

Jun 11, 2025

De Novo Assembly of Sequences from Eurofins with Geneious

DOI

<https://dx.doi.org/10.17504/protocols.io.eq2lyn5wqvx9/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.eq2lyn5wqvx9/v1>

Protocol Citation: Dakota Betz 2025. De Novo Assembly of Sequences from Eurofins with Geneious. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.eq2lyn5wqvx9/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: June 24, 2022

Last Modified: June 11, 2025

Protocol Integer ID: 65271

Keywords: eurofins with geneious, de novo assembly of sequence, de novo assembly, eurofin, geneious, sequence

Abstract

Our protocols are constantly evolving and old versions will be deleted.

The documents here are not intended to be cited in publications



1 Download the results from Eurofins. You should get an email like this:

GenomicsSupport@eurofins.com 9:28 AM (3 hours ago) ☆ ↶ ⋮
to ahatch, marruda, ahiley, me, pproctor, k4green ▼

Dear Greg Rouse,

The sequencing results generated from your samples are now ready and may be downloaded via the following link:

[Download Results](#)

If you have any questions, concerns, or need any technical insight regarding your sequencing projects, our Technical Support Team is here to help. Please e-mail us at GenomicsSupport@eurofins.com or call 1-800-688-2248 (option 2) and we will be happy to assist you.

Additionally, please refer to the file attached to this e-mail titled "Viewing Your Sequencing Result Files" for technical clarifications regarding the extraction and subsequent analysis of these data.

Thank you for choosing Eurofins Genomics LLC for your DNA sequencing needs.

Best Regards,
Your Technical Support Team

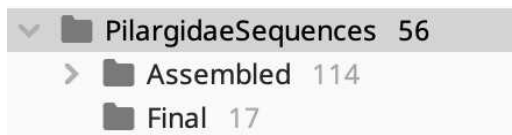
eurofins | Genomics
12701 Plantside Drive | Louisville, Kentucky 40299
Phone: 1-800-688-2248
GenomicsSupport@eurofins.com

Click on the underlined and blue highlighted "Download Results" in the second line of the email content. This will download a .zip file of the sequencing results.

Then, to keep your computer organized, move that file from your downloads into whichever folder you've designated for it, i.e. Documents > RouseLab > EurofinsRawReads > "DateOfResultsReceived".

Double click on the zip file to expand it.

- 2 Open Geneious (either Prime or the shared license, or on a lab computer). Either navigate to or create a folder named "OrganismYouAreWorkingWithSequences" - mine, for instance, is "PilargidaeSequences". Within that folder, I also have an "Assembled" folder, which is where I put the Eurofins raw reads once I'm done with them, and a "Final" folder, where I put the sequences I get from the raw reads.





- 3 In the unzipped folder you downloaded from Eurofins, sort it by file type. You only want to import files with the ending ".ab1". If your computer allows it (mine doesn't like this, but it should work), you can also search for your four letter code (for instance mine is sonj, Marina's is mari, whatever your specific code is) within the folder to filter out the sequences that are yours.

Name
 A4257BsonjPolyHCO_...late GR1623_D03.ab1
 A4257BsonjPolyLCO_...Plate GR1623_F01.ab1
 A4257CsonjPolyHCO...Plate GR1623_E03.ab1
 A4257CsonjPolyLCO_...Plate GR1623_G01.ab1
 A4257DsonjPolyHCO...Plate GR1623_F03.ab1
 A4257DsonjPolyLCO_...Plate GR1623_H01.ab1
 A4257EsonjPolyHCO_...late GR1623_G03.ab1
 A4257EsonjPolyLCO_...Plate GR1623_A02.ab1
 A4257FsonjPolyHCO_...late GR1623_H03.ab1
 A4257FsonjPolyLCO_...Plate GR1623_B02.ab1
 A4257GsonjPolyHCO...Plate GR1623_A04.ab1
 A4257GsonjPolyLCO_...late GR1623_C02.ab1

- 4 In Genious, navigate to the working folder you want to drag your sequences into. Then, in Finder, select the raw reads that are yours and end in .ab1, and drag them into Geneious. The app should look something like this:

