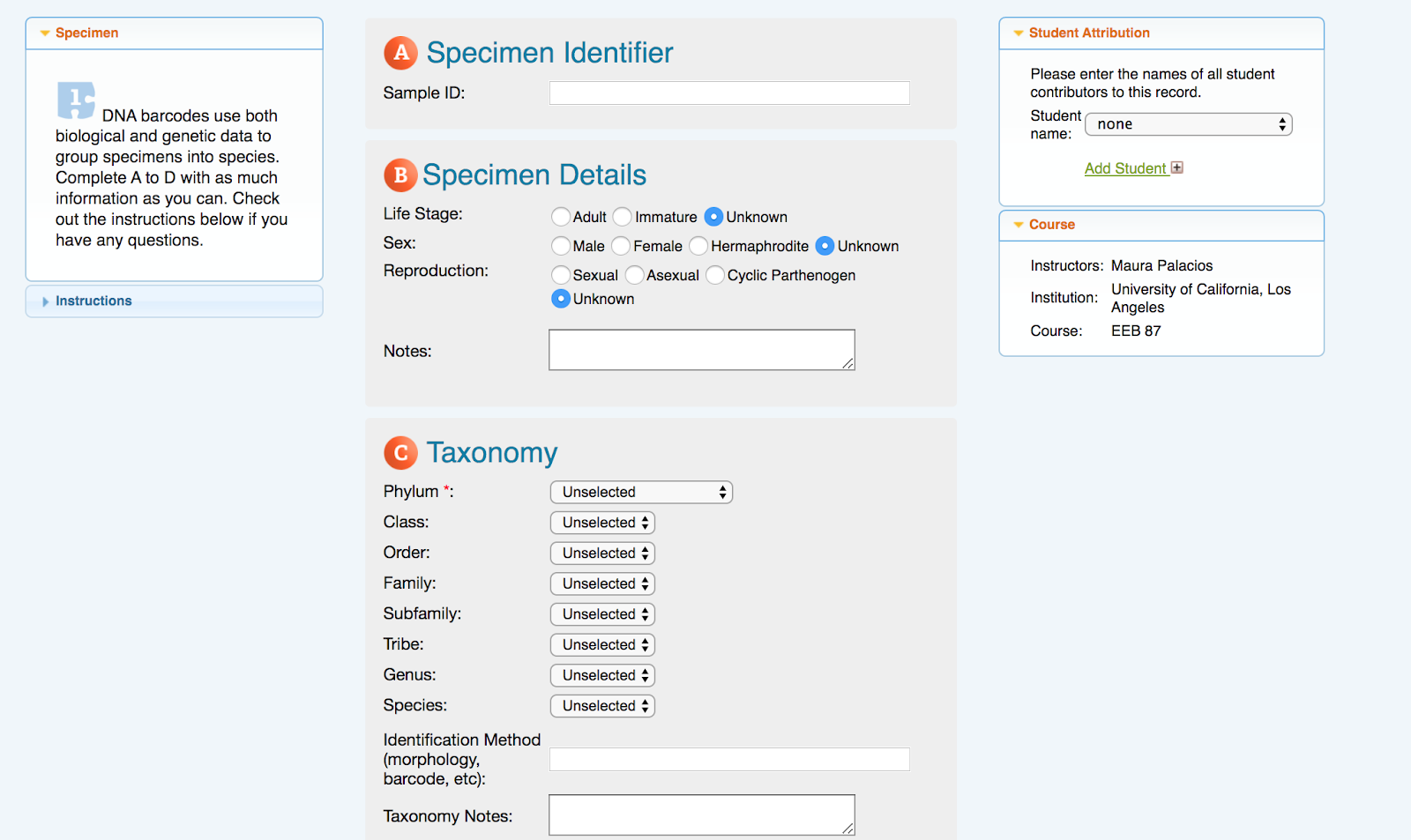
****

For the login information, type in **class-878** for the username and **DTL77** as the password (Be sure to type in). This will take you the main console. We will be working in the data management console. **Click new specimen**

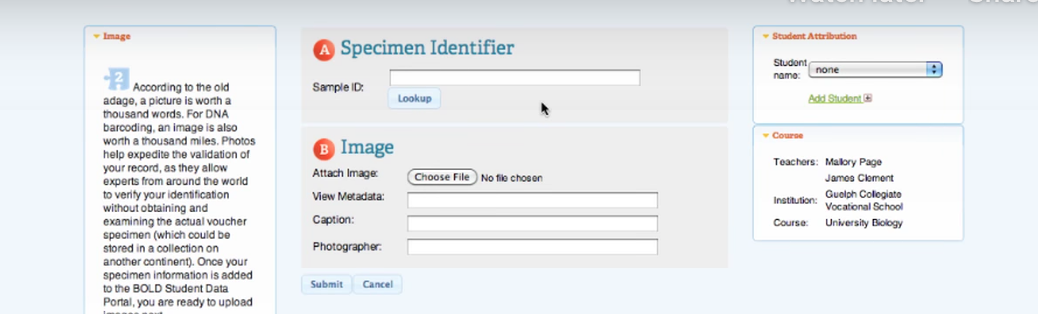


That will take you to this window. For Specimen voucher, we will be using the format year\_initialoftheschool\_yourinitials\_samplenumber. Some of you may have the larval or adult stage of the organism. Input the taxonomy from the GBIF or Genbank. Samples were collected on October 5th and the coordinates are as follows: **Lat:** 34.202856 **Lon:** -118.166464

