



Jun 11, 2025

Mitogenome (or other Multi-gene) Maximum Likelihood Phylogenetic Analyses

DOI

dx.doi.org/10.17504/protocols.io.bp2l613kzvqe/v1

Dakota Betz¹

¹ucsd

Rouse Lab



Sonja Huc

UNCW

OPEN  ACCESS



DOI: dx.doi.org/10.17504/protocols.io.bp2l613kzvqe/v1

Protocol Citation: Dakota Betz 2025. Mitogenome (or other Multi-gene) Maximum Likelihood Phylogenetic Analyses .
protocols.io <https://dx.doi.org/10.17504/protocols.io.bp2l613kzvqe/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: April 28, 2022

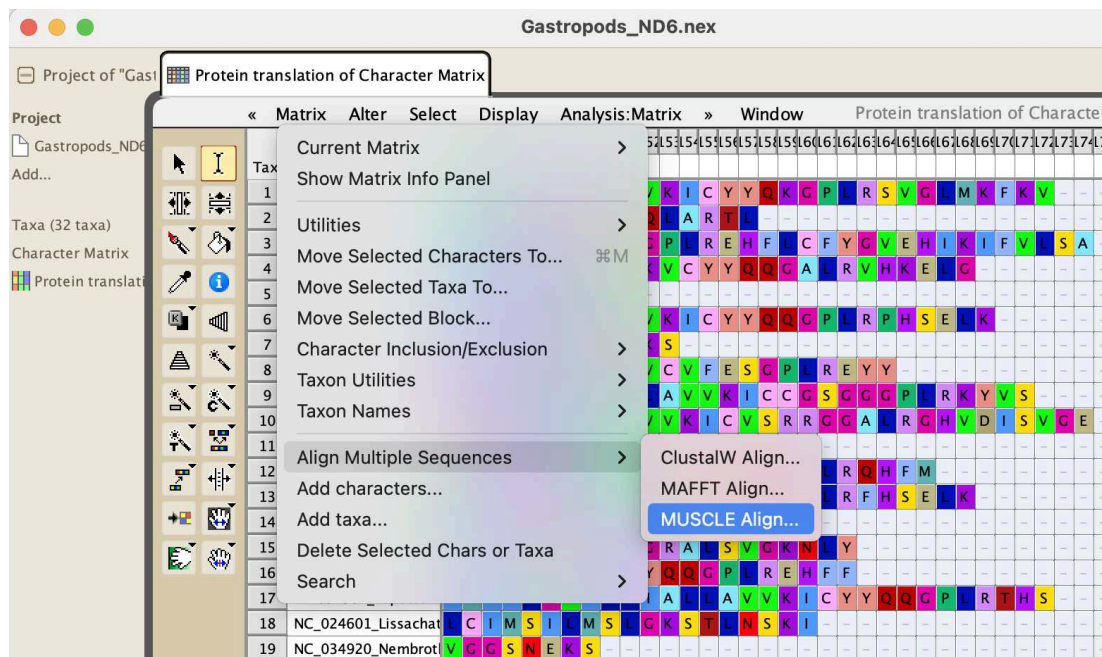
Last Modified: June 11, 2025

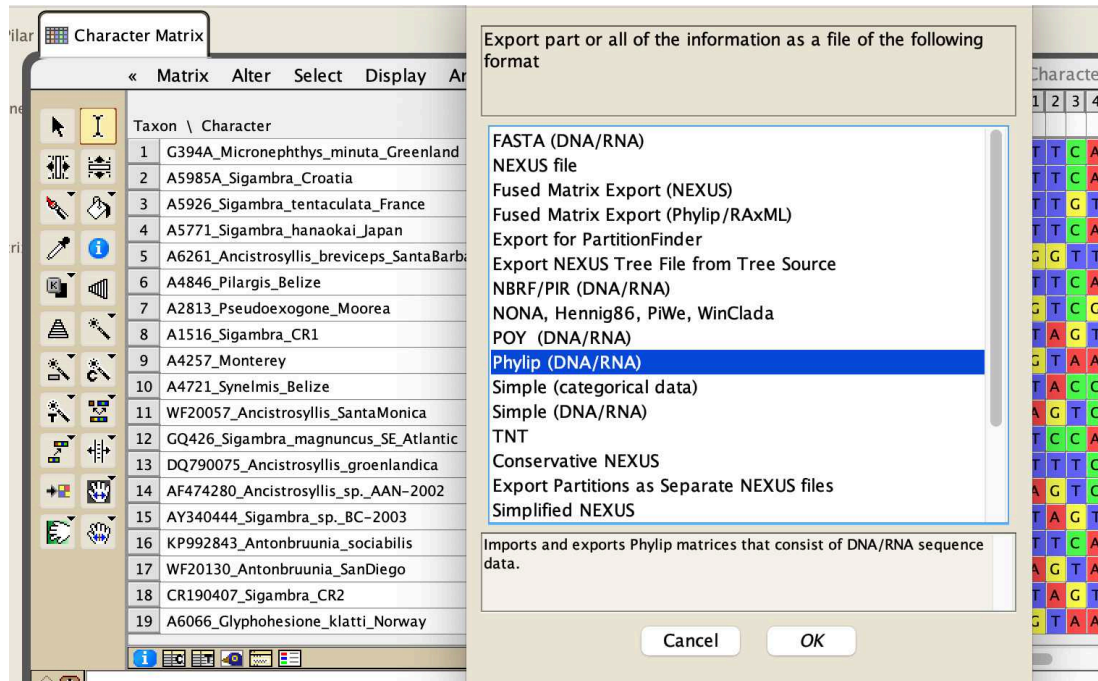
Protocol Integer ID: 61648

Abstract

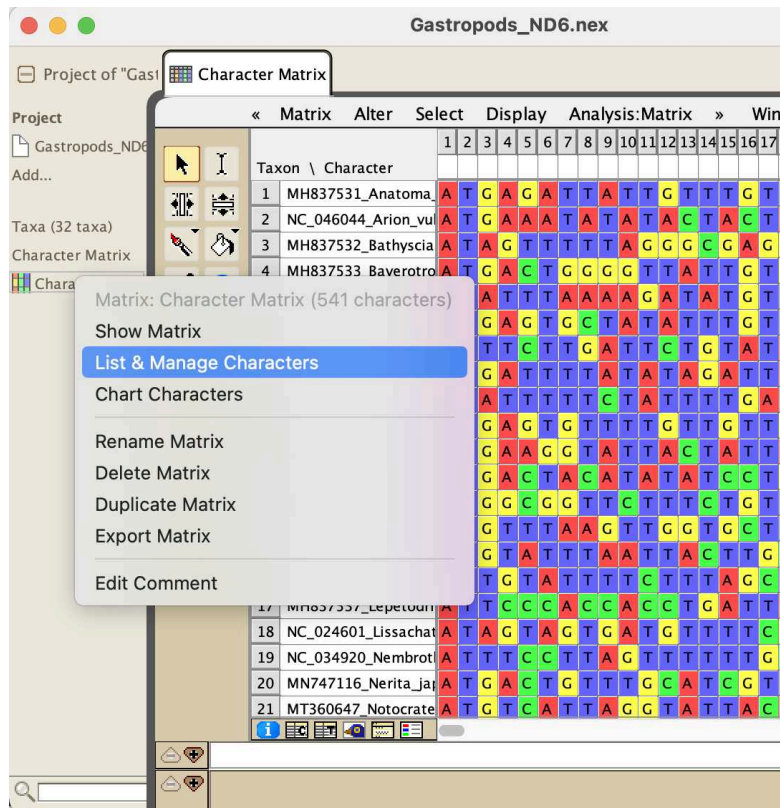
A guide to doing maximum likelihood whole mitogenome (or multigene) analyses in RaXML. This was written after Avery's instructions!

- 1 Create text files in FASTA format with all of each of the genes you want to use (i.e. Pilargidae_12S.txt, Gastropods_ATP6.txt, etc.). Make sure the same organism has the same name for ALL of the genes (spaces included - there should be none, but sometimes they sneak up on you).
- 2 Open Mesquite, create a new file, make sure to check "Make Character Matrix" and then select "DNA Data" as the type of data. Drag any one of your gene text files into it and import as FASTA, click "Don't Adjust". Delete the top empty taxa that were there by default.
- 3 For non-protein-coding genes (12S, 16S) or if you're doing a nucleotide only analysis, align the sequences using MUSCLE (Click on "Matrix", "Align Multiple Sequences", then "MUSCLE Align". Select "No" for running on multiple threads, and then click okay. Check to see if everything is fine.), then export in phylip (DNA/RNA) format. Repeat this for all non-protein-coding genes, or all genes for a nucleotide only analysis.