Name	Description	Modified	% HQ	Sequence L...	Post-Trim Le...	Topology	Molecule
A4257BsonPolyHCO_PREMIX_Plate_Plate GR1623_D03.ab1	-	24 Jun 2022 12:22 pm	84.1%	762	762	linear	DNA
A4257BsonPolyLCO_PREMIX_Plate_Plate GR1623_F01.ab1	-	24 Jun 2022 12:22 pm	84.9%	743	743	linear	DNA
A4257CsonPolyHCO_PREMIX_Plate_Plate GR1623_F03.ab1	-	24 Jun 2022 12:22 pm	89.0%	718	718	linear	DNA
A4257CsonPolyLCO_PREMIX_Plate_Plate GR1623_G01.ab1	-	24 Jun 2022 12:22 pm	91.3%	690	690	linear	DNA
A4257DsonPolyHCO_PREMIX_Plate_Plate GR1623_H01.ab1	-	24 Jun 2022 12:22 pm	90.9%	695	695	linear	DNA
A4257DsonPolyLCO_PREMIX_Plate_Plate GR1623_H03.ab1	-	24 Jun 2022 12:22 pm	90.5%	687	687	linear	DNA
A4257EsonPolyHCO_PREMIX_Plate_Plate GR1623_G03.ab1	-	24 Jun 2022 12:22 pm	92.5%	692	692	linear	DNA
A4257EsonPolyLCO_PREMIX_Plate_Plate GR1623_A02.ab1	-	24 Jun 2022 12:22 pm	92.7%	687	687	linear	DNA
A4257FsonPolyHCO_PREMIX_Plate_Plate GR1623_H03.ab1	-	24 Jun 2022 12:22 pm	93.2%	692	692	linear	DNA
A4257FsonPolyLCO_PREMIX_Plate_Plate GR1623_B02.ab1	-	24 Jun 2022 12:22 pm	91.4%	685	685	linear	DNA
A4257GsonPolyHCO_PREMIX_Plate_Plate GR1623_A04.ab1	-	24 Jun 2022 12:22 pm	93.3%	691	691	linear	DNA
A4257GsonPolyLCO_PREMIX_Plate_Plate GR1623_C02.ab1	-	24 Jun 2022 12:22 pm	93.7%	687	687	linear	DNA
A5985BsonPolyHCO_PREMIX_Plate_Plate GR1623_G02.ab1	-	24 Jun 2022 12:22 pm	93.1%	698	698	linear	DNA
A5985BsonPolyLCO_PREMIX_Plate_Plate GR1623_A01.ab1	-	24 Jun 2022 12:22 pm	92.0%	687	687	linear	DNA
A5985CsonPolyHCO_PREMIX_Plate_Plate GR1623_H02.ab1	-	24 Jun 2022 12:22 pm	92.9%	700	700	linear	DNA
A5985CsonPolyLCO_PREMIX_Plate_Plate GR1623_B01.ab1	-	24 Jun 2022 12:22 pm	92.3%	687	687	linear	DNA
A5985DsonPolyHCO_PREMIX_Plate_Plate GR1623_A03.ab1	-	24 Jun 2022 12:22 pm	93.1%	693	693	linear	DNA
A5985DsonPolyLCO_PREMIX_Plate_Plate GR1623_C01.ab1	-	24 Jun 2022 12:22 pm	92.3%	684	684	linear	DNA
A5985EsonPolyHCO_PREMIX_Plate_Plate GR1623_B03.ab1	-	24 Jun 2022 12:22 pm	93.8%	694	694	linear	DNA
A5985EsonPolyLCO_PREMIX_Plate_Plate GR1623_D01.ab1	-	24 Jun 2022 12:22 pm	93.1%	686	686	linear	DNA
A5985FsonPolyHCO_PREMIX_Plate_Plate GR1623_C03.ab1	-	24 Jun 2022 12:22 pm	91.9%	695	695	linear	DNA

- 5 If they aren't already, select all of the raw reads (you can do this easily by clicking the box just to the left of the "Name" field). Then, click on the "Align/Assemble" tool, and select "De Novo Assemble..."