**Accuracy:** 66m



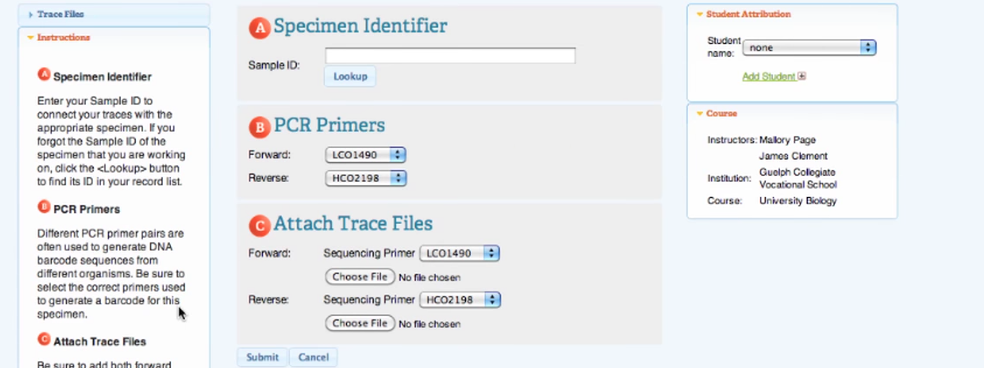
After you submit do not view record and you will be taken back to the data management console. **Click upload images and you will be taken to this screen.**



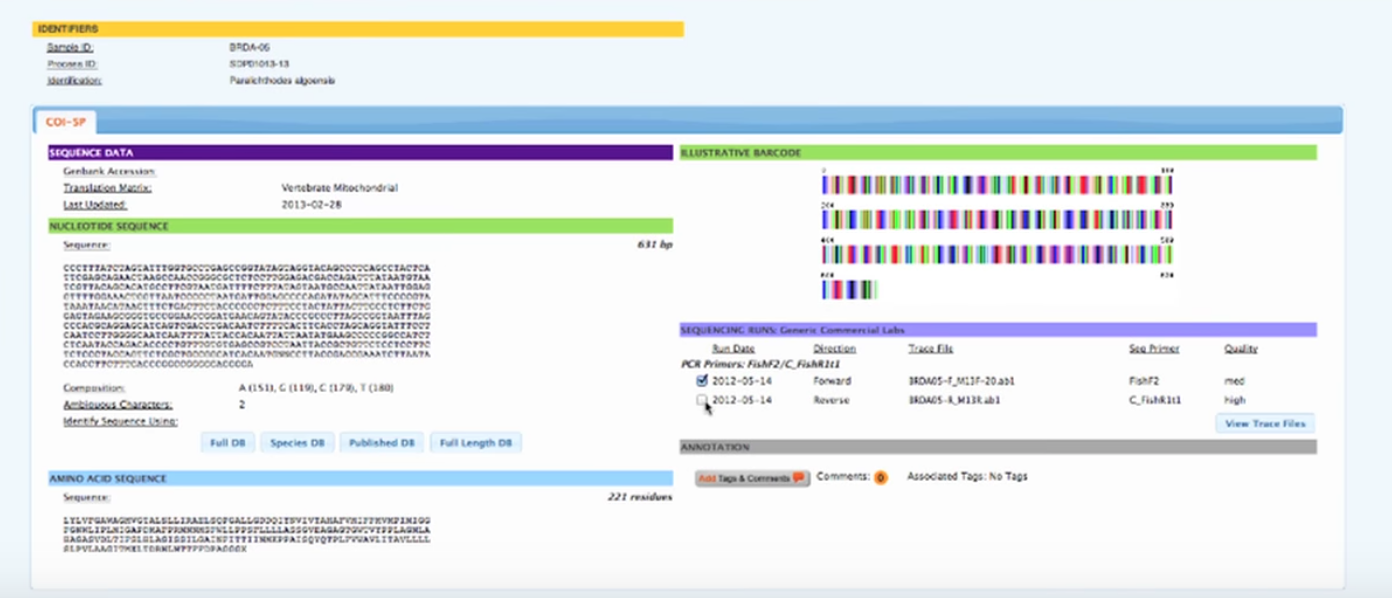
Search for the image of your specimen here (<https://drive.google.com/open?id=1hd3zupVR0cDNgtsxuLgaOXB0ebSbfJPC>) and upload. To change the view of the images, change from list to gridview. Download your image to desktop and then upload. Be sure to input **Caroline Schreck as the photographer.**

You can then view your submission or return to data management console. **From the main console click upload traces and you will be taken to this screen.**

Search for your trace files here (<https://drive.google.com/open?id=1DMMLlA2wHXGTRw2J8HDa29UCOgeJ05p2>) and upload corresponding sequence. Change view to list instead of gridview for easier access. Download your sequence files by clicking the three dots and select download to your Desktop and upload to the server. \* Note you may only have one sequence, in that case submit a file with N’s for the sequence direction you are missing



From the main console **click add sequence** and you will be taken to this screen. Follow the prompts to finalize sequence submission.



**CONGRATULATIONS!!!** You just submitted new data to the BOLD database.

**References:**

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. "Basic local alignment search tool." *Journal of Molecular Biology* 215, no. 3 (1990): 403-410.

GBIF.org (2019), GBIF database. Available from: https://www.gbif.org [21 November 2019].

Ratnasingham, Sujeevan, and Paul DN Hebert. "BOLD: The Barcode of Life Data System (http://www. barcodinglife. org)." *Molecular Ecology Notes* 7, no. 3 (2007): 355-364.

**Assessment Pages**

**Summative Assessment**

*Submission assessment*

Revise BOLD uploads to make sure everyone has submitted

Evaluate information input in the google spreadsheet

*Reflection questions on the activity*

What did you find useful from your DNA mini-barcoding project?

What questions or curiosities do you have after the DNA barcoding activity?

What recommendations do you have for improving this DNA barcoding mini-project?

Pre and post-survey administered by Cold Springs Harbor Lindsay Barone (lbarone@cshl.edu)