# Protocol: Mitogenome Annotation and GenBank Submittal

This protocol builds upon the [NEFC mitochondrial genome sequencing protocol](#), and is designed to clean up and finalize mitogenome annotation, and for generation of a Sequin file for GenBank submission.

1. Go to https://submit.ncbi.nlm.nih.gov/genbank/template/submission/ and complete the form and save the resulting template.sbt file.
2. Export the sequence data from Geneious as a FASTA sequences/alignment (*.fasta) file with the name of the id, e.g. SNARCC_18-003, uncheck the 'Include sequence description' box. NOTE: You could also do a 'Batch Export' to do all files at once saved as separate files with a specified .fsa extension.
3. Rename the extension of the file in the previous step to .fsa if necessary.
4. Open the .fsa file and add the following to the definition line (updated with your sample information) after the seqID: [topology=circular] [completeness=complete] [location=mitochondrion] [moltype=genomic DNA] [mgcode=2] [organism=Ptychocheilus lucius] [common=Colorado pikeminnow] [isolate=SNARCC18-003] [isolation-source=New Mexico_SNARRC Hatchery] [country=USA]
   NOTE: You can export the entire sample table in Geneious as a tab delimited file and then open it up in Excel in order to simplify gathering of this information.
5. Export the annotation from Geneious as a Genbank sequences (flat file) (*.gb)
   NOTE: This step could also be performed as a batch export.
6. Go to http://174.55.94.160/annotate and click on 'Prepare GenBank File'.
7. Import the *.gb file from step 5 in order to add source and locus tag information, standardize CDS, rRNA, and tRNA nomenclature, remove unnecessary feature attributes, and summarize findings for error checking.
8. Enter the locus_tag prefix.
   NOTE: A locus_tag is necessary to use this app at present. To receive a registered locus_tag from GenBank, submit a [BioSample](#) request for the sample organism.
9. Click the 'Annotate' button.
10. Download the edited file from the previous step, edit for any missing genes (should be 22 tRNA (8 complement), 13 CDS (1 complement), 2 rRNA, and 1 D-loop), and verify all translation exceptions (transl_except) by confirming that the 'AA' bases are not present (view in Geneious), and that the 'pos:' and base pair end range value for both the CDS and the accompanying gene are equal to the base pair number of the final 'T' base.
    NOTE: At present a single 'A' following the 'T' is ignored and not included in the CDS and gene ranges.

11. To verify annotation or examine a second annotation source, the *.fsa file can also be run through MitoAnnotator, and the output *.txt file compared to the *.gb file downloaded in step 10.
12. Once the *.gb file is deemed to be complete and error free, go to http://174.55.94.160/annotate and click on 'Create Sequin File'.
13. Import a GenBank template file (*.sbt), a FASTA file (*.fsa), and a GenBank flat file (*.gb) and click 'Run'.
14. If warnings or errors were encountered during the run, the red triangular exclamation symbol (see below) will be visible. Hover the cursor over the symbol to see the warnings and errors as a popup, or click the symbol to download the associated validation file (.val) produced by the tbl2asn run.
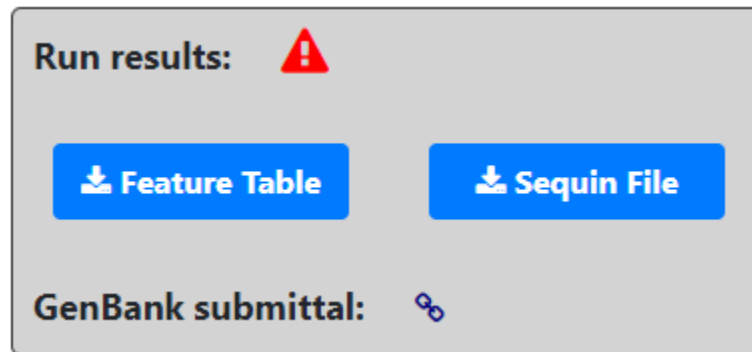
15. If errors are present, resolve them in your input files and rerun. Definitions of common errors can be found at https://www.ncbi.nlm.nih.gov/genbank/genome_validation/.
16. Once a run completes without any errors, download the resulting Sequin file (*.sqn).
17. Go to https://www.ncbi.nlm.nih.gov/LargeDirSubs/dir_submit.cgi, complete the form, upload the *.sqn file(s), and submit to GenBank.