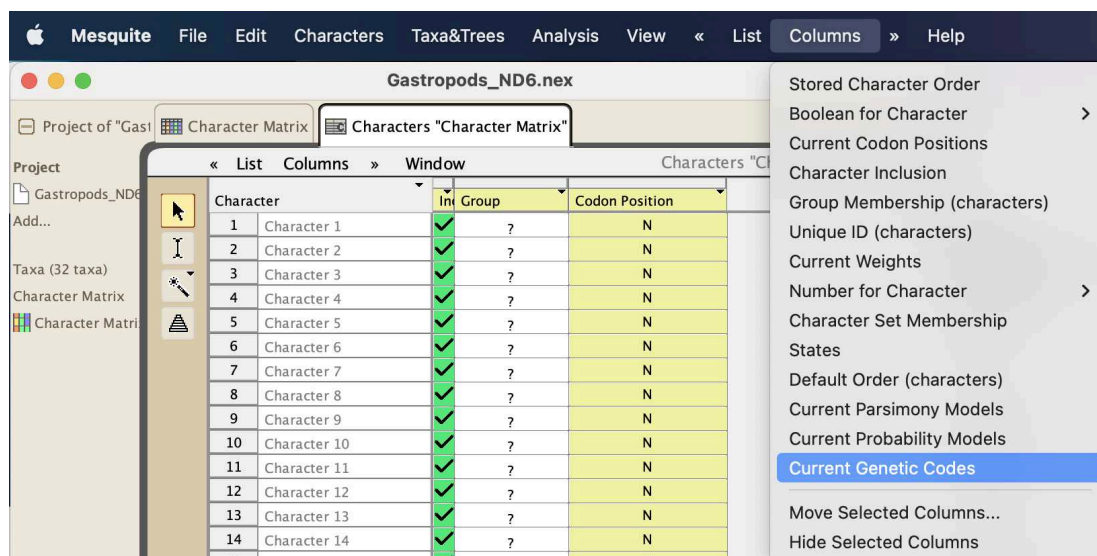




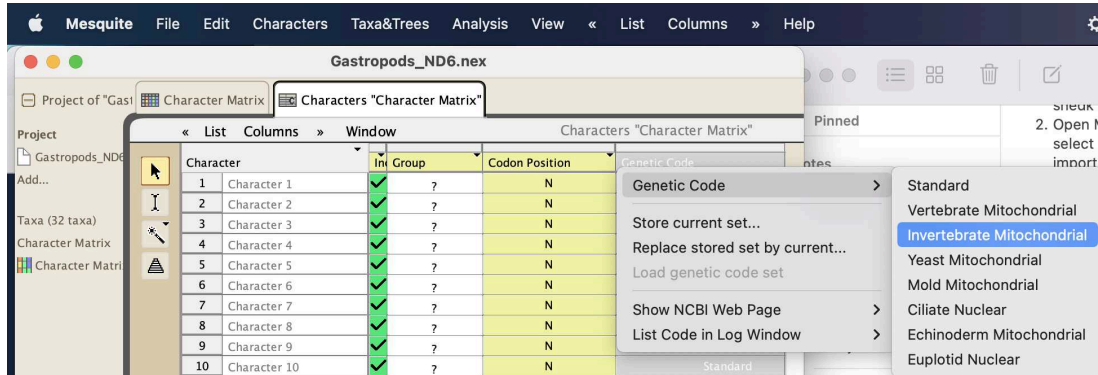
- 4 If you're doing an amino acid only or nucleotide for non-protein-coding genes and amino acid for protein-coding genes analysis, follow these next steps:
 - 4.1 Create a nexus file like before, DO NOT yet align it, save it, then duplicate it. Name the duplicated file something like "Gastropods_ATP6_NT.nex". The file you have open can stay "Gastropods_ATP6.nex" or be something like "Gastropods_ATP6_AA.nex". As long as you know which one will be the nucleotide only and which one will be the protein only file.
 - 4.2 2. Click on "Character Matrix" on the left sidebar and select "List & Manage Characters"



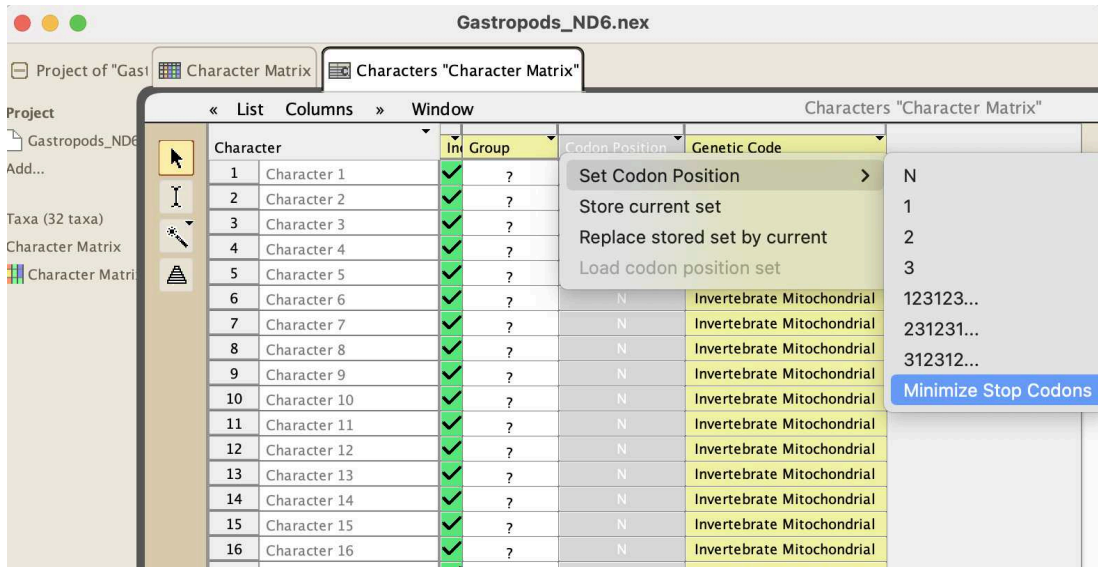
4.3 In the toolbar, select "Columns", then "Current Genetic Codes".