Name	Description
A4257BsonPolyHCO_PREMIX_Plate_Plate GR1623_D03.ab1	-
A4257BsonPolyLCO_PREMIX_Plate_Plate GR1623_F01.ab1	-
A4257CsonPolyHCO_PREMIX_Plate_Plate GR1623_F03.ab1	-
A4257CsonPolyLCO_PREMIX_Plate_Plate GR1623_G01.ab1	-
A4257DsonPolyHCO_PREMIX_Plate_Plate GR1623_H01.ab1	-
A4257DsonPolyLCO_PREMIX_Plate_Plate GR1623_H03.ab1	-
A4257EsonPolyHCO_PREMIX_Plate_Plate GR1623_G03.ab1	-
A4257EsonPolyLCO_PREMIX_Plate_Plate GR1623_A02.ab1	-
A4257FsonPolyHCO_PREMIX_Plate_Plate GR1623_H03.ab1	-
A4257FsonPolyLCO_PREMIX_Plate_Plate GR1623_B02.ab1	-
A4257GsonPolyHCO_PREMIX_Plate_Plate GR1623_A04.ab1	-
A4257GsonPolyLCO_PREMIX_Plate_Plate GR1623_C02.ab1	-
A5985BsonPolyHCO_PREMIX_Plate_Plate GR1623_G02.ab1	-
A5985BsonPolyLCO_PREMIX_Plate_Plate GR1623_A01.ab1	-

- 6 Then, in the first part of the window that pops up, change the "part of name, separated by" to whatever YOUR four letter code is, and click "OK". The rest of the parameters should be default, but check to make sure yours looks like this, too.

57CsonjPolyLCO_PREMIX_Plate_Plate GR1623_G01.ab1 - 24 Jun 2022 12:22 pm 91.3

57DsoniPolvHCO_PREMIX_Plate_Plate GR1623_F03.ab1 - 24 Jun 2022 12:22 pm 90.9

57 De Novo Assemble

57 Data

57 ☒ Assemble by: 1st part of name, separated by sonj

57 ☐ Assemble each sequence list separately ☐ Assemble each paired read separately

57 Method

85 Assembler: Geneious ?

85 Sensitivity: Highest Sensitivity / Slow ?

85 Memory Required: 24 MB of 13 GB

85 Trim Before Assembly

85 ☐ Use existing trim regions

85 ☐ Remove existing trim regions from sequences

85 ☒ Trim sequences Options

85 ☐ Do not trim

85 Results

85 Assembly Name {Reads Name} Assembly ...

85 ☒ Save assembly report

85 ☒ Save list of unused reads

85 ☐ Save in sub-folder

85 ☒ Save contigs

85 ☐ Save consensus sequences Options

85 More Options

85 Cancel OK

7 Create an excel file and create a table for your sequences. Mine is called "PilargidaeDeNovoAssemblyStats". Have column headers that are: "Sample ID", "File Name/Assembly", "HQ% Before", "HQ% After", "Sequence Length Before", "Sequence Length After", "Ambiguities Before", "Ambiguities After", "Gaps Before", "Gaps After", and "Date Received".

