

Jun 11, 2025

Finalizing Mitogenome Annotation in Geneious

DOI

<https://dx.doi.org/10.17504/protocols.io.yxmvmkjrbg3p/v1>

Dakota Betz¹

¹ucsd

Rouse Lab



Sonja Huc

UNCW

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.yxmvmkjrbg3p/v1>

Protocol Citation: Dakota Betz 2025. Finalizing Mitogenome Annotation in Geneious. **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.yxmvmkjrbg3p/v1>

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

We use this protocol and it's working

Created: August 16, 2021

Last Modified: June 11, 2025

Protocol Integer ID: 52360

Keywords: mitogenome annotation in geneious, finalizing mitogenome annotation, geneious, annotation

Disclaimer

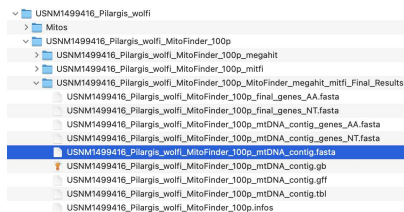
Our protocols are constantly evolving and old versions will be deleted.
The documents here are not intended to be cited in publications

Abstract

Our protocols are constantly evolving and old versions will be deleted.
The documents here are not intended to be cited in publications

ð

- 1 BEFORE you start annotating your mitogenomes, you want to run them through another annotating software, MITOS.
- 1.1 Go to the MITOS web server: <http://mitos.bioinf.uni-leipzig.de/index.py> and upload your sequence.
Enter your name, email, if you want you can add a job identifier (i.e. what is the mitogenome you're uploading for annotation), **MAKE SURE** to change the genetic code to the appropriate choice (in our lab it's generally always "05 - Invertebrate"), then navigate to your **assembled mitogenome** in Finder. Expand the **MitoFinder** folder, then expand the "..._Final_Results" folder, and find the **"Sample_MitoFinder_100p_mtDNA_contig.fasta"** file



Opening Geneious and Setting Up Workspace

- 2 **Log into Geneious** - (or, if you have your own version, just open that) Use the floating license server. IT IS VERY IMPORTANT YOU DO NOT UPDATE GENEIOUS - the floating license will not work with newer versions.



- 2.1 When prompted, enter the following server IP address and port, and make sure to select "Use floating license server".

Command

Server IP

132.239.80.207

Command

Port

27001

Enter your License Details

Want to trial Geneious? [Register online](#) and we'll email you a license to paste in below:

☐ Use license key

Licensee name:

License Key:

☒ Use floating license server

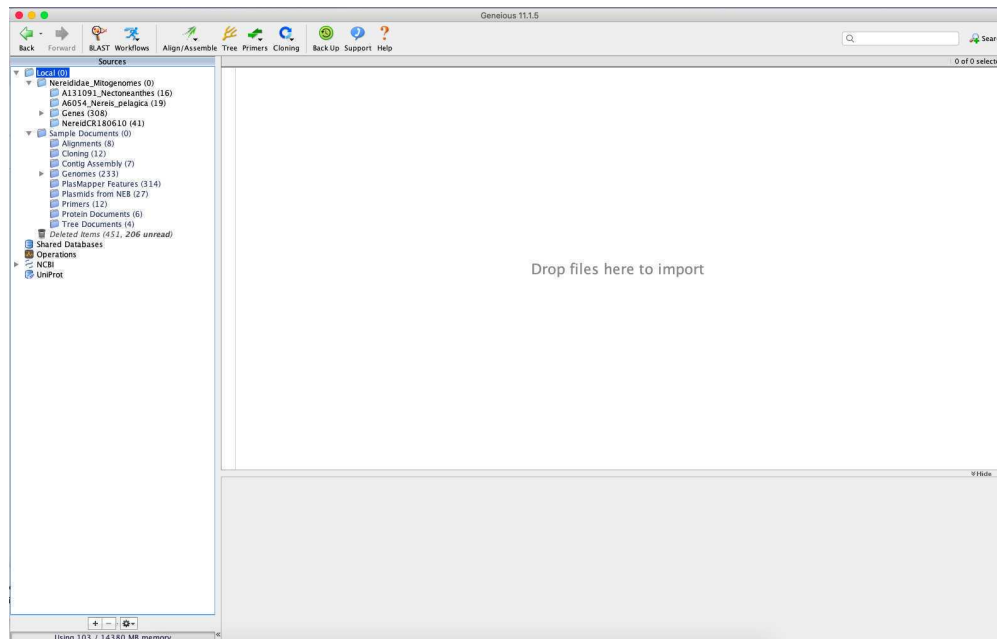
Server:

Port:

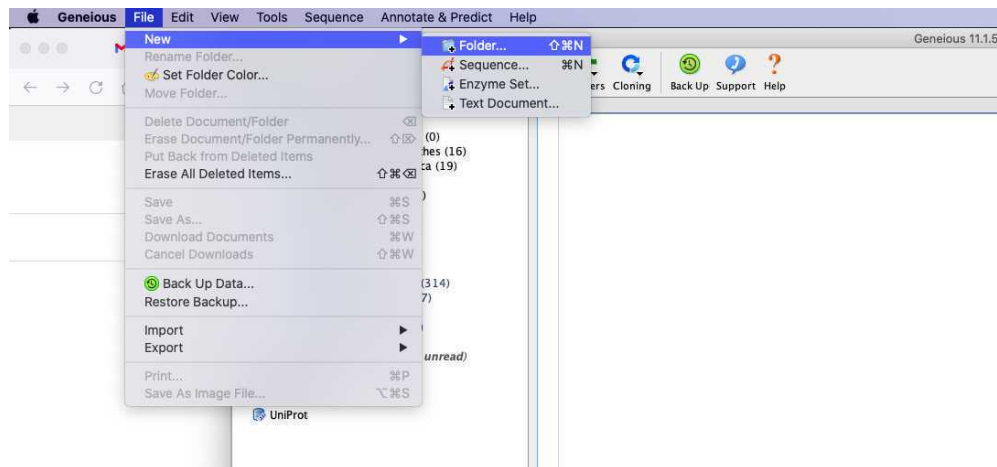
☐ Use Sassafra KeyServer

☐ Log in to Geneious Account

2.2 Create A New Folder: When you first open Geneious, you should have a blank workspace with some sample documents.



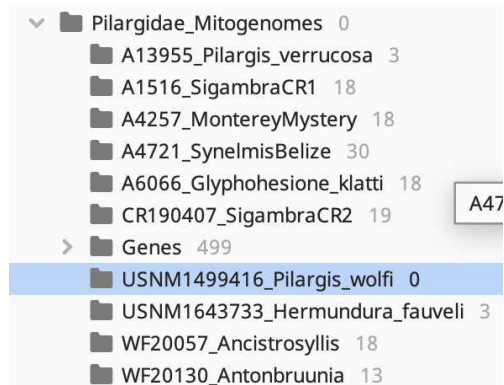
You will want to **create a new folder** for the mitogenome you will work on. To do this, go to the tool bar at the top of your monitor and click "**File**", then "**New**", and "**Folder...**"



2.3 Name it something like "SpecimenFamily_Mitogenomes" (i.e. "Pilargidae_Mitogenomes"), or whatever is useful to you. It's good practice to not use any spaces in the folder and file names.



- 2.4 If you will be working on **multiple specimens in the same family** or greater taxonomic category than just one sample from one species, **make another folder** within the folder you just made i.e. with your specific sample name.



Example - highlighted is the new folder.

- 2.5 **Create another folder** within the family or greater taxonomic category folder named **"Genes"**, and within *that* folder, **create 28 new folders**. For each protein coding gene, you want 2 folders; one that is **"ProteinCodingGene_AA"** for the amino acid translation sequence, and the other that is **"ProteinCodingGene_NT"** for the nucleotide sequence (i.e. ATP6_AA and ATP6_NT). 12S (rrnS) and 16S (rrnL) are not protein coding genes, so for them, you only need one folder each: **"rrnS_NT"** and **"rrnL_NT"**.



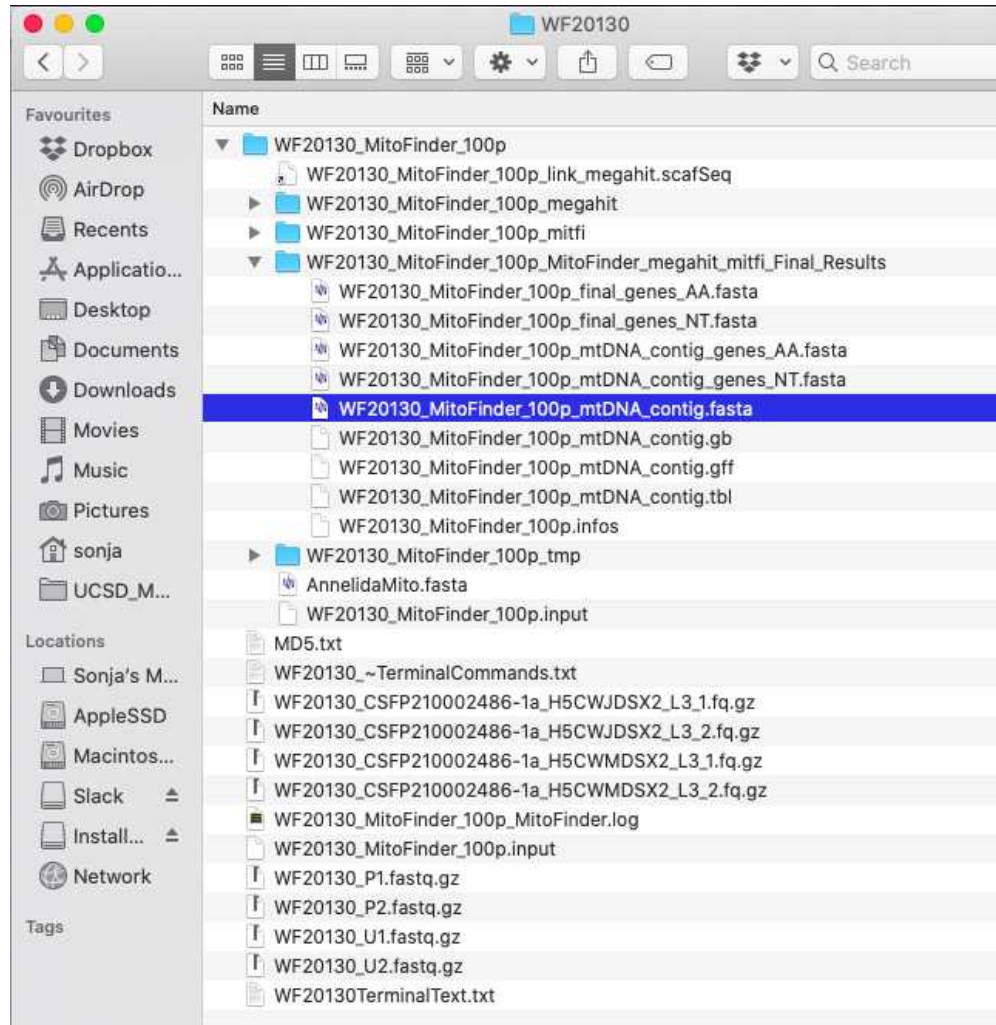
- ▼ **Pilargidae_Mitogenomes** 0
 - A13955_Pilargis_verrucosa
 - A1516_SigambraCR1 18
 - A4257_MontereyMystery 18
 - A4721_SynelmisBelize 30
 - A6066_Glyphohesione_klatti
 - CR190407_SigambraCR2 19
 - G394A_Micronephthys_minu
- ▼ **Genes** 0
 - ATP6_AA 20
 - ATP6_NT 20
 - ATP8_AA 19
 - ATP8_NT 19
 - COX1_AA 20
 - COX1_NT 20
 - COX2_AA 20
 - COX2_NT 20
 - COX3_AA 20
 - COX3_NT 20
 - CYTB_AA 20
 - CYTB_NT 20
 - ND1_AA 20
 - ND1_NT 20
 - ND2_AA 20
 - ND2_NT 20
 - ND3_AA 20
 - ND3_NT 20
 - ND4_AA 20
 - ND4_NT 20
 - ND4L_AA 20
 - ND4L_NT 20
 - ND5_AA 20
 - ND5_NT 20
 - ND6_AA 19
 - ND6_NT 20
 - rrnL_16S_NT 17
 - rrnS_12S_NT 21

