

May 08, 2020

Version 1

NCBI_COI_Submission V.1

DOI

<https://dx.doi.org/10.17504/protocols.io.bf6ijrce>

Avery Hatch¹

¹UCSD - Scripps Institution of Oceanography

Rouse Lab



Avery S Hiley

UCSD - Scripps Institution of Oceanography

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.bf6ijrce>

Protocol Citation: Avery Hatch 2020. NCBI_COI_Submission. protocols.io <https://dx.doi.org/10.17504/protocols.io.bf6ijrce>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

This protocol is for uploading COI sequences to GenBank (NCBI) via the "Submission Portal".

Created: May 08, 2020

Last Modified: May 08, 2020

Protocol Integer ID: 36778

- 1 First and foremost, gather all your COI sequences that need to be uploaded and align them in either Geneious, Mesquite, or using the online MAFFT version 7 server: <https://mafft.cbrc.jp/alignment/server/>.
- 2 Blast one of the sequences and make sure that the first record it matches with on NCBI says 'Plus/Plus'. This indicates that the sequence is in the correct forward 5' direction. If it says 'Plus/Minus', then reverse complement your entire alignment.
- 3 Important note: Please make sure there are no ambiguities and/or gaps in your sequences. If a COI sequence has lots of bases that aren't ACGT (e.g. N, ?, K, M, other IUPAC codes, etc.):
 - 3.1 Please look at your original De Novo assembly in Geneious and check if you can make a nucleotide base call. Compare the amplitudes of the Forward/Reverse peaks at the corresponding nucleotide positions in the chromatogram (.ab1 files), and choose the nucleotide with the greater amplitude.
 - 3.2 Check if the ends of the sequence assembly need to be trimmed due to low/poor quality.
- 4 Make any nucleotide edits and/or deletions as necessary before proceeding.
- 5 Open your final edited COI alignment in Mesquite, and click on the colorful rubix cube icon on the left that says 'Character Matrix'. Then click 'List & Manage Characters'.
- 6 Highlight/select all the columns & rows. Click the 'Codon Position' column heading > 'Set Codon Position' > 'Minimize Stop Codons'.
- 7 In the top task bar, click 'Columns' > 'Current Genetic Codes'.
- 8 Click the 'Genetic Code' column heading > 'Invertebrate Mitochondrial'.
- 9 In the top task bar, click 'Characters' > 'Make New Matrix from' > 'Translate DNA to Protein'.
- 10 View the protein translation of your character matrix, and make sure there are no black stop codons with an asterisk.
- 11 If the translation looks good (no stop codons), proceed to view the original nucleotide character matrix again.



- 12 In the top task bar, click 'Display' > 'Color Matrix Cells' > 'Color Nucleotide by Amino Acid'.
- 13 If any of your COI sequences do not start on codon position 1, delete the corresponding one or two nucleotides at the front of the sequence that are preventing this.
- 14 Once all COI sequences start on codon position 1 and do not have ambiguities and/or gaps, export the alignment as a fasta file.
- 15 Open this fasta file in TextWrangler or BBEdit. Edit all of your sequence headers to match this format:

 >A9919 [organism=Peinaleopolynoe mineoi] Peinaleopolynoe mineoi voucher
 SIO:BIC:A9919 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial
- 15.1 In this example, "A9919" is the BIC number; use either the BIC number or a different institution's catalog number here to represent the sequence ID. If catalog numbers are not assigned yet, you may use a lab code (e.g. S25170) as the sequence ID here.
- 15.2 "[organism=Peinaleopolynoe mineoi]" is obviously where you insert the species name for the corresponding sequence ID.
- 15.3 "Peinaleopolynoe mineoi": Repeat the species name here.
- 15.4 "voucher SIO:BIC:A9919": Follow this proper format to identify the corresponding BIC number with our institution abbreviations. If the specimen is deposited at a different institution, this will instead mimic the abbreviations of that place before listing the catalog number (e.g. MNHN for Muséum national d'Histoire naturelle). If your catalog numbers aren't assigned yet, exclude this section altogether.
- 15.5 "cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial" should be listed at the end of the aforementioned information.
- 16 After all of your sequence headers follow this format, save the fasta file again.
- 17 Create an online account for NCBI's **Submission Portal** platform:
<https://submit.ncbi.nlm.nih.gov/>.
- 18 Start a new submission.



- 19 **Submission Type:** 'Metazoan (multicellular animal) Mitochondrial COX1'.
- 20 **Submitter:** Fill out the corresponding information (see screenshot below for our lab details). Make sure to check 'Update my contact information in profile' in order for future submissions to use this specific info by default.

Submitter

Affiliation

The information you give here will be displayed in the final sequence records.
For address details, provide the primary address where work was done to generate the data in this submission.