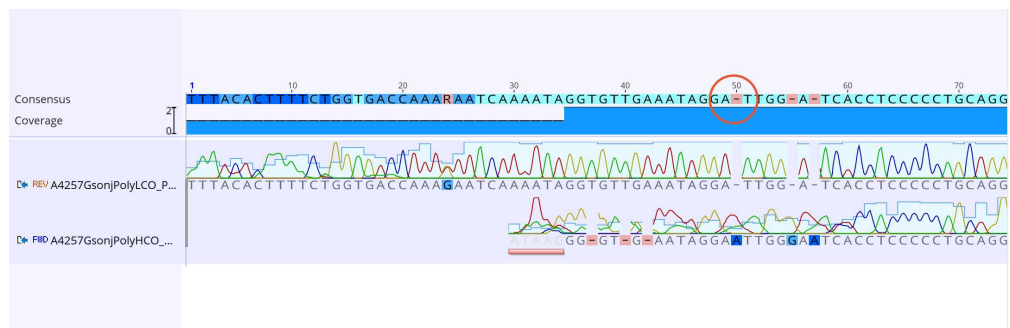
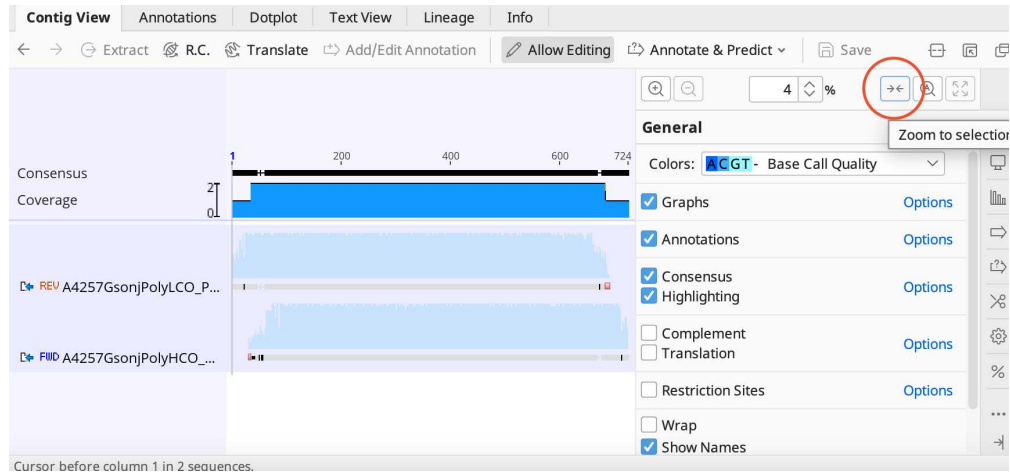
Mine looks like this I like to also add the stats for each raw read, not just the assembly, but the assembly by itself should be enough.

Sample ID	File Name/Assembly	HQ% Before	HQ% After	Seq Length Before	Seq Length After	Ambiguities Before	Ambiguities After	Gaps Before	Gaps After	Date Received
A2813	A2813 Assembly Contig arl/BrH	91.5%	99.3%	572	449	3	1	0	0	Aug 6 2021
3	A2813sonjBrH_PREMIX_Plate_Plate 01_C03.ab1	49.8%		548		0	0	0	0	Aug 6 2021
4	A2813sonjArL_PREMIX_Plate_Plate 01_A03.ab1	82.3%		552		1	0	0	0	Aug 6 2021
5	A2813 Assembly Contig 185	96.5%	99.0%	1,841	1,735	1	0	1	1	Aug 6 2021
6	A2813sonj1F_PREMIX_Plate_Plate 01_G04.ab1	68.0%		952		0	0	0	0	Aug 6 2021
7	A2813sonj3F_PREMIX_Plate_Plate 01_B04.ab1	72.6%		944		0	0	0	0	Aug 6 2021
8	A2813sonj5R_PREMIX_Plate_Plate 01_A05.ab1	59.1%		908		0	0	0	0	Aug 6 2021
9	A2813sonj9R_PREMIX_Plate_Plate 01_H05.ab1	61.3%		666		0	0	0	0	Aug 6 2021
10	A2813sonjA2_PREMIX_Plate_Plate 01_C05.ab1	42.2%		665		0	0	0	0	Aug 6 2021
11	A2813sonjB1_PREMIX_Plate_Plate 01_E03.ab1	74.8%		912		0	0	0	0	Aug 6 2021
A4846	A4846 Assembly Contig arl/BrH	98.7%	100.0%	564	530	0	0	0	0	Aug 6 2021
13	A4846sonjArL_PREMIX_Plate_Plate 01_B03.ab1	93.4%		541		0	0	0	0	Aug 6 2021
14	A4846sonjBrH_PREMIX_Plate_Plate 01_D03.ab1	87.0%		544		0	0	0	0	Aug 6 2021
15	A4846 Assembly Contig 185	98.4%	100.0%	1,828	1,782	0	0	1	1	Aug 6 2021
16	A4846sonj1F_PREMIX_Plate_Plate 01_G07.ab1	97.2%		943		0	0	0	0	Aug 6 2021
17	A4846sonj3F_PREMIX_Plate_Plate 01_C04.ab1	96.4%		945		0	0	0	0	Aug 6 2021
18	A4846sonj5R_PREMIX_Plate_Plate 01_H07.ab1	96.9%		946		0	0	0	0	Aug 6 2021
19	A4846sonj9R_PREMIX_Plate_Plate 01_A06.ab1	97.5%		660		0	0	0	0	Aug 6 2021
20	A4846sonjA2_PREMIX_Plate_Plate 01_D05.ab1	92.4%		665		0	0	0	0	Aug 6 2021
21	A4846sonjB1_PREMIX_Plate_Plate 01_F03.ab1	97.8%		942		0	0	0	0	Aug 6 2021
A5771	A5771 Assembly Contig arl/BrH	92.7%	99.6%	562		0	0	0	0	Aug 6 2021
23	A5771sonjArL_PREMIX_Plate_Plate 01_A07.ab1	93.8%		536		0	0	0	0	Aug 6 2021
24	A5771sonjBrH_PREMIX_Plate_Plate 01_D07.ab1	75.8%		545		0	0	0	0	Aug 6 2021
25	A5771 Assembly Contig 185	98.6%	99.9%	1,847	1,804	0	0	12	12	Aug 6 2021
26	A5771sonjB1_PREMIX_Plate_Plate 01_G03.ab1	91.1%		948		0	0	0	0	Aug 27 2021