Importing Sequences

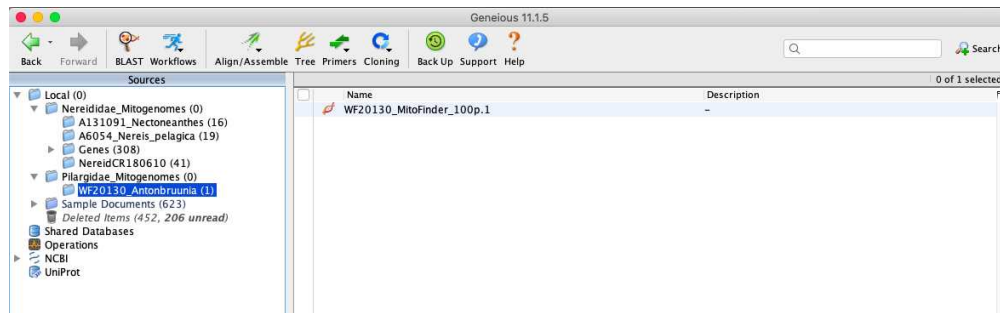
3 Importing Your Data

- 3.1 **Drag In Your "...contig.fasta" File:** Go to wherever you have your **assembled mitogenome** in Finder. Expand the **MitoFinder** folder, then expand the

"...Final_Results" folder, and find the "**Sample_MitoFinder_100p_mtDNA_contig.fasta**" file.



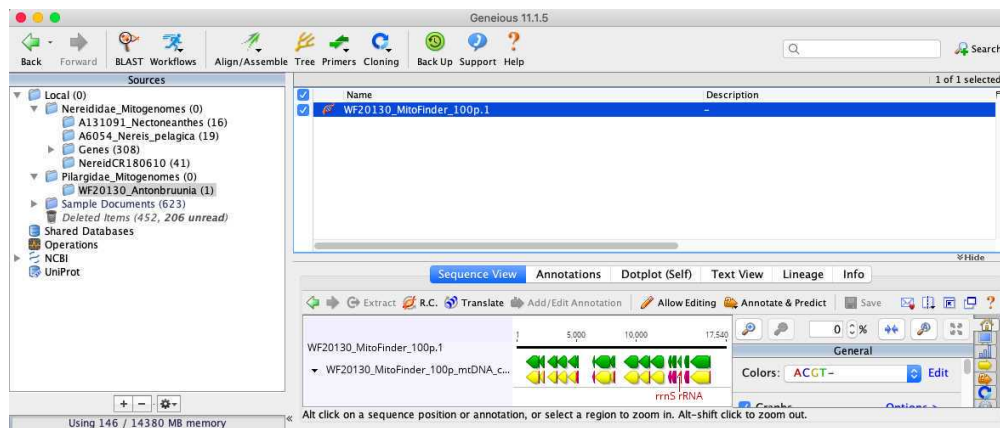
- 3.2 Drag the "**Sample_MitoFinder_100p_mtDNA_contig.fasta**" file into Geneious, where it says "Drop files here to import". Your file should show up like this. If you accidentally placed it in the wrong folder, you can click on it and move it around, like in Finder.



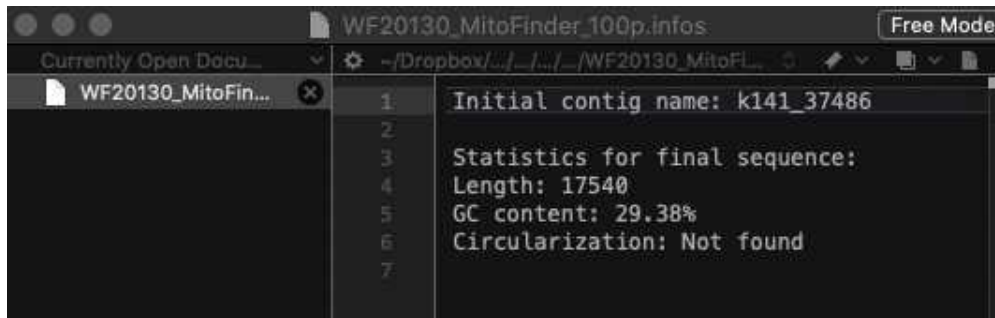
- 3.3 Drag In Your "contig.gff" File:** In the same folder as your "contig.fasta" file, locate the "**Sample_MitoFinder_100p_mtDNA_contig.gff**" file, and drag it into Geneious, into the same area where your "contig.fasta" file shows up. This will visually map the genes MitoFinder found.

Sometimes Geneious will ask you questions about this - i.e. what you want to map it to, just say yes for everything.

Sometimes after importing your data won't visually look like this - usually, deleting the files from your Geneious working folder and re-importing them fixes this issue (sometimes a you have to re-import multiple times before it works)

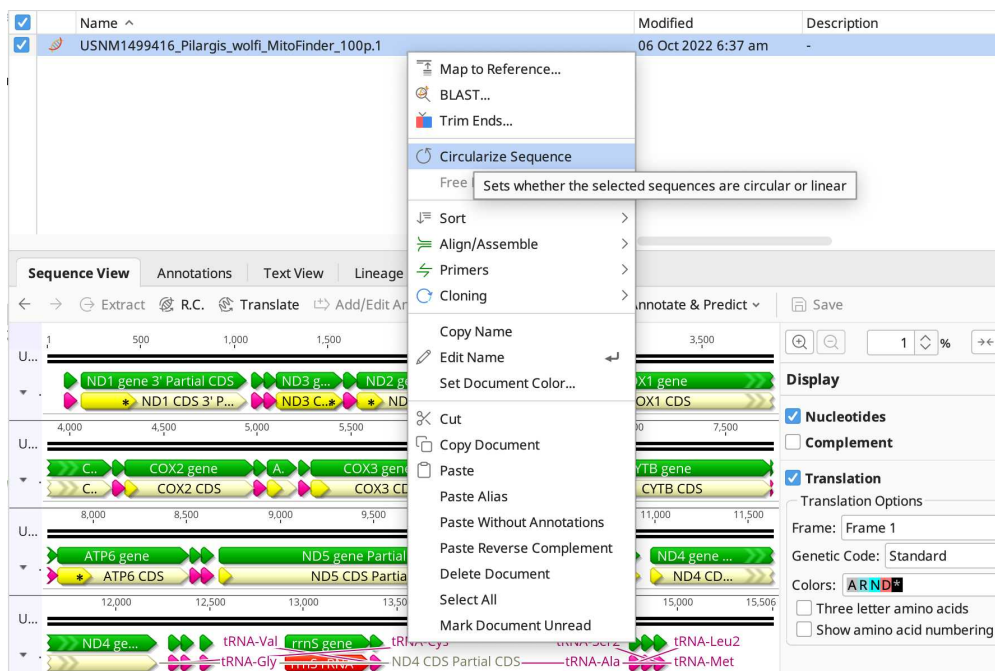


- 3.4 Circularization:** The mitogenome should be circular. However, MitoFinder most often does not find the circularization and leaves your genes linear. If you want to check, locate the "**Sample_MitoFinder_100p.infos**" file and open it in a text editor.

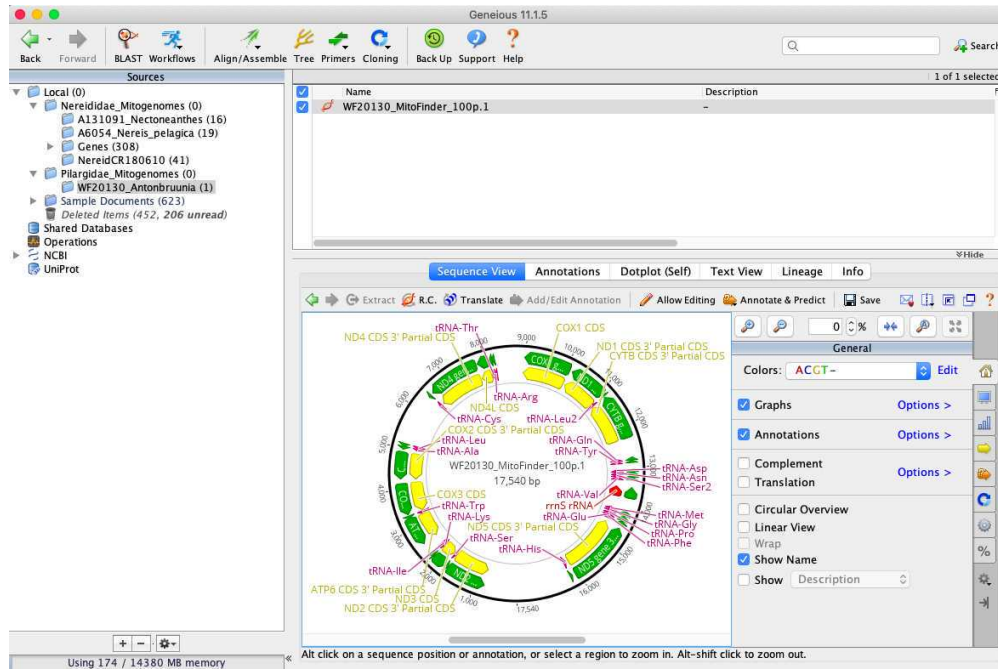


If circularization is not found (which you can also see by seeing Geneious does not automatically circularize it), we want to manually force it. The non-coding region (D-loop) can be very repetitive and it's hard for the program to decide where the genome ends.