★ Submitting organization	★ Department			
Scripps Institution of Oceanography	Marine Biology Research Department			
★ Street	★ City	★ State/Province	★ Postal code	★ Country
8750 Biological Grade, Hubbs Hall Rm. 235C	La Jolla	CA	92037	United States

Contact information

GenBank may use this information to contact you about your submission, it will not be displayed in the final sequence records.

★ Email (primary)	Email (secondary)	Please provide an alternate email address if you have one.
seahatch3@gmail.com	ahatch@ucsd.edu	
★ First (given) name	Middle name	★ Last (family) name
Avery	Sea	Hatch
Phone	Fax	

☒ Update my contact information in profile

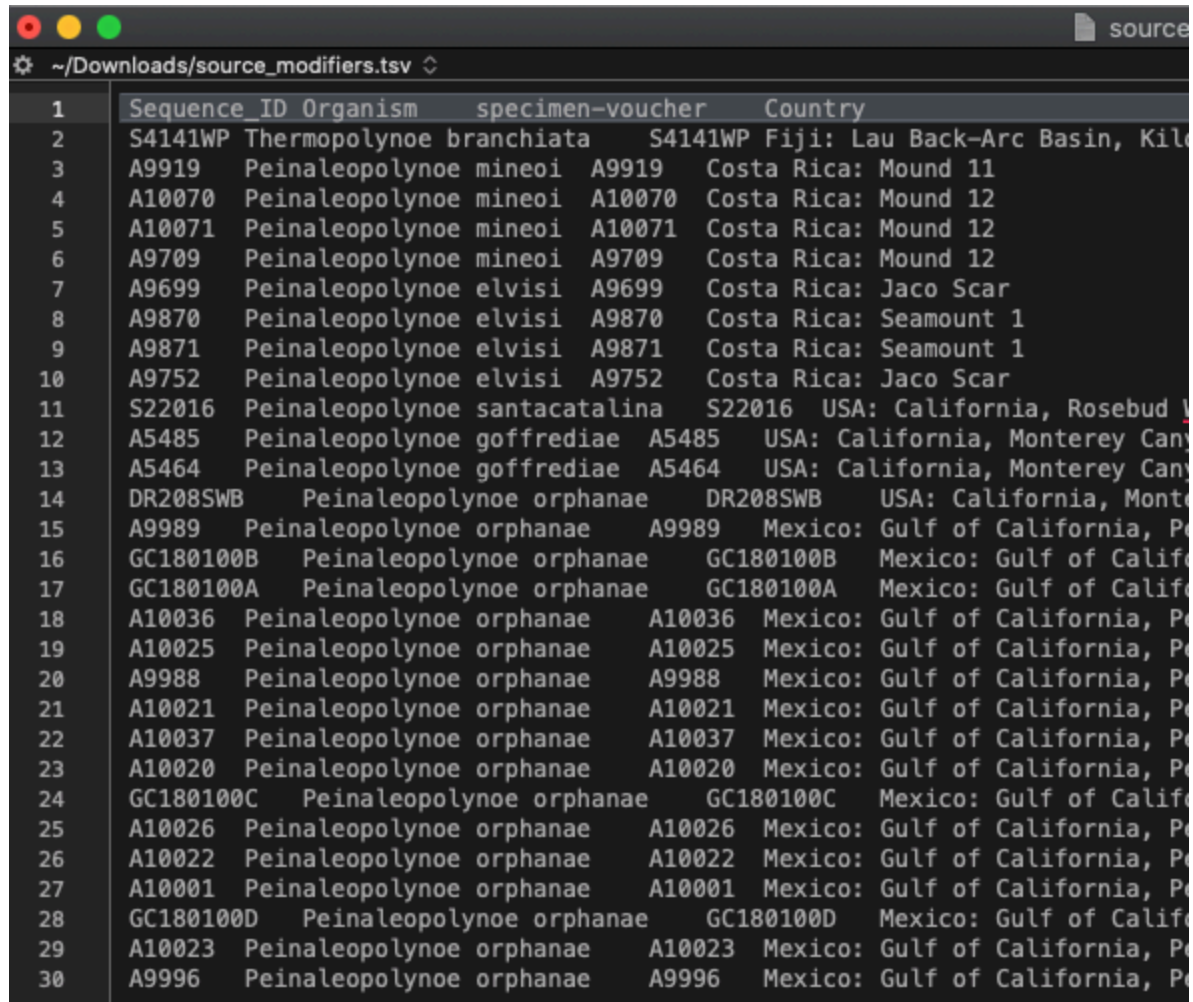
- 21 **Sequencing Technology:** Select 'Sanger dideoxy sequencing' & 'Assembled sequences (each sequence was assembled from two or more overlapping sequence reads)'.
- 22 **Sequences:** Select 'Release on specified date or upon publication, whichever is first' or release immediately if you are late in uploading your sequences (this should almost never be the case). Typically you will want to choose a year in advance to be safe. Next, upload your COI fasta file that was completed in step 16.