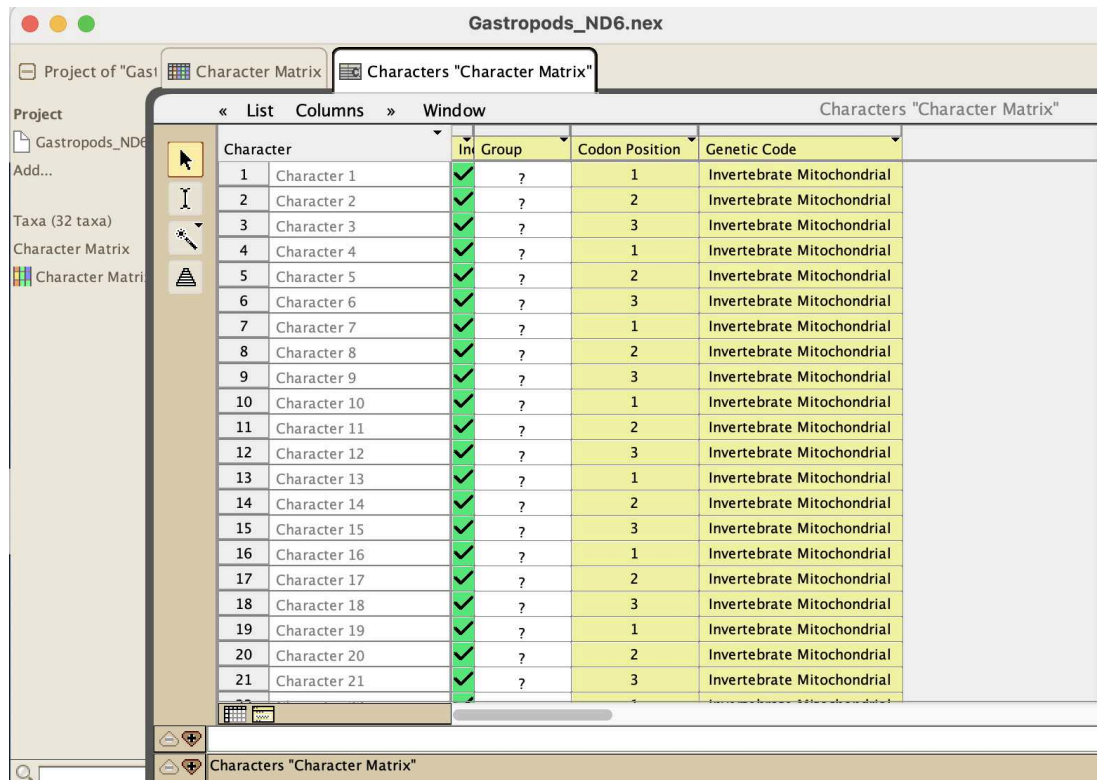
- 4.4 Click on the "Genetic Code" tab in your file window, then under "Genetic Code" choose Invertebrate Mitochondrial (or whatever applies, but in our lab it should mostly be that one). Make sure you actually selected the column, so the values all change from displaying "Standard" to "Invertebrate Mitochondrial".



- 4.5 Click on "Codon Position", select "Minimize Stop Codons". When you do this, MAKE SURE that the column is selected! The values should change from "N" to numbers 1-3, depending on the position.



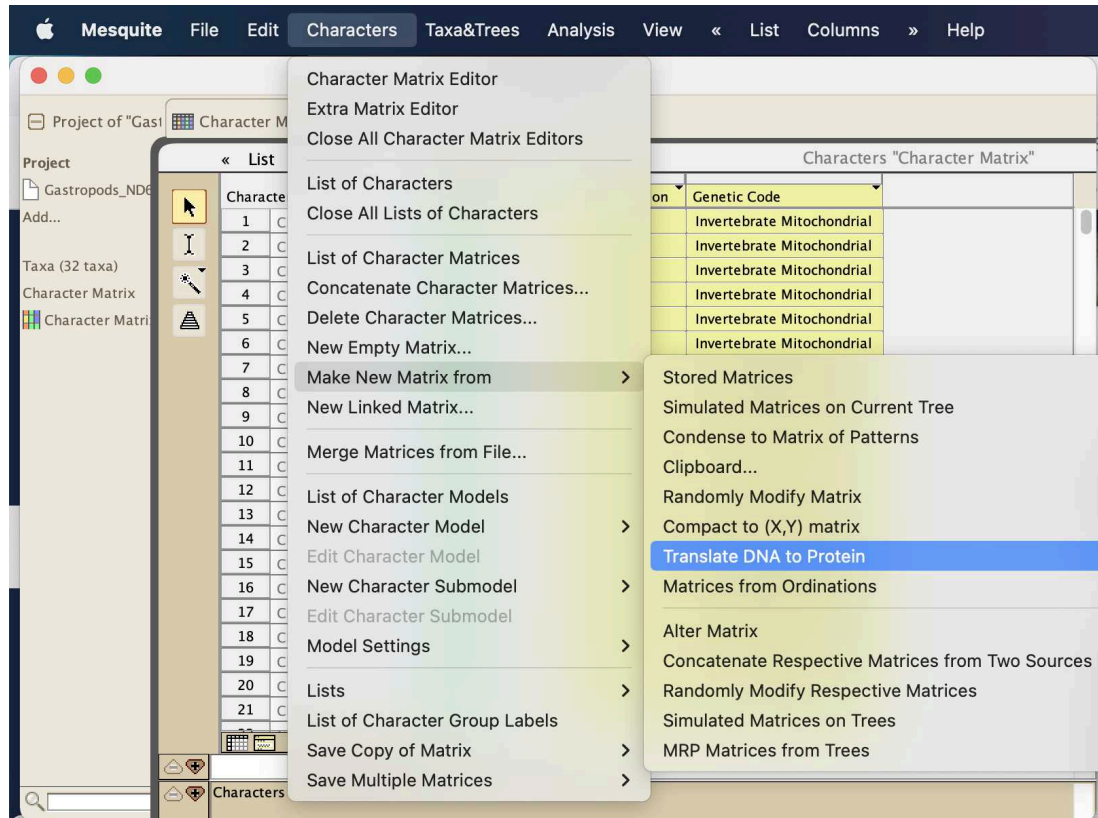
Your window should look like this:



The screenshot shows the 'Gastropods_ND6.nex' application window. The 'Characters "Character Matrix"' tab is active, displaying a table with 21 rows and 5 columns. The columns are 'Character', 'Int', 'Group', 'Codon Position', and 'Genetic Code'. The 'Int' column contains green checkmarks, 'Group' contains question marks, 'Codon Position' contains numbers 1, 2, or 3, and 'Genetic Code' contains 'Invertebrate Mitochondrial'.