3.5 Forcing Circularization: Right click on your sequence in Geneious and click on "Circularize Sequence".



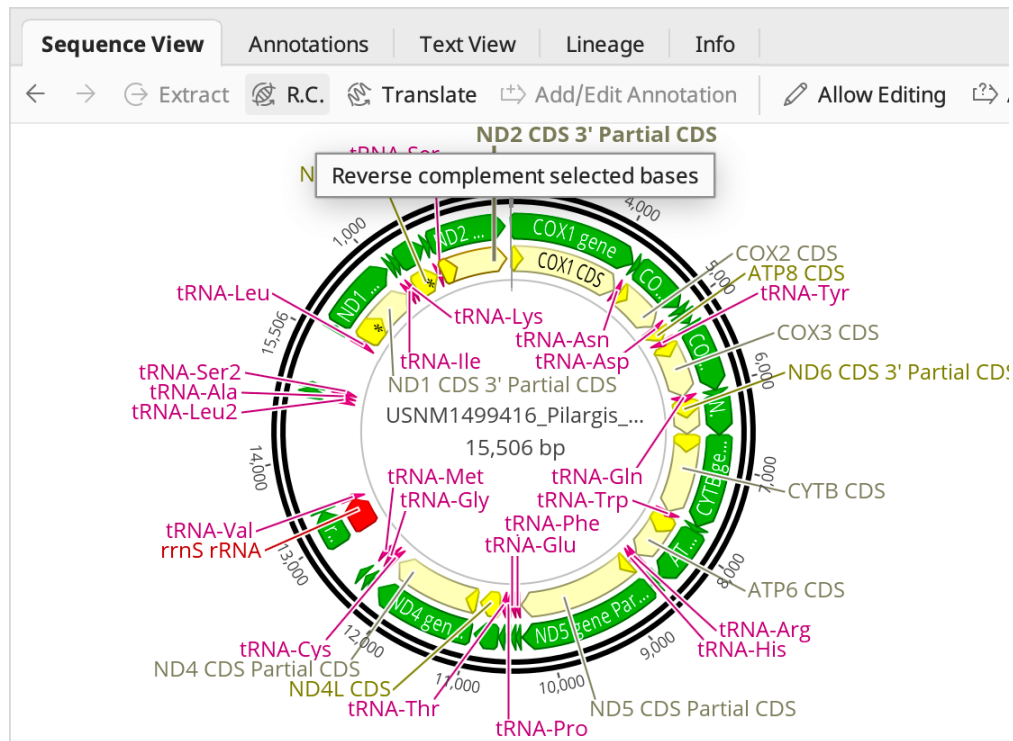
The change should show up in your bottom part of Geneious, where the sequence is no longer mapped in a straight line, but now becomes a circle.



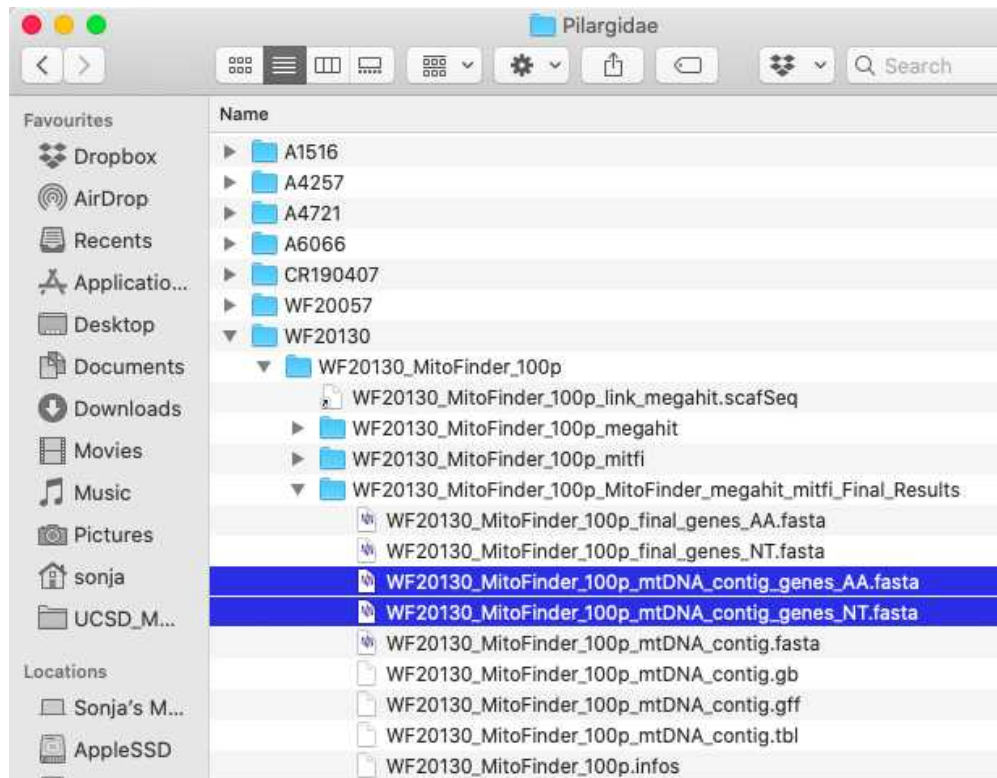
Save changes.

Note

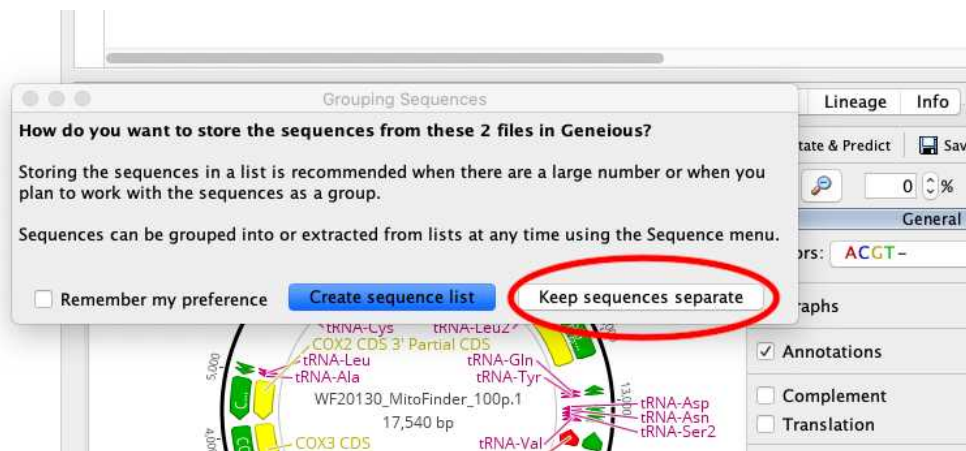
NOTE: You want COX1 to be going from left to right (clockwise)! If it is not, you have to reverse complement the entire mitogenome! Do this by clicking "R.C.". (In this screenshot, COX1 is fine, but in the above one, it is not and must be reverse complemented).



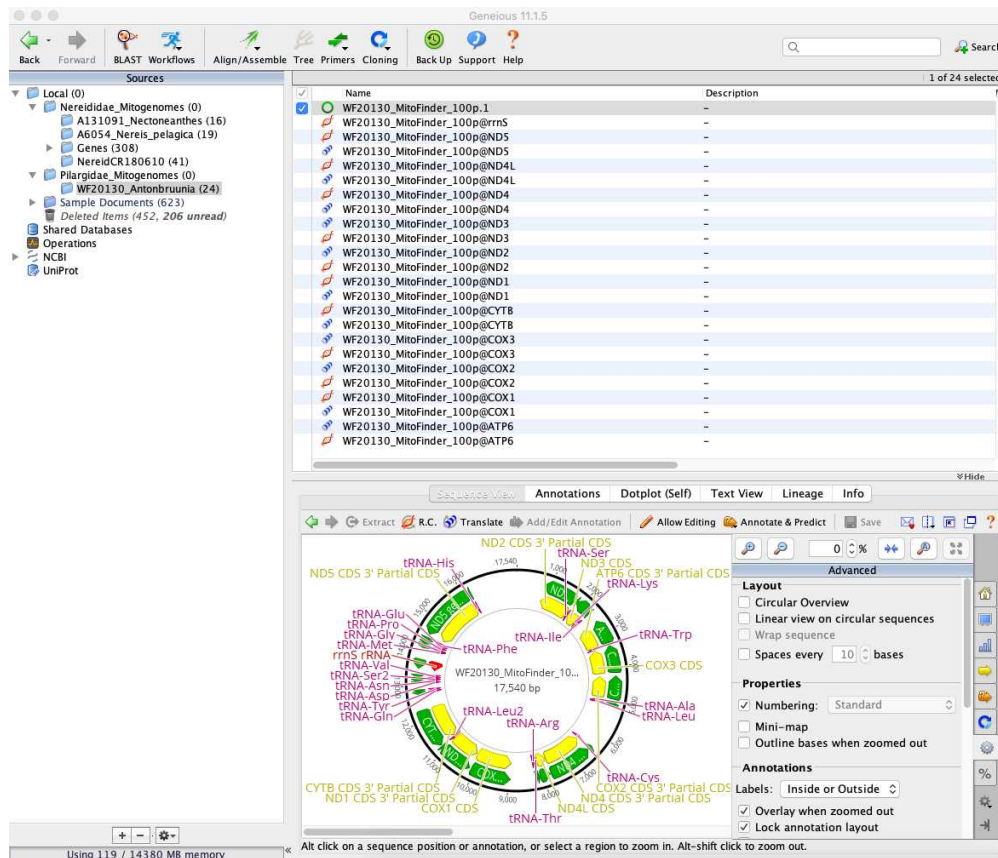
- 3.6 **Import Amino Acids and Nucleotides:** In Finder, in the same folder as before, locate and drag into Geneious the "**Sample_MitoFinder_100p_mtDNA_contig_genes_AA.fasta**" and the "**Sample_MitoFinder_100p_mtDNA_contig_genes_NT.fasta**" files. You don't want to import the "..._final_genes_..." files, because they aren't necessarily correct.



- 3.7 Geneious will ask how you want the sequences from the files stored. Choose **"Keep sequences separate"**.



Your Geneious workspace should then look something like this:



- 3.8 There should be **13 AA files**, which are noted by the blue icons - these are the **Amino Acid** translations of the protein coding genes, and **15 NT files**, which are noted by the orange double helix icons - these are the **NucleoTide** sequences of both the protein coding genes and the non-coding genes (12S/rrnS and 16S/rrnL).

HOWEVER, sometimes MitoFinder doesn't find all of the genes - that's okay; that's why we've run MITOS, and if even MITOS doesn't find them, then we can troubleshoot using steps outlined later on in the protocol.

First Pass Check

- 4 The first way we will check how correct MitoFinder is, is by checking for start and stop codons in the AA sequences of the protein coding genes.
- 4.1 **Click** on one of your AA files (blue icon). In the bottom part of Geneious, you should see a sequence of letters, which are the single letter codes for amino acids. Ideally, the sequence will **start** with the letter '**M**', which signifies a start codon, and **end** with an **asterisk** '*', which signifies the stop codon.

- 4.2 Wherever you are keeping your notes, **write down** for each gene if it did or did not start with an M and end with an asterisk.
- 4.3 **Do this for EACH AA file.** It is very likely that many of your genes will *not* start and end with start and stop codons. This is how I like to write my notes, but organize your own however is most helpful to you:

**Nereis pelagica A6054 Mitogenome**

July 19 2021

Geneious - checking through the AA genes for completion.