- 8 In Geneious, click on an assembly. In the spreadsheet, note what the sample ID is, the assembly or file name, and then scroll to the right to see what the HQ% is and write that down ("HQ% Before"), the next column over should have the sequence length and note that ("Sequence Length Before"), and 13 columns to the right should be "Ambiguities" ("Ambiguities Before"), so note that number, too.

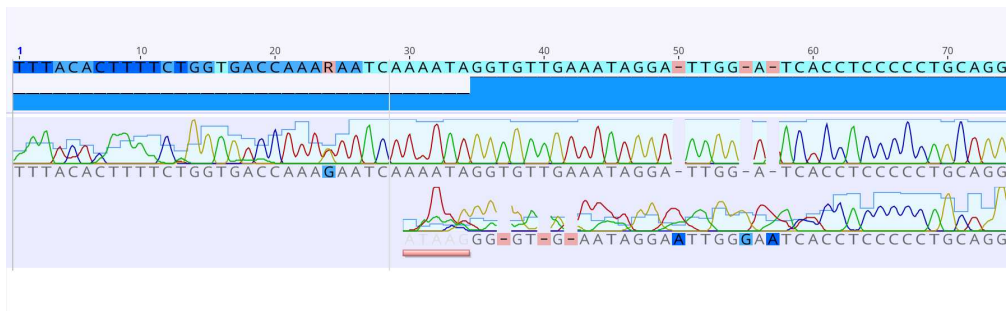
The screenshot displays the Geneious software interface. On the left is a project tree with folders like 'Local', 'LevinagryProject', 'MolluscSequences', 'Nereididae_Mitogenomes', 'Pilargidae_18S', 'Pilargidae_Minimap2_Skri', 'Pilargidae_Mitogenomes', 'PilargidaeSequences', 'Pilargis_verrucosa', 'Sample Documents', 'Reference Features', 'Geneious Plasmid Feat', 'Deleted Items', 'Shared Databases', 'Operations', 'NCBI', and 'UniProt'. The main window shows a table of assemblies with columns for Name, Description, Modified, and % HQ. The table lists various assemblies, including 'USNM1643733sonjA2_PREMIX_Plate_P...', 'USNM1643733 Assembly Contig 2', 'USNM1643733sonj16Sb_PREMIX_Plate_P...', 'USNM1643733sonjAnnF_PREMIX_Plate_P...', 'WF20129 Assembly', 'WF20129sonjPolyHCO_PREMIX_Plate_P...', 'WF20129sonjPolyLCO_PREMIX_Plate_P...', 'WF20132 Assembly', 'WF20132sonjPolyHCO_PREMIX_Plate_P...', 'WF20132sonjPolyLCO_PREMIX_Plate_P...', 'WF20136 Assembly', 'WF20136sonjPolyHCO_PREMIX_Plate_P...', 'WF20136sonjPolyLCO_PREMIX_Plate_P...', '56 documents Assembly Report', 'A4257G Assembly', 'A4257GsonjPolyLCO_PREMIX_Plate_P...', 'A4257GsonjPolyHCO_PREMIX_Plate_Pla...', and 'A4257GsonjPolyHCO_PREMIX_Plate_Pla...'. The bottom panel shows a detailed view of the 'A4257G Assembly' with a 'Contig View' tab selected. It displays a consensus sequence, coverage, and a sequence viewer with a blue highlight and a red dash indicating a gap. The right sidebar contains a 'General' tab with options for 'Colors', 'Base Call Quality', 'Graphs', 'Annotations', 'Consensus', 'Highlighting', 'Complement', 'Translation', 'Restriction Sites', 'Wrap', and 'Show Names'.

