Read me for measurements, counts and taxonomic IDs of Zooscan fish from R/V Cook

Helena McMonagle

Nov 10, 2023

The majority of Cook fish (n = 191 in final dataset with Myctophidae, Sternoptychidae, and Gonostomatidae) were photographed at sea with a ruler, and HM measured their standard lengths using ImageJ

Remaining Cook fish were photographed using Zooscan (see Ecotaxa database, project “EXPORTS\_2021”) and can be found by filtering for Actinopterygii (but not those that are eggs or scales, rather, whole fish). In Ecotaxa, there are 57 individual images of fish (whole fish) available. After searching “EXPORTS\_2021” in Project field, click “Update view and apply filters”, and then click on link at the top for the “dedicated page” for this project. Then click on Actinopterygii (n = 135 as of April 25, 2024).

HM took screen shots of each of these fish images, which include the sample label in the bottom right hand corner. The file names were renamed to include this sample label. HM also included “not in xlsx” if the sample label was not present in EXPORTS 2021 long.xlsx, which is a spreadsheet of fish Ecotaxa data generated using Hannah Grossner’s code (technician in Dr. Amy Maas’s lab at BIOS) using Hannah’s code. This spreadsheet includes some individuals labeled as Actinopterygii which do not appear in the EXPORTS 2021 long spreadsheet—not sure why there were 77 rows in the “EXPORTS 2021 long.xslx” spreadsheet and only 57 images in the Ecotaxa images, but the remaining might be fish scales at eggs (n = 17), which nearly accounts for this difference.

Naming convention for sample labels, from email from Amy Maas:

e.g., 201009\_2000\_1\_2\_1\_tot\_1\_86

20=year 10=tow 09=net 2000=size fraction (200,500,1000,2000,5000), rest of the numbers are something about splits and fractions, and the last part is the image number.

Use R code from Hannah to get from the image to the split. We’ll duplicate individuals if they were in a split (e.g., if they were imaged from a quarter split, we’ll multiply them by 4 so that there will be 4 duplicate individuals in the sample to approximately represent what the entire catch might have looked like) 🡪 update: note that some splits were not entered correctly into Ecotaxa originally. Joe Cope in Deb Steinberg’s lab has fixed this, and the correct spilts for Cook fish are available in fishdata.csv in “split\_correct\_Cook” column.

Edited EXPORTS 2021 long.xlsx to include a Std\_length\_mm column, editing tow to be Tow\_number, edited net to be Net\_number, and removed the m from before all the tows (e.g., m10 was changed to 10) and removed the n from before all the net numbers (e.g., n9 was changed to 9). This way the columns in the spreadsheet will match with the main dataframe of fish data from the rest of the samples on the cruise for both Cook and Sarmiento fish.

Protocol:

Renamed each image file according to the label at the bottom left of the image

Pick up with opening each image in ImageJ (File> Open)

Set scale bar to 1 mm using scale bar in bottom left of each Ecotaxa image screen shot

Measure (might need multiple lines)

Save as CSV with same file name as the image file

Use CSVs to get standard lengths of each fish in mm. \*\* no need to do images that are clearly < 10 mm, because we are excluding any fish smaller than 10 mm in length

Measured standard length from images unless a fish was obviously damaged. For individual 209010\_5000\_1\_4\_1\_tot\_1\_47, doubled the length from image because this bristlemouth looked to be roughly broken in half with the other half missing (potentially in a different split for which we don’t have images)

If standard lengths were less than 10 mm, lengths from ImageJ were not saved and “small” was added to the end of the sample label in the image file name

CSV files from image J measurements were named with the same file name as associated images, but with .csv file type instead of .png

Within CSV files, I summed line segment lengths and entered this into Exports 2021 long.xlsx along with morph ID family name

Recall that sample labels marked as tow 12 are really tow 11 (combined tows 11 and 12, see MOCNESS event log notes)

See EXPORTS 2021 long.xlsx. There were 77 fish in the spreadsheet from Hannah Gossner’s code. 25 of those rows in the spreadsheet had data but no image in Ecotaxa, so we removed those (could have been a glitch with using the code but applying it to fish data instead of zoop data). This brings the total to 52. Then, there were 7 images that were in Ecotaxa images but not in spreadsheet (6 of which were too small to include anyway, < 10mm, and one of which was a fish that could not be IDed due to damage but could be Paralepididae, which is an excluded rare taxon). This would bring the total to 59 and there were only 57 fish. Not a perfect accounting of all small fish, but certainly better than leaving out all the Ecotaxa fish from the dataset. Of those that had an Ecotaxa image, 6 of these fish were < 10 mm, so we’ll use a total of n = 45. The fish that were too small or were missing an Ecotaxa image were removed from the spreadsheet we used, titled “Ecotaxa\_small\_fish\_Cook.csv”.

Made small additional edits to Ecotaxa\_small\_fish\_Cook.csv to match column names with other Cook datasheet (see details of data cleaning in load\_data() function in 00-functions.R) and fixed depth interval for net 8 to be 150-100 m instead of 150-0 (this appears to be a typo in how the Ecotaxa\_small\_fish\_Cook.csv data was before)

Note: use vol filtered data from the Cook’s MOCNESS metadata, not from Ecotaxa\_small\_fish\_Cook. Found a discrepancy: Ecotaxa\_small\_fish\_Cook says vol filtered for tow 48 net 2 was 757, whereas Cook’s MOCNESS metadata says it’s 1408. Also, Ecotaxa\_small\_fish\_Cook says tow 48 net 8 filtered 383, whereas Cook’s MOCNESS metadata says 720. Some discrepancy between vol filtered for tow 48 nets 9 and 10, too.

**Update on April 25, 2024**

Met with Joe Cope, who alerted us that there were some issues with the raw data from their lab/Ecotaxa. Some fish had not been uploaded in Ecotaxa, and the information we were using for splits was not correct (E.g., it looked like some fish came from say a quarter split, when actually they came from a whole sample). So, the former issue was leading to missing fish (and thus underestimating biomass), whereas the splits issue was leading to extra fish that shouldn’t have been in our final datasheet (thus overestimating biomass). The following edits were made

**No edits made to fishdata.csv because the code will do this. Instead, edit “Ecotaxa\_small\_fish\_Cook.csv” and “Cook MOC fish data.csv” and let the code edit to create the final fishdata.csv file.**

**in Images in folder Ecotaxa images and ImageJ data**

* New folder added called “images added April 2024” with new fish added. This includes 5 fish in a folder “not in Ecotaxa” that Joe said were not uploaded to Ecotaxa. Three of these are not large enough (>10 mm) to be included in the analysis. In folder “images added April 2024”, we will also add the Ecotaxa images that Joe recently uploaded (Amy and Hannah’s lab also did some of this sample processing to get the files ready for Ecotaxa).
* **Also be sure to add new images (in Images added April 2024) to the SeaBASS folder with all the images to be uploaded to the database**

Added column to FishIDs-v3.csv from Joe Cope that reads “Added\_to\_Ecotaxa\_images\_folder” so I can be sure that I have all the Ecotaxa images now, including the ones that were just uploaded by Joe.

* 201010\_5000\_1\_2\_1\_tot\_1\_83 was already added (and thus has already been measured), but need to delete duplicate (this actually came from a whole sample, not half split)

Delete all 4 rows with 204802\_5000\_1\_4\_1\_63 --these are not fish