ND6 - starts with M, does NOT end with asterisk
ND5 - starts with M, does NOT end with asterisk
ND4L - Starts with M, ends with *
ND4 - starts with M, ends with *
ND3 - starts with M, ends with *
ND2 - starts with M, ends with *
ND1 - starts with M, does NOT end with an asterisk
CYTB - starts with M, does NOT end with an asterisk
COX3 - starts with M, ends with *
COX2 - starts with M, ends with *
COX1 - starts with M, does NOT end with an asterisk
ATP8 - stats with M, ends with *
ATP6 - starts with M, ends with *

- 4.4 Now put away these AA and NT files into their corresponding folders within the "Genes" folder you made. For extra clarity, you can select all of them,

Manually Fixing Mitogenome Annotations with MITOS (default)

5

This is the default way you should do mitogenome annotation. If neither MitoFinder nor MITOS found all 15 genes, then you would refer to "Obtaining Reference Mitogenomes for Troubleshooting, if all 15 genes are not present" onward.

6 Importing MITOS results

- 6.1 Hopefully by now, MITOS should have gotten results back to you.

By clicking the links on the left, download the GFF file, FAS file, and the protein plot. Put them in a new corresponding file somewhere that makes sense to you (i.e. a "MITOS" file where you have the FitoFinder file).



MITOS WebServer

1 Enter Input Parameters 2 View Results

Downloads:

[BED file](#)
[GFF file](#)
[TBL file](#)
[Gene Order file](#)
[FAS file](#)
[FAA file](#)

Raw data:

[protein plot](#)
[ncRNA plot](#)
[raw data](#)

Misc:

[Job settings](#)new job | [CREx](#) | [tRNAdb](#) | [Bioinformatik](#)
Jobid: Pilargis_wolfi (USNM1499416_Pilargis_wolfi_MitoFinder_100p.1)

Name	Start	Stop	Strand	Length	Structure
trnL2(tta)	89	155	+	67	svg ps
nad1	174	1079	+	906	
trnI(ata)	1091	1156	+	66	svg ps
trnK(aaa)	1157	1220	+	64	svg ps
nad3	1236	1568	+	333	
trnS1(aga)	1584	1652	+	69	svg ps
nad2	1674	2648	+	975	
cox1	2666	4189	+	1524	
trnN(aac)	4226	4289	+	64	svg ps
cox2	4291	4956	+	666	
trnD(gac)	4983	5048	+	66	svg ps
atp8	5050	5202	+	153	
trnY(tac)	5218	5282	+	65	svg ps
cox3	5290	6063	+	774	
trnQ(caa)	6072	6140	+	69	svg ps
nad6	6151	6600	+	450	
cob	6623	7714	+	1092	
trnW(tga)	7746	7809	+	64	svg ps
atp6	7867	8499	+	633	
trnR(cga)	8517	8576	+	60	svg ps
trnH(cac)	8577	8638	+	62	svg ps
nad5	8678	10348	+	1671	

Scroll to the bottom of the page. If MITOS found anything unusual, it will tell you here.
This is an example:

nad5	8678	10348	+	1671	
trnF(ttc)	10320	10381	+	62	svg ps
trnE(gaa)	10382	10446	+	65	svg ps
trnP(cca)	10447	10510	+	64	svg ps
trnT(aca)	10511	10574	+	64	svg ps
nad4l	10578	10859	+	282	
nad4	10886	12205	+	1320	
trnC(tgc)	12216	12277	+	62	svg ps
trnG(gga)	12278	12339	+	62	svg ps
trnM(atg)	12340	12403	+	64	svg ps
rrnS	12402	13220	+	819	svg ps
trnV(gta)	13215	13278	+	64	svg ps
rrnL	13224	14591	+	1368	svg ps
trnL1(cta)	14542	14603	+	62	svg ps
trnS2(tca)	14604	14670	+	67	svg ps

■ tRNA gene ■ rRNA gene ■ protein coding gene



Warning(s) and peculiarities:

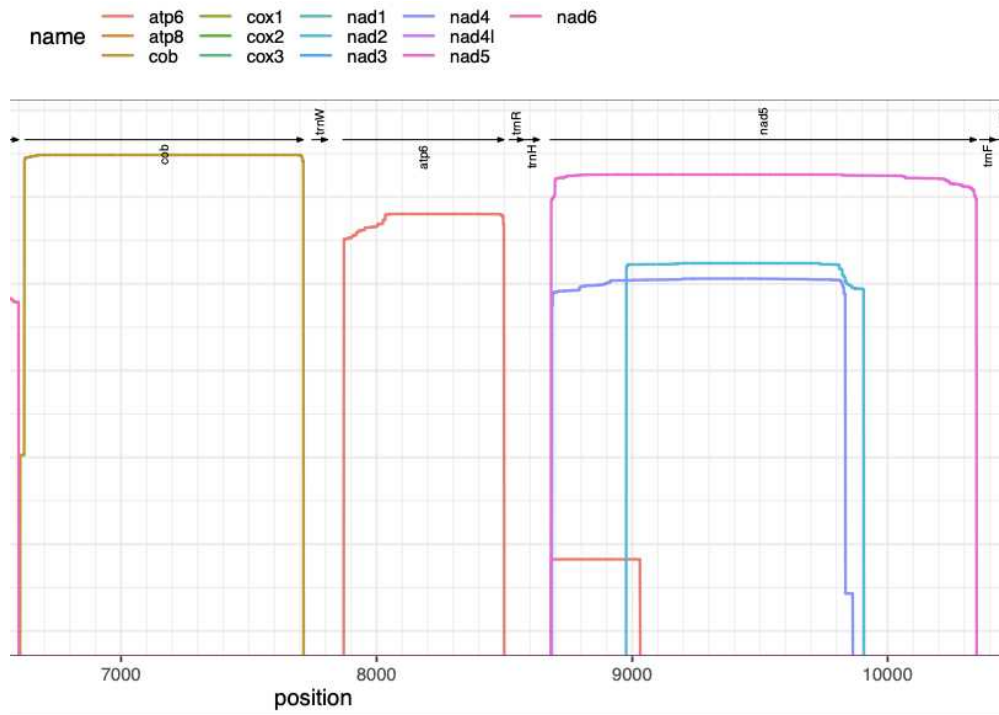
- Genes not found: trnA

Tip(s):

- Consult the protein and ncRNA plots if there is a signal for the genes that MITOS could not determine automatically.

Sometimes MITOS will find 2 or more of a gene, or not find some, and that's when the protein plot comes in handy.

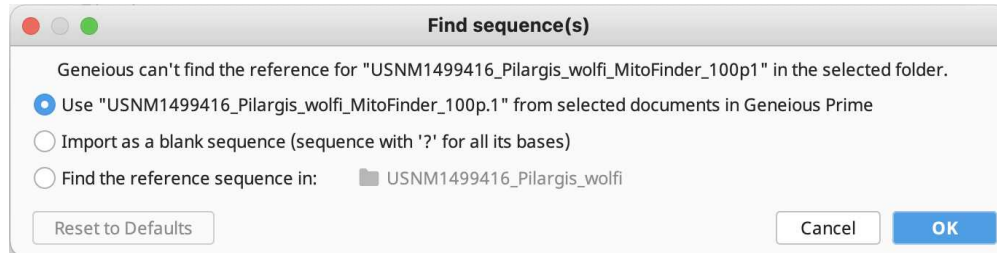
This is an example of a zoomed in protein plot:



Notice, how under the pink line (nad5), there are 3 other lines, orange, blue and purple (atp8, nad2, and nad4). Sometimes, some genes have certain parts that could be similar, but not very, so it still picks up some signal. The line that has the highest plateau is the one you want to trust. Sometimes, when there are say, 2 atp8 genes that MITOS found, what happened is that the low orange line was found somewhere where there was no other strong signal line above it, but it was still just as low as it is here. Then, you would get 2. But if you checked the protein plot, you would know which one is correct - the one that reaches the highest on the graph.

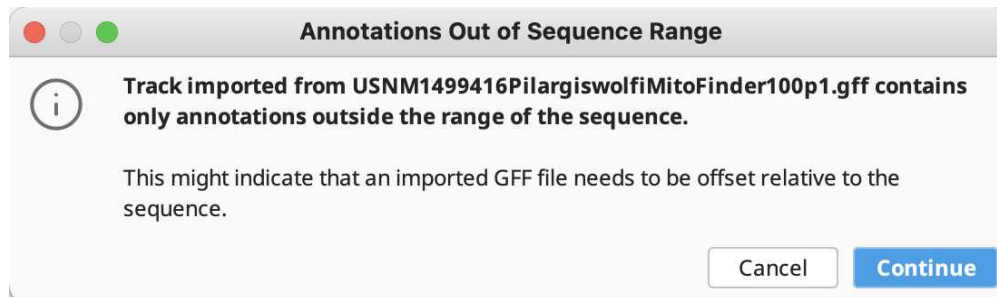
6.2 In Geneious, once again **drag in the SAME "...conting.fasta" file** as before.

6.3 Then, **drag in the MITOS "____.gff" file**.
It might say this:



If so, tell it to use what it found, and click **"OK"**.

It might also say this:



Select **"Continue"**.

This is what your Geneious should now look like:



Name	Modified	Description
USNM1499416_Pilargis_wolff_MitoFinder_100p.1	22 Oct 2022 9:50 am	-
USNM1499416_Pilargis_wolff_MitoFinder_100p.1	20 Oct 2022 11:55 am	-

Sequence View Annotations Text View Lineage Info

Extract R.C. Translate Add/Edit Annotation Allow Editing Annotate & Predict

General

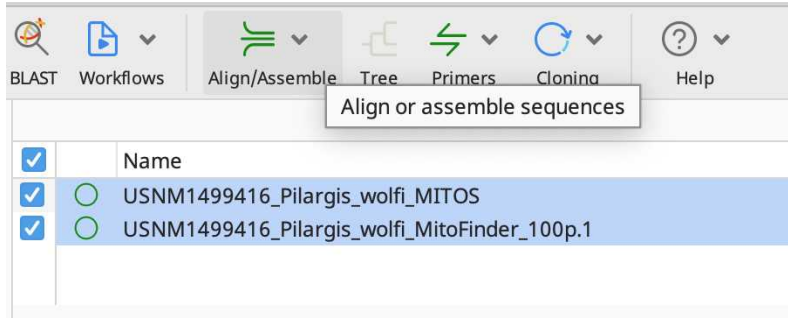
- Colors: A C G T - Edit
- ☒ Graphs Options
- ☒ Annotations Options
- ☐ Complement Options
- ☒ Translation Options
- ☐ Restriction Sites Options
- ☐ Circular Overview
- ☒ Linear View
- ☒ Wrap
- ☒ Show Name
- ☐ Show Description

6.4 **Rename** the sequence so you know it's from MITOS (i.e. change "MitoFinder_100p.1" to "MITOS").