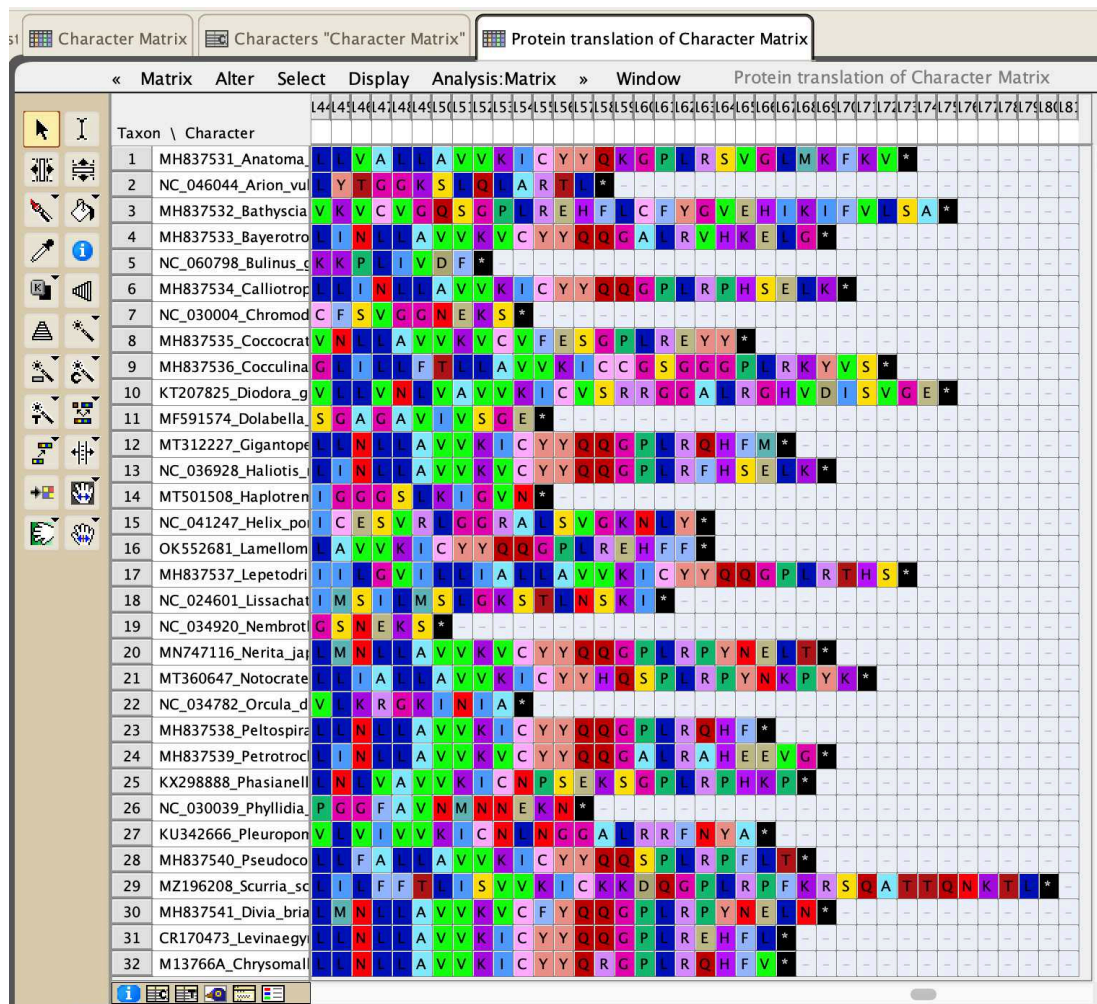
Character	Int	Group	Codon Position	Genetic Code
1 Character 1	✓	?	1	Invertebrate Mitochondrial
2 Character 2	✓	?	2	Invertebrate Mitochondrial
3 Character 3	✓	?	3	Invertebrate Mitochondrial
4 Character 4	✓	?	1	Invertebrate Mitochondrial
5 Character 5	✓	?	2	Invertebrate Mitochondrial
6 Character 6	✓	?	3	Invertebrate Mitochondrial
7 Character 7	✓	?	1	Invertebrate Mitochondrial
8 Character 8	✓	?	2	Invertebrate Mitochondrial
9 Character 9	✓	?	3	Invertebrate Mitochondrial
10 Character 10	✓	?	1	Invertebrate Mitochondrial
11 Character 11	✓	?	2	Invertebrate Mitochondrial
12 Character 12	✓	?	3	Invertebrate Mitochondrial
13 Character 13	✓	?	1	Invertebrate Mitochondrial
14 Character 14	✓	?	2	Invertebrate Mitochondrial
15 Character 15	✓	?	3	Invertebrate Mitochondrial
16 Character 16	✓	?	1	Invertebrate Mitochondrial
17 Character 17	✓	?	2	Invertebrate Mitochondrial
18 Character 18	✓	?	3	Invertebrate Mitochondrial
19 Character 19	✓	?	1	Invertebrate Mitochondrial
20 Character 20	✓	?	2	Invertebrate Mitochondrial
21 Character 21	✓	?	3	Invertebrate Mitochondrial

- 4.6 In the toolbar, go to "Characters", select "Make New Matrix From", and then "Translate DNA to Protein". Click "OK". This should open a new tab with your data displayed as amino acids. You should only have stop codons (black with asterisks) at the ends of your sequences if they are complete, but sometimes if the sequence doesn't start on the first codon, there may be more stop codons throughout. This only means this one sequence needs to be removed, not that all of the others are wrong, too. They're fine!