- 23 **Source Info:** Under 'Do your sequence IDs represent one of these?', select 'Specimen-Voucher' if you followed the headers format in step 15. If catalog numbers aren't assigned yet and your sequence IDs were listed as lab codes (e.g. S25147), select 'Isolate' here.
- 24 **Source Modifiers:** This is the section where you add key details that you would like to be attached to the sequences (e.g. locality and depth). However, only the organism name and specimen-voucher are required for submission. You should have already included these in the previous steps, so you may continue if you do not wish to apply additional source modifiers.

*Since all the specifics will be in your publication, you may typically stick with locality (column name = 'Country') and depth (column name = 'Altitude').

*If you would like to add a couple key source modifiers, then choose one of the following two options under 'How do you want to apply source modifiers?':

- 24.1 Option 1. 'Upload a tab-delimited table': Proceed to download the source modifier template table. You can edit this file in several programs (TextEdit, TextWrangler, BBEdit, or even Microsoft Excel). In a text edit program, separate the information in each column by inserting 1 tab. The locality information in the 'Country' column must go from broad to specific. For example, "Costa Rica: Mound 12" follows this format. The depths entered in the 'Altitude' column must be negative (e.g. "-1800 m"). An example of a tab-delimited table is attached below:



1	Sequence ID	Organism	specimen-voucher	Country
2	S4141WP	Thermopolynoe branchiata	S4141WP	Fiji: Lau Back-Arc Basin, Kilo
3	A9919	Peinaleopolynoe mineoi	A9919	Costa Rica: Mound 11
4	A10070	Peinaleopolynoe mineoi	A10070	Costa Rica: Mound 12
5	A10071	Peinaleopolynoe mineoi	A10071	Costa Rica: Mound 12
6	A9709	Peinaleopolynoe mineoi	A9709	Costa Rica: Mound 12
7	A9699	Peinaleopolynoe elvisi	A9699	Costa Rica: Jaco Scar
8	A9870	Peinaleopolynoe elvisi	A9870	Costa Rica: Seamount 1
9	A9871	Peinaleopolynoe elvisi	A9871	Costa Rica: Seamount 1
10	A9752	Peinaleopolynoe elvisi	A9752	Costa Rica: Jaco Scar
11	S22016	Peinaleopolynoe santacatalina	S22016	USA: California, Rosebud V
12	A5485	Peinaleopolynoe goffrediae	A5485	USA: California, Monterey Cany
13	A5464	Peinaleopolynoe goffrediae	A5464	USA: California, Monterey Cany
14	DR208SWB	Peinaleopolynoe orphanae	DR208SWB	USA: California, Monte
15	A9989	Peinaleopolynoe orphanae	A9989	Mexico: Gulf of California, Pe
16	GC180100B	Peinaleopolynoe orphanae	GC180100B	Mexico: Gulf of Calif
17	GC180100A	Peinaleopolynoe orphanae	GC180100A	Mexico: Gulf of Calif
18	A10036	Peinaleopolynoe orphanae	A10036	Mexico: Gulf of California, Pe
19	A10025	Peinaleopolynoe orphanae	A10025	Mexico: Gulf of California, Pe
20	A9988	Peinaleopolynoe orphanae	A9988	Mexico: Gulf of California, Pe
21	A10021	Peinaleopolynoe orphanae	A10021	Mexico: Gulf of California, Pe
22	A10037	Peinaleopolynoe orphanae	A10037	Mexico: Gulf of California, Pe
23	A10020	Peinaleopolynoe orphanae	A10020	Mexico: Gulf of California, Pe
24	GC180100C	Peinaleopolynoe orphanae	GC180100C	Mexico: Gulf of Calif
25	A10026	Peinaleopolynoe orphanae	A10026	Mexico: Gulf of California, Pe
26	A10022	Peinaleopolynoe orphanae	A10022	Mexico: Gulf of California, Pe
27	A10001	Peinaleopolynoe orphanae	A10001	Mexico: Gulf of California, Pe
28	GC180100D	Peinaleopolynoe orphanae	GC180100D	Mexico: Gulf of Calif
29	A10023	Peinaleopolynoe orphanae	A10023	Mexico: Gulf of California, Pe
30	A9996	Peinaleopolynoe orphanae	A9996	Mexico: Gulf of California, Pe

- 24.2 Option 2. 'Use an editable table': This is self explanatory. You can manually add columns with corresponding information for each sequence ID via the online interface. However, please copy and paste, and **save** your work somewhere else for your records.
- 25 **References:** Under 'Sequence authors', add yourself (unless someone else did the lab work). Select the corresponding 'Publication status' that applies to you. Add the title of your paper and 'Specify new authors' to add the names of all authors on your paper.
- 26 **Review & Submit:** Check the details of your submission here. If everything looks good, submit your sequences.