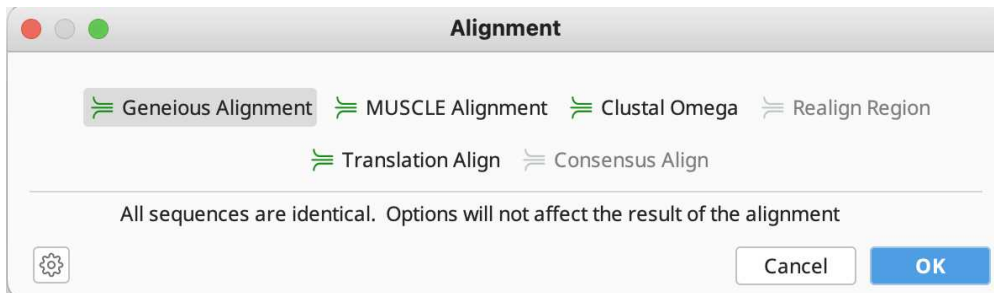
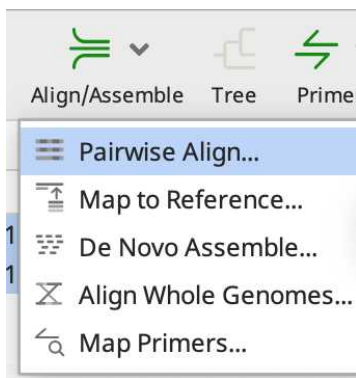
6.5 **Force circularize** for this sequence, too, like before for the MitoFinder sequence.

7 Align the MitoFinder and MITOS sequences for manual annotation and comparison!

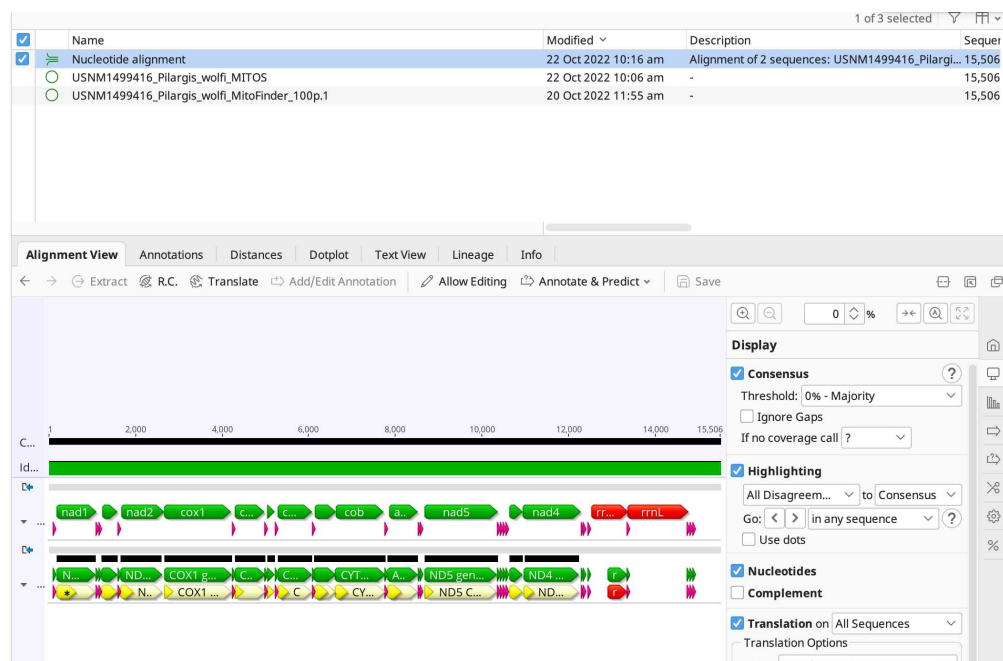
7.1 **Select** both the MitoFinder and MITOS files. Then **click "Align/Assemble"**.



7.2 Choose "**Pairwise Align...**", and choose "**OK**".

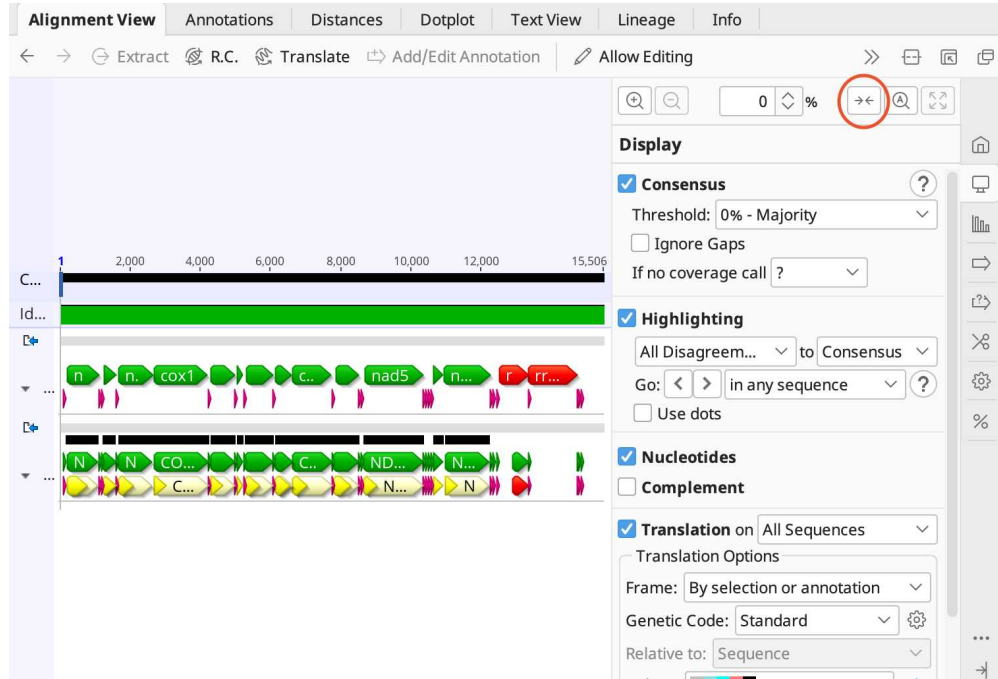


7.3 Your Geneious workspace should look like this:



8 Now, it's time to scroll through the mitogenome and manually annotate it based on comparing it with MITOS.

8.1 Place your cursor at the front of the alignment (click there), and press the zoom in button.

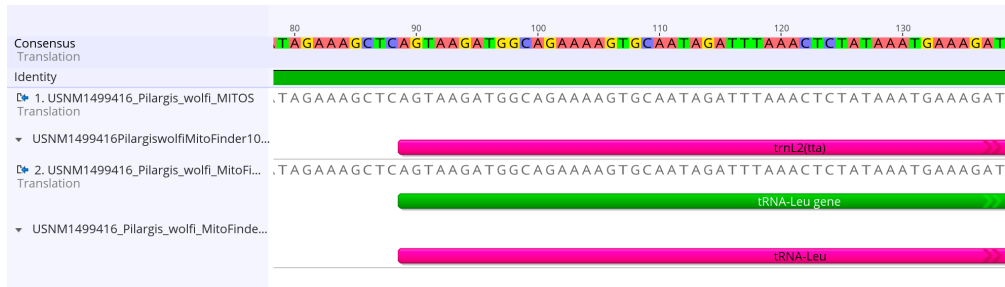


- 8.2 **Scroll to the right** until you hit your first annotation. Once you deal with that one, keep scrolling until the next one, and then the next one, etc. until you get through your entire mitogenome!
- 8.3 If it's a tRNA, just check if the MITOS and MitoFinder annotations are the same (if they start and stop with the same nucleotide).
Also check for overlapping - you don't want them to overlap either with each other or with other genes. Once you make sure the gene a tRNA is overlapping with is correct, shorten the MitoFinder tRNA annotation so they don't overlap, by clicking and dragging the pink and green bars (each separately) until they don't.



Note

NOTE: You should have and tRNA-Leu1 and tRNA-Leu2, and a tRNA-Ser1 and tRNA-Ser2. When you are interested in gene order, you **write down 1 or 2 based on MITOS, not MitoFinder!** MITOS has historically been more correct with this.



So in this case, this gene is tRNA-Leu2.

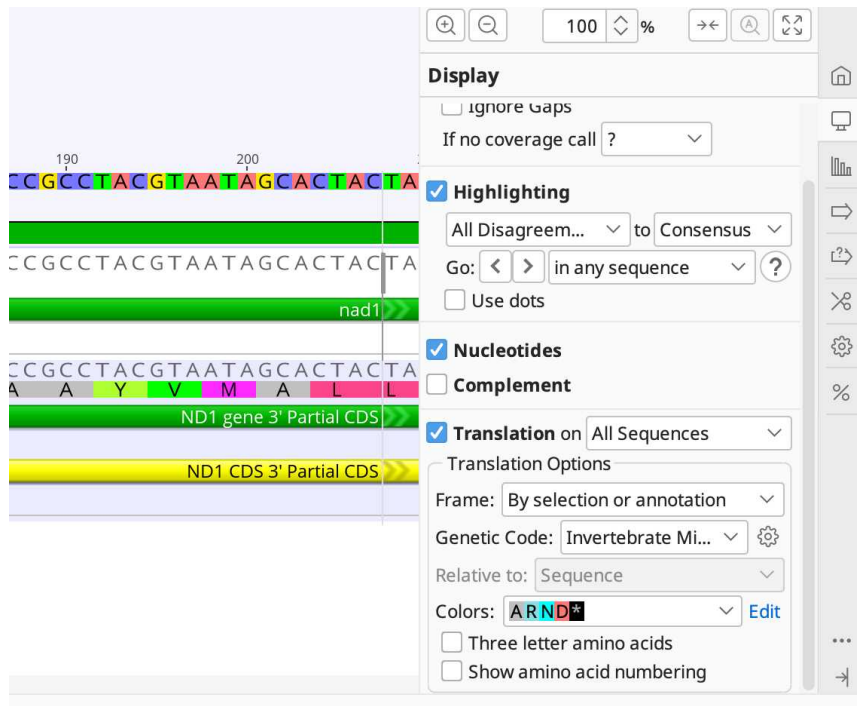
8.4 If it's a protein coding gene or rRNA, stop before the beginning. These are the genes you notes whether they start with an M (the start codon) and end with an asterisk (the stop codon).

First, I like to turn on the Translation - it helps me visually.

1. Go to the second right sidebar tab that looks like a screen

2. Check the box by "Translation"

3. MAKE SURE to change the "Genetic Code" to whatever is relevant for you - in our lab, that's mostly "Invertebrate Mitochondrial"



You can check the translation of a part of sequence by clicking where you want to start checking and then dragging to where you want to see - but **MAKE SURE you drag from left to right! Drag direction matters!**

Note

HOWEVER - **just because you have the translation, doesn't mean you can blindly trust the colors** - sometimes some codons can be interpreted as an M or something else, and if it's at the beginning of the annotated gene, Geneious likes to show it as an M, even if you wrote down that the gene does NOT start with an M.

- 8.5 For annotation, **you are changing ONLY the MitoFinder sequence**, but you are looking at MITOS to help you.