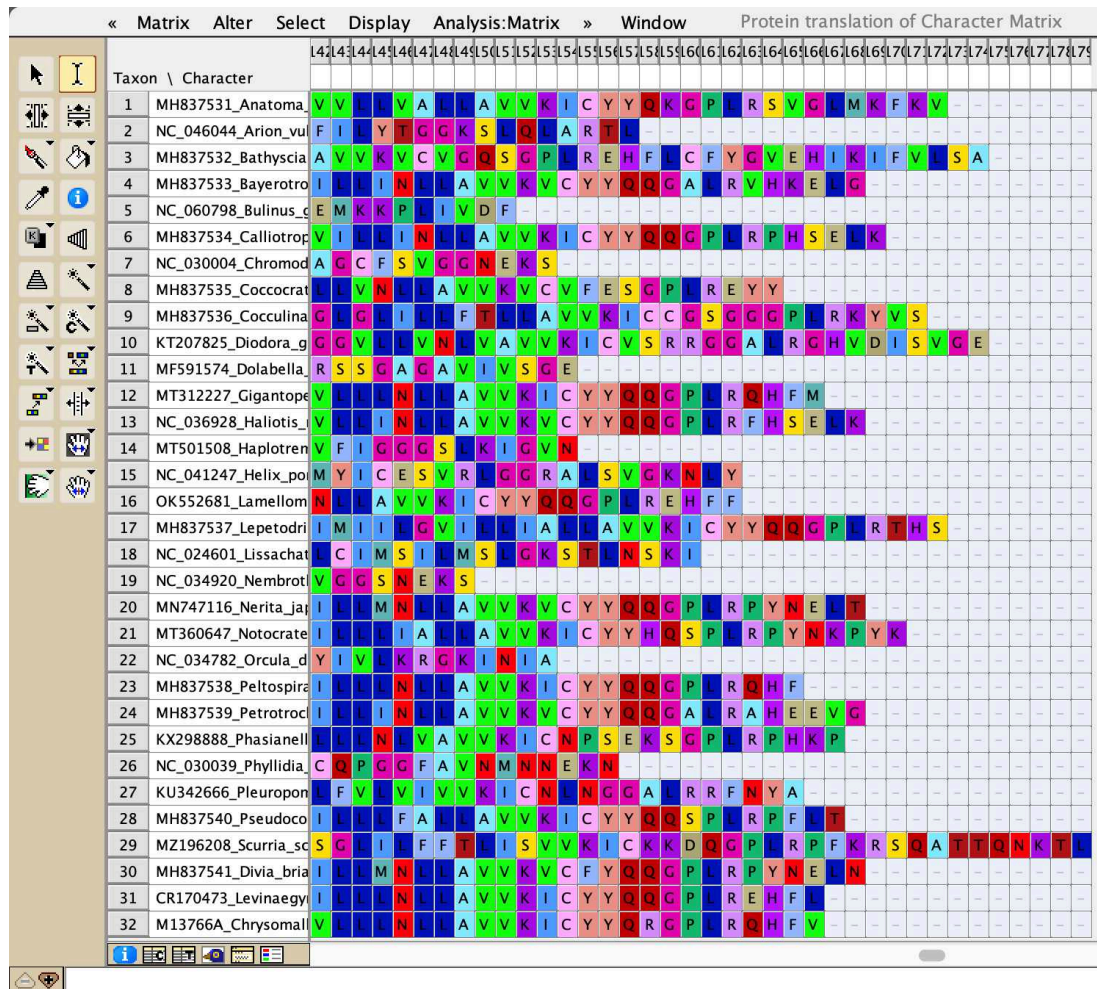


4.7 It's a good idea to save your file here, in case you run into issues in this next step, because Mesquite doesn't have a useful undo feature. Go to the end of the sequences. Delete any empty columns, and then delete all of the stop codons, because they interfere with the RaXML analysis. You should be able to delete the stop codons by just selecting them and deleting them using the "Select" tool, but if that doesn't work for some reason, you can use the "Edit" tool and click on each stop codon and type "-" instead of "*" and the end result is the same. Make sure you get ALL of them! Even those from shorter sequences that may be hiding more towards the beginning of the sequences.