A guiding principle is that **you want MINIMAL OVERLAP between genes** - if you can entirely avoid it, do. (But at least in annelids, **ND4 and ND4L will always overlap ~10 bp!**)

To change a MitoFinder annotation, click on the front or back end of the specific green gene bar, and **drag it** to where you want it (this may take a bit of practice to click correctly for dragging and not just selecting). Then, do the same for the specific yellow

CDS bar (coding sequence) or red rrna bar (16S and 12S), and drag that to where you want it, as well.

Note

BE EXTREMELY CAREFUL NOT TO ACCIDENTALLY ADD A STOP CODON ANYWHERE in the middle of a gene! This is the bit that the "Translation" function really helps me with.

- 8.6 At the **beginning** of a gene, look for **the start codon M** - the corresponding nucleotides are **ATA** or **ATG**.

(Rarely, it could be another set of nucleotides. For Nephytidae, I've run into COI starting with GTG instead).

a.) If the gene starts with an M and the MITOS annotation is either the same, or starts further to the right, keep the MitoFinder annotation as is.

b.) If the gene starts with an M but the MITOS annotation is further to the left and either the MITOS annotation starts with an M or there is an M even further to the left that does not overlap with other genes, change the MitoFinder annotation to whatever the furthest to the left non-overlapping start codon (M) would be.

c.) If the gene does not start with an M, but MITOS' annotation does start with an M, either to the right or to the left, change the MitoFinder annotation to what MITOS found.

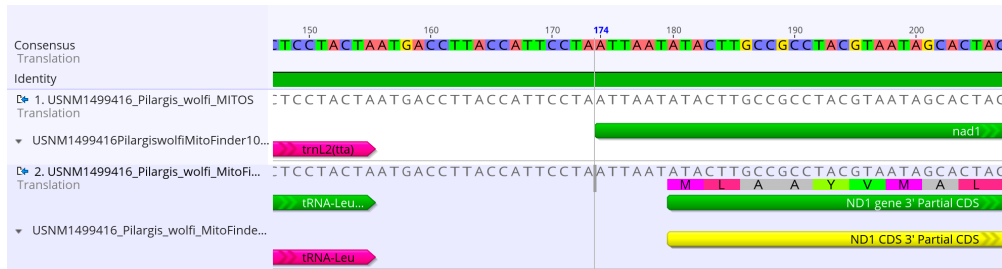
d.) If neither MitoFinder nor MITOS annotations start with an M: search for one. Start with the left and go to the right if no luck on the left. Compare gene length to other known genes of the taxonomic group (i.e. if a gene is usually 1500 bp, you'll know something is weird if your gene has 800 bp.).

Below are some examples of editing the beginning of a gene. Feel free to add your own if you don't think they fit into one of the already added examples!

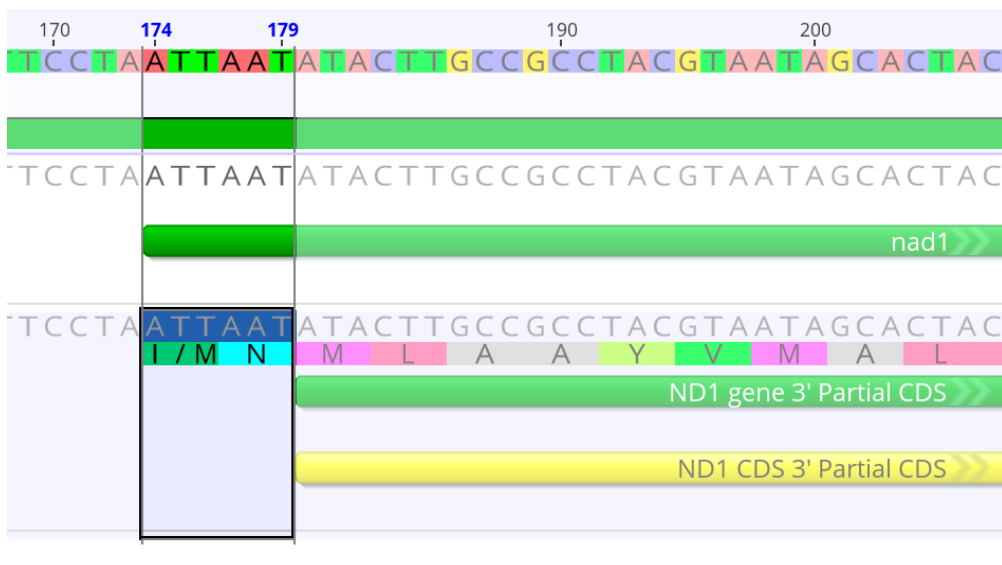
Note

Example 1: Starts with an M, but MITOS found more that doesn't start with an M.

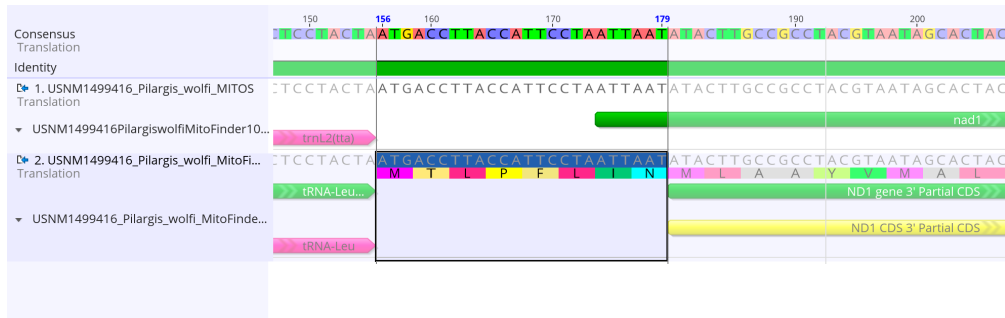
Here, I've run into ND1. It **starts with an M**.



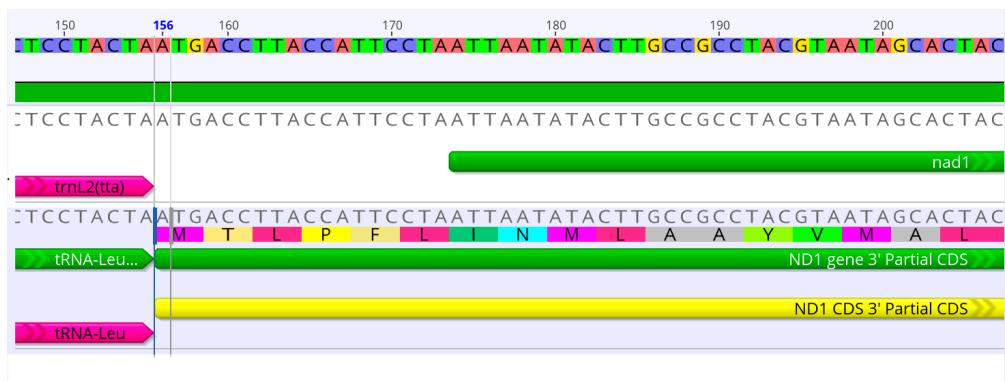
However, **MITOS found more of the gene further back to the left**. But, if I click at the beginning of the MITOS annotation in the mitofinder track and drag it to the right to the beginning of the MitoFinder annotation, I can see that **it does NOT start with an M** (even though Geneious says I/M, it is not ATA or ATG).



But, if I start checking before then, I can see that there is a start codon right after the end of the previous gene (in this case, a tRNA). So, I will change it to that! But *only because MITOS found more gene to the left*.



This is what my ND1 now looks like:



8.7 At the **end** of a gene, look for the **stop codon *** - the corresponding nucleotides are **TAA** or **TAG**.

9

10

Obtaining Reference Mitogenomes for Troubleshooting, if all 15 genes are not present

11 **Obtaining Reference Mitogenomes:** When manually fixing genes, we help ourselves with reference mitogenomes. We want to import and simultaneously look at our new mitogenome and already published mitogenomes, because those help us make informed decisions when it comes to expanding or shortening a gene.

- 11.1 Go to **GenBank** and search for the mitogenomes of closely related species by typing in the following:

Command

Search for mitogenomes of related species

"SampleFamily" & "mitogenome"

Make sure to replace "**SampleFamily**" with whatever your family of organisms is, so you would search for i.e. " "**Pilargidae**" & "**mitogenome**" ". Include the quotation marks, because that means GenBank will show you only exact matches.

- 11.2 If your search gives no results, try the searching by your sample's suborder or order, etc. until you get results. You can use **WoRMS** to help you.
- 11.3 Once you have the results, select all of them by clicking the boxes before them and choose "**Send to**", select "**Gene Features**", keep the default "**FASTA Nucleotide**" as the format, and click "**Create File**".



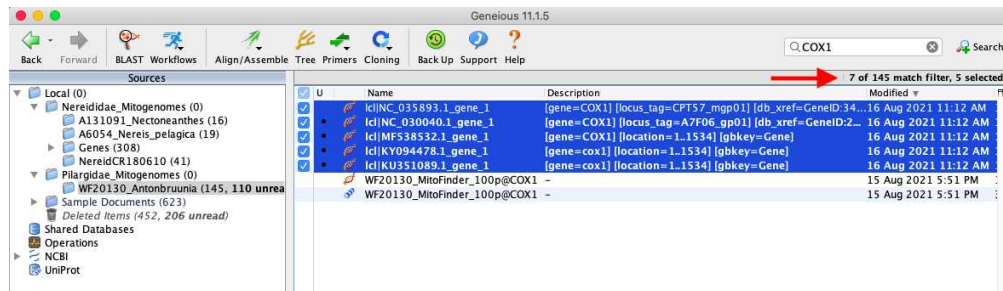
11.4 **Rename** the file accordingly and move it into your sample's folder. Then, **drag it into Geneious**, again keeping the sequences **separate**. A lot of new NT files should appear in your workspace, making it look like this:

1 of 135 selected				
U	Name	Description	Modified	
•	lcl NC_035893.1_gene_37	[locus_tag=CPT57_mgt22] [db_xref=GenID:34568765] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_36	[gene=ND2] [locus_tag=CPT57_mgp02] [db_xref=GenID:3...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_35	[locus_tag=CPT57_mgt21] [db_xref=GenID:34568766] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_34	[gene=ND3] [locus_tag=CPT57_mgp03] [db_xref=GenID:3...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_33	[locus_tag=CPT57_mgt20] [db_xref=GenID:34568792] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_32	[locus_tag=CPT57_mgt19] [db_xref=GenID:34568764] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_31	[gene=ND1] [locus_tag=CPT57_mgp04] [db_xref=GenID:3...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_30	[locus_tag=CPT57_mgt18] [db_xref=GenID:34568790] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_29	[locus_tag=CPT57_mgt17] [db_xref=GenID:34568789] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_28	[locus_tag=CPT57_mgt16] [db_xref=GenID:34568788] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_27	[locus_tag=CPT57_mgt15] [db_xref=GenID:34568787] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_26	[locus_tag=CPT57_mgt01] [db_xref=GenID:34568786] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_25	[locus_tag=CPT57_mgt14] [db_xref=GenID:34568785] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_24	[locus_tag=CPT57_mgt02] [db_xref=GenID:34568767] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_23	[gene=ND4] [locus_tag=CPT57_mgp05] [db_xref=GenID:3...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_22	[gene=ND4L] [locus_tag=CPT57_mgp06] [db_xref=GenID:...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_21	[locus_tag=CPT57_mgt13] [db_xref=GenID:34568783] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_20	[locus_tag=CPT57_mgt12] [db_xref=GenID:34568782] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_19	[locus_tag=CPT57_mgt11] [db_xref=GenID:34568781] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_18	[locus_tag=CPT57_mgt10] [db_xref=GenID:34568769] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_17	[gene=ND5] [locus_tag=CPT57_mgp07] [db_xref=GenID:3...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_16	[locus_tag=CPT57_mgt09] [db_xref=GenID:34568779] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_15	[locus_tag=CPT57_mgt08] [db_xref=GenID:34568758] [lo...	16 Aug 2021 11:12 AM	
•	lcl NC_035893.1_gene_14	[gene=ATP6] [locus_tag=CPT57_mgt08] [db_xref=GenID:...	16 Aug 2021 11:12 AM	

- 11.5 Sometimes, the uploaded mitogenomes won't have **rrnS/12S** and **rrnL/16S** included in the FASTA download. In those cases, you'll have to specifically search for both rrnS and rrnL FASTA files on GenBank of each of the mitogenomes you imported. Download, rename, and drag in those files, too, if you don't already have them.

Manually Checking and Fixing Protein Coding Genes when all 15 genes are not present

- 12 The first pass check gave us a general idea of how complete our genes are. Now, we want to carefully check each one (**even those that *do* start and stop with start and stop codons!**) and fix those that need fixing.
- 12.1 **Aligning the Sequences:** Start by **typing "COX1" into the search bar** in Geneious. Then, select all of the COI sequences that were uploaded from GenBank. You want to make sure those are the only sequences selected - you can check this by looking at the upper right hand corner below the search bar (I've noted it with the arrow) - the number of selected sequences should be the same as the number of imported mitogenomes.



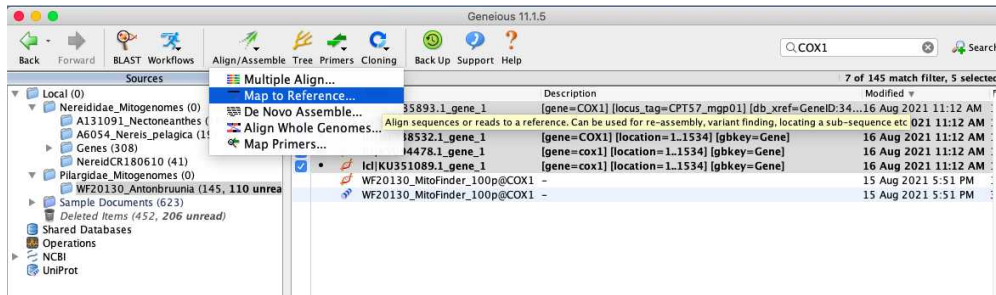
Note

Sometimes, uploaders to GenBank will use COI/CO1 instead of COX1 as the gene name. If you're not getting the right amount of genes given the amount of mitogenomes you uploaded for reference, try spelling it differently. Other examples of differently spelled genes are:

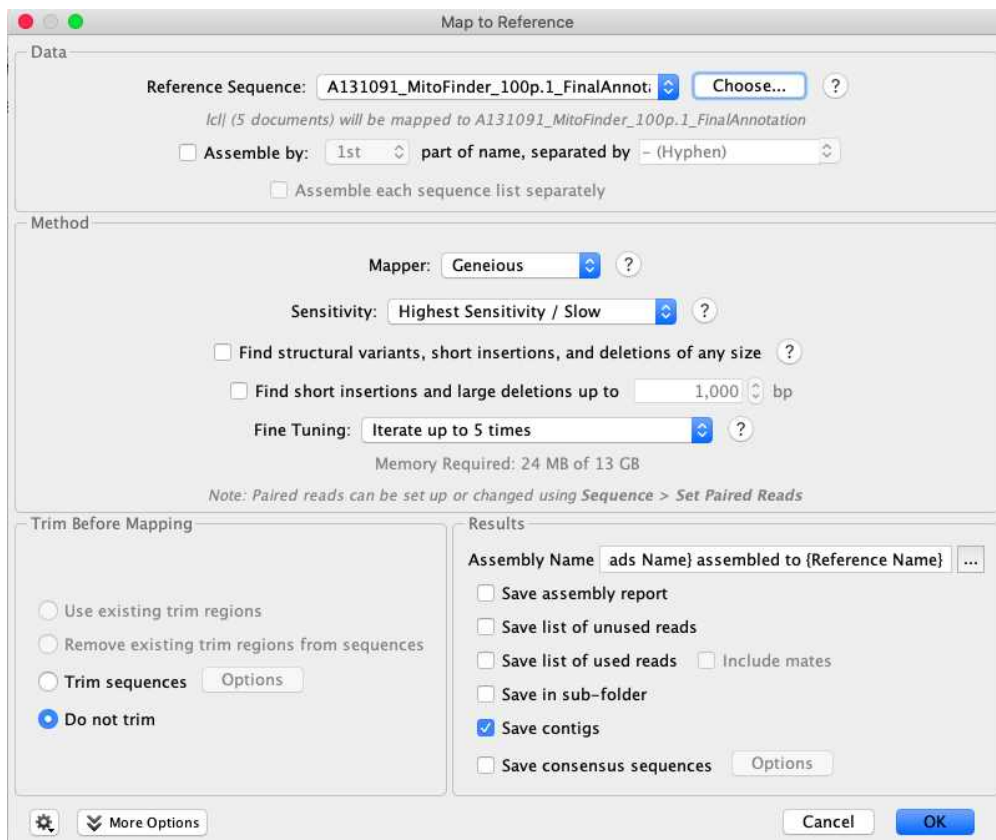
- NAD1-6 instead of ND1-6
- COB instead of CYTB

When aligning these, make sure when you change the search from say ND5 to NAD5, you click directly on the box where the check mark shows up to let you know you've selected the sequence. This keeps all of the sequences you've selected so far selected, even if you're in a different search result. Then, click the 'x' in the search bar, and proceed with aligning. If you click anywhere else when selecting a sequence, it will select *only that specific sequence* and deselect the rest.

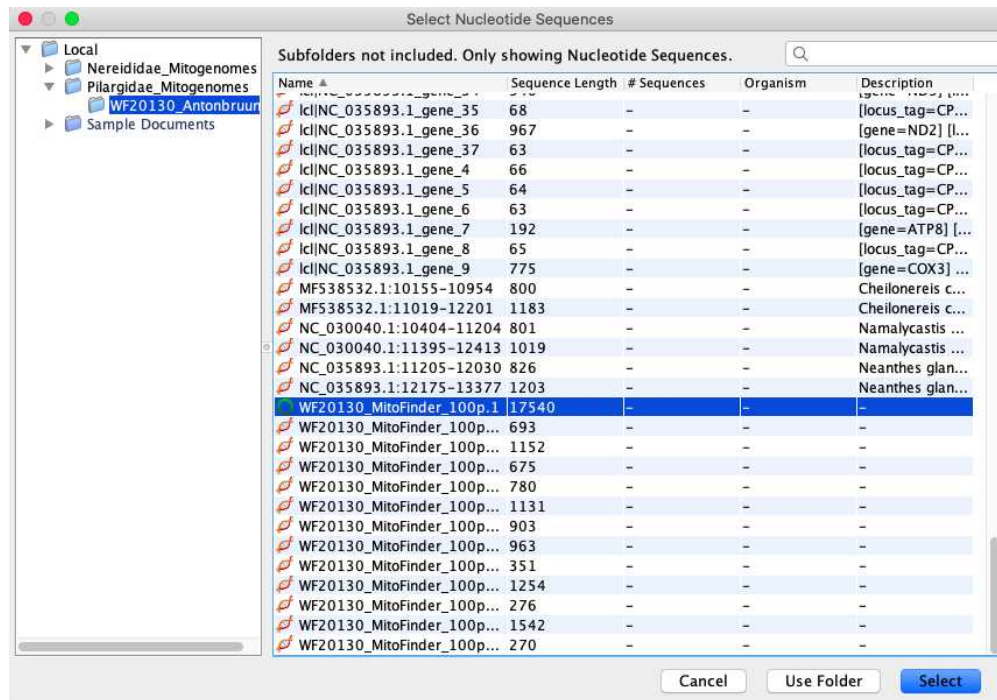
12.2 Aligning the Sequences: Then click "Align/Assemble" in the tool bar, and choose "Map to Reference..."



This window will pop up:



12.3 Aligning the Sequences: To specify the reference sequence, which we want to be your sample, click "Choose..." and navigate to your circularized file (green circle icon), then click "Select".

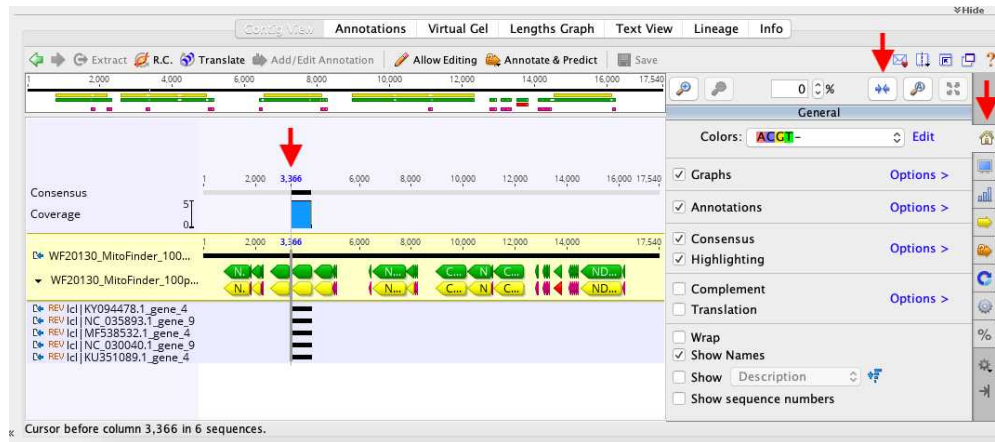


12.4 Aligning the Sequences: The rest of the default options are fine, so click "OK". Once Geneious is done, clear the search, and at the top of your files, you should see an alignment (orange icon of three parallel lines). Your workspace should look like this:

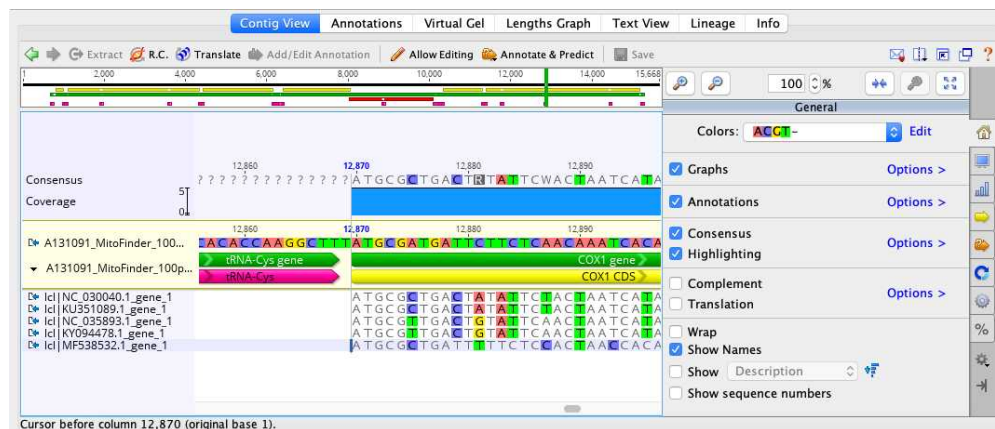
The screenshot displays the Geneious 11.1.5 interface. The top panel shows a list of sources on the left and a table of sequences in the center. The bottom panel shows a detailed view of a sequence alignment with annotations, a virtual gel, and a text view. The alignment shows a consensus sequence with coverage and various annotations including gene names and coordinates.