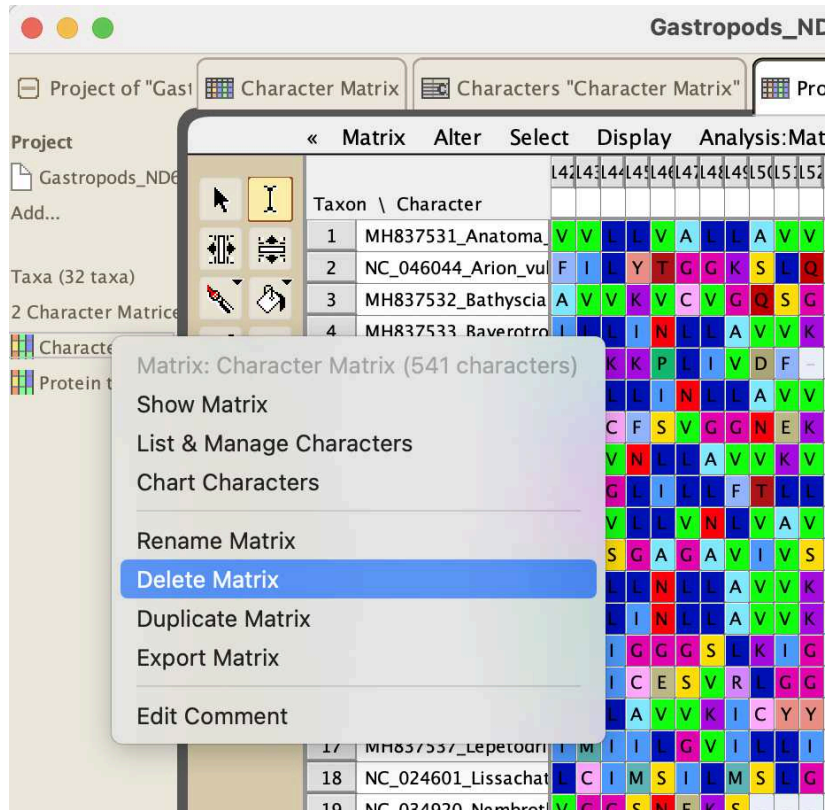
You go from this:



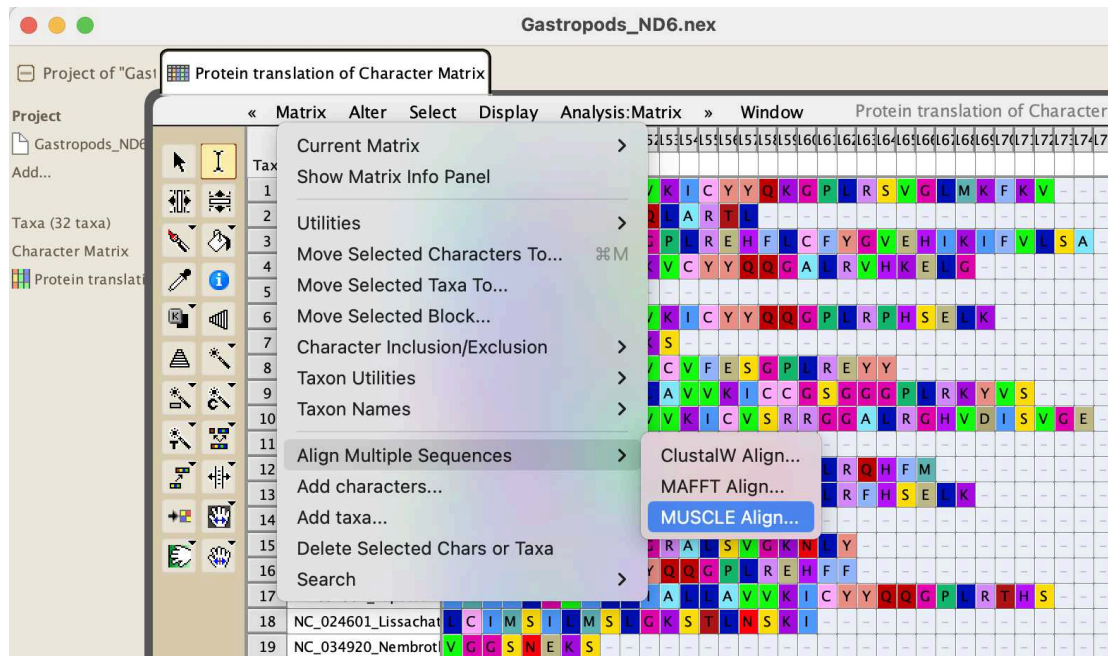
To this:



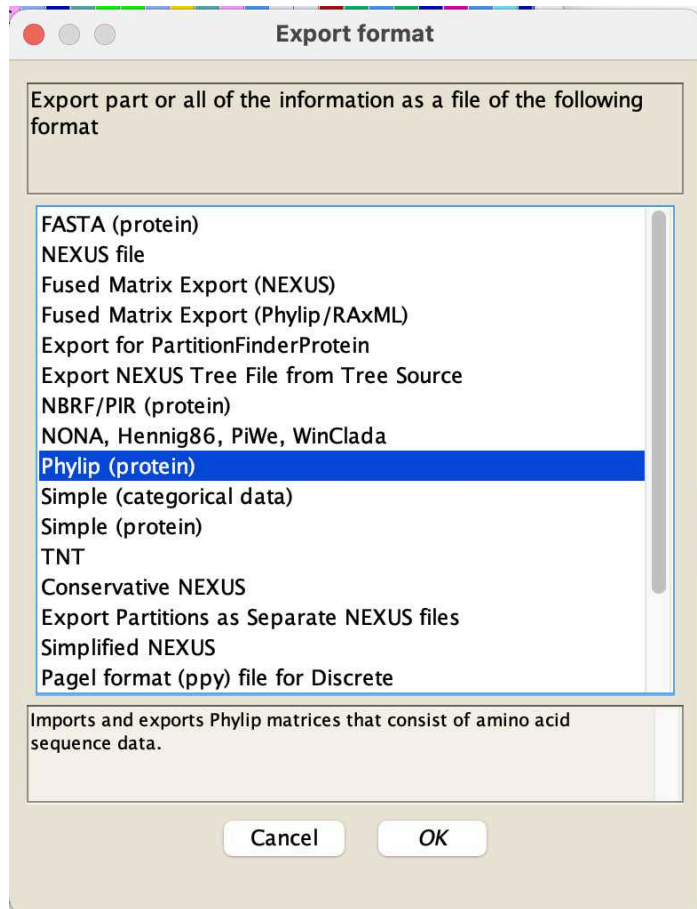
4.8 You want a file with just the nucleotides and a file with just the amino acids to not confuse RaXML. Because you duplicated this file in the beginning and already have a nucleotide version, go ahead and delete the character matrix from the side bar by clicking on it, then selecting "Delete Matrix". You should be left with just the "Protein Translation of Character Matrix" tab.



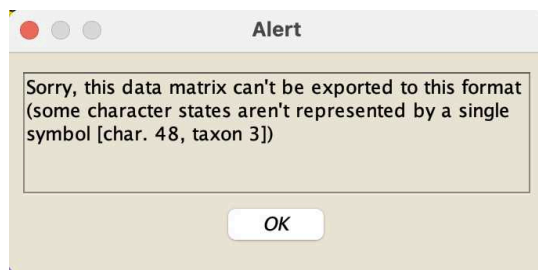
- 4.9 Now align the sequences. Click on "Matrix", "Align Multiple Sequences", then "MUSCLE Align". Select "No" for running on multiple threads, and then click okay. Check to see if everything is fine (depending on how related your organisms are, the alignment might also look nicer or not as nice). Then, delete any empty columns at the end of the alignment. Save the file.



- 4.10 Export as a Phylip (protein) file. MAKE SURE it says "protein" in parentheses next to it! You can name it something like "Gastropods_ATP6_AA.phy". Change the number for "Maximum length of taxon names" so that you will get the whole name of the sequence in the phylip file. I usually put something large like 100 or 200 just to be safe.



You might get an error saying "Sorry, this data matrix can't be exported to this format (some character states aren't represented by a single symbol [char. 91, taxon 24])". In that case, find the character and change the value (which will be something like "W/M/V/C") to a "?" using the "Edit" tool. Make sure to save the file again before retrying to export.



4.11 Repeat this for every protein coding gene you want to use amino acids for.

- Open RaXML GUI and load all of your alignments - the two non protein coding gene nucleotide phylip files, and the protein coding gene amino acid phylip files (or whatever applies for your analysis). Drag the first one in, then select the rest of them with the "ADD ALIGNMENT" feature.

Note

You can opt to do this part on the lab computer remotely, if you want to spare your computer! I will either do that, or have the analysis run overnight on my laptop. Depending on the capacity of your computer, this might take a really long time, it only takes a few hours on the lab computer.

- Select "RUN MODELTEST" for all of them (each gene you've added should have a separate option to click "RUN MODELTEST"). Make sure you have 15 genes if you're doing a mitogenome analysis, and make sure you have really ran the model test function for all of them.

The screenshot shows the raxmlGUI 2.0 interface. The 'INPUT' tab on the left shows a list of alignment files, including 'Gastropods_ATP6_AA.phy' and 'RAxML_Gastropods_12S_concat.txt'. The 'MODELTEST' tab in the center shows options for substitution models, stationary frequencies, and proportion of invariant sites. The 'COMMANDS' tab on the right displays the generated command line for running the analysis, including the path to the raxml-ng binary and the output directory.

- Set the analysis to "ML + thorough bootstrap + consensus", set Runs to 10, and keep Reps. at 100 (optionally, you can increase both numbers). Leave "<none>" as the outgroup and whatever seed RaXML decided on.

The screenshot displays the raxmlGUI 2.0 interface, which is used for running RAxML analyses. The interface is organized into several panels:

- Input Panel (Top Left):** Shows the input file 'Gastropods_ATP6_AA.phy' (32 sequences, length 303). It includes options for substitution models (+F (empirical), +GAMMA (mean)), stationary frequencies, and proportion of invariant sites. A 'RUN MODELTEST' button is present.
- Command Line Panel (Top Right):** Displays the command line for raxml-ng, including the path to the input file and various options like --all, --msa, --prefix, --tree, and --bs-metric.
- Concatenated Alignment Panel (Bottom Left):** Shows the concatenated alignment 'RAXML_Gastropods_125_concat.txt' (33 sequences, length 7925 from 15 alignments). It includes options for filling gaps and GTR+I+G models.
- Analysis Settings Panel (Bottom Right):** Shows the analysis settings, including the model 'ML + thorough bootstrap + consensus', the number of runs (10), and the number of repetitions (100).
- Output Panel (Bottom):** Shows the output file name '800081' and the outgroup 'Seed'.

- 8 Run the analysis (bottom right corner of the top right window - sometimes you need to expand the program window to find it).
I usually open a youtube window and set an i.e. 10h timer video so my computer doesn't go to sleep while it's computing.
- 9 Once the analysis is done, find the output file that ends in ".support" and change the suffix to ".tre", so you can open the file in Figtree.
- 10 In Figtree, to see the results with the bootstrap results, check off "Node Labels" and in the drop-down menu for "Display", choose "labels".