In the bottom part of Geneious, you now have the alignment. The black bars are where the GenBank sequences aligned with your sequences that MitoFinder found, which are the green and yellow bars.

- 12.5 **Check the Beginning of the Gene:** First zoom in to the front of your gene. Do this by clicking on the **house icon** on the bottom right side of Geneious. Then, **place your cursor** close to the beginning of the gene you are looking at and click on the **two blue arrows** pointing towards each other. You can zoom out by clicking on the same icon.



This will give you such a view:



The invertebrate start codons are **ATG** and **ATA**. However, just because MitoFinder decided your gene starts with a start codon, this **DOES NOT MEAN** that this is the **FIRST** start codon - make sure to **check to the left of the gene** to see if there are any more ATG/ATA codons. Take care to go by **three nucleotides at a time**, otherwise you shift the frame of translation.

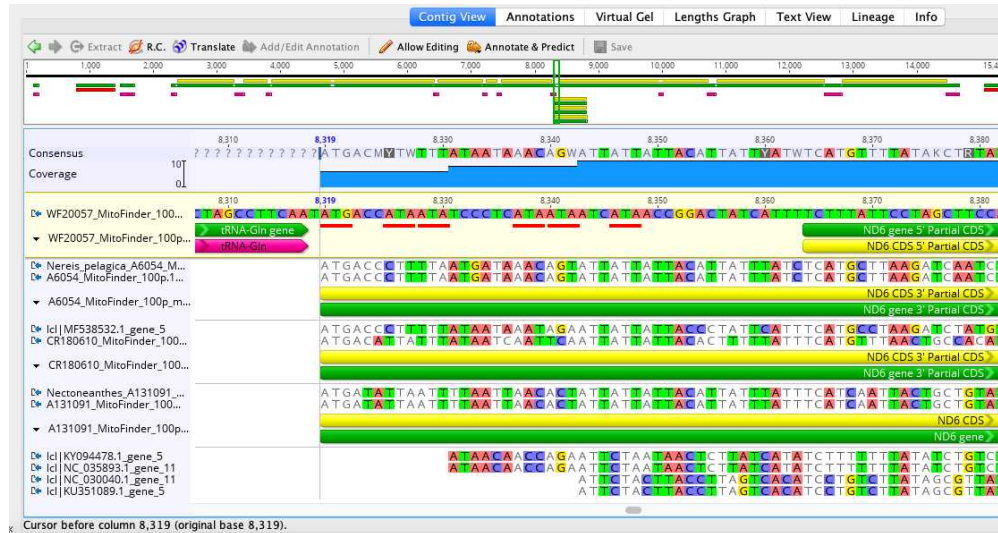
- 12.6 **Editing the Beginning:** If you found that the current way MitoFinder annotated the beginning is incorrect (either no start codon or not the right one), click on the green and yellow bars and **drag** them one at a time to the nucleotide they should really start with. **Make sure you extend/shorten the gene only by a number of nucleotides divisible by three!**

Note

For later reference, I like having "before" and "after" screenshots of the sequences, so I'll definitely know what I changed where, even months/years later.

12.7 Editing the Beginning Example:

MitoFinder didn't find the start codon. When I zoomed into the beginning of the gene, I began to look for start codons. I actually found *six* potential start codons (underlined red)! However, the one farthest to the left is the correct option.



Note

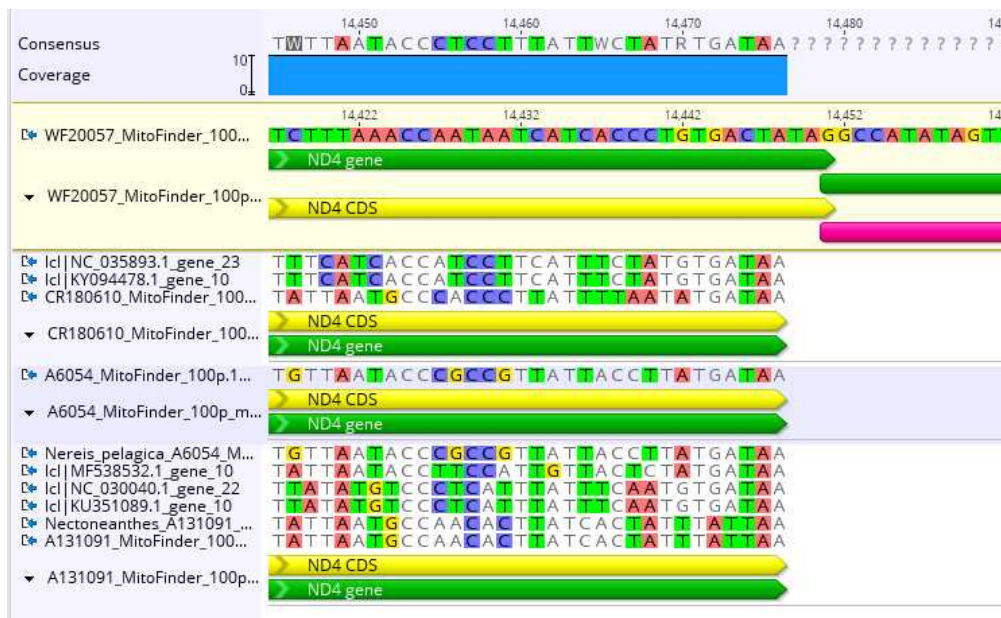
Note: The top few sequences I aligned look different than the bottom few because the bottom ones are from GenBank and the top ones with the graphic representations of genes are from Geneious - sequences I have already manually checked and annotated.

I then dragged the gene annotations so they now look like this when corrected:



I made sure the amount of nucleotides I added was divisible by three.

- 12.8 Check the End of the Gene:** Either zoom out and place your cursor at the end of the gene and zoom in again, or just scroll to the end of the gene. The stop codons you are looking for are **TAA** or **TAG**. Sometimes, when MitoFinder finds the stop codon, your sequence will end like the others you are aligning it to, and other times it won't. However, it can still be correct, like this:



There's no obvious way to change this sequence, it doesn't align with the other reference sequences, but there is a stop codon, so I left it as is.

- 12.9 **Editing the End:** If you found that the current way MitoFinder annotated the end is incorrect (no stop codon), click on the green and yellow bars and **drag** them one at a time to the nucleotide they should really end with. **Make sure you extend/shorten the gene by a number of nucleotides divisible by three!**

Note

For later reference, I like having "before" and "after" screenshots of the sequences, so I'll definitely know what I changed where, even months/years later.

- 12.10 **Editing the End Example:**
MitoFinder didn't find the stop codon. When I zoomed into the end of the gene, I found two:

HOWEVER! Stop codons are different than start codons. There can't be a stop codon somewhere within the gene, so even if there is another potential stop codon farther to the right, that is not a stop codon for this gene. You have to choose the stop codon closest to the end of the gene you're fixing. Therefore, I edited this gene like this:

I made sure I extended the gene by a number of nucleotides divisible by three - in this case 45.

- 12.11 **I Can't Find a Stop Codon?**
AA residue thing

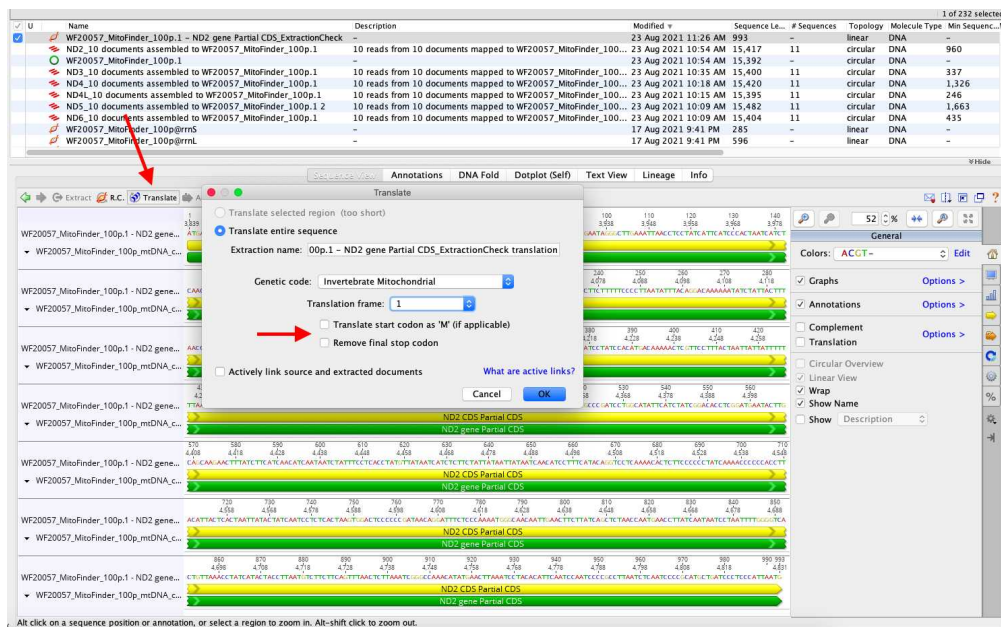
- 12.12 **Checking the Validity of Edits:** After you've edited the annotations of a protein coding gene, there is one more way to check that you have done so correctly. This makes sure you've edited the nucleotides comprising the gene in numbers divisible by three, and also ensures you didn't accidentally add any extra stop codons anywhere.

First, click on the green annotation (the gene), and then click '**Extract**'.



You can add something to the end of the name, i.e. "...AnnotationCheck" so that you know where the sequence came from later on, then click **"OK"**.

12.13 Checking the Validity of Edits: Then click on the sequence it gives you as an output and click on the **"Translate"** button. Change the Genetic code to **"Invertebrate Mitochondrial"**, leave the translation frame as 1, and make sure to **UNCHECK both of the checked boxes** - we don't want to force Geneious to translate the first three nucleotides as a start codon and we don't want it to remove the final stop codon.





13