Geneious doesn't tell you how many gaps there are, but you can find out for yourself, by clicking the two blue arrows that point towards each other, and then scrolling through the consensus sequence - it's at the top and highlighted in blue - to find gaps. They are highlighted in red and are marked with a dash. Count them, and note that number, as well.

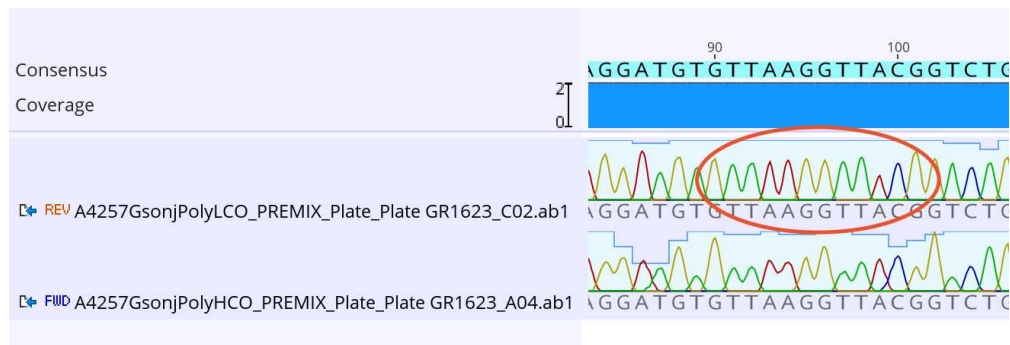


The red highlighted R is an ambiguity, it means Geneious can't quite reliably make a call on which nucleotide should be in that space.

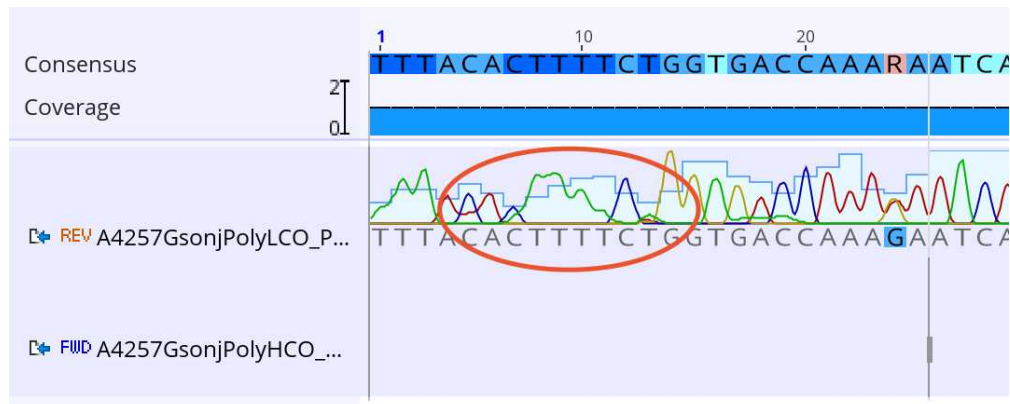
- 9 Once you are done noting the "Before" stats for the assemblies, it's time to edit them, so their HQ% goes up as close to 100% as you can get it. But don't worry if it's not exactly 100%! Sometimes it's better to have a longer sequence and a lower HQ%, especially if you look at the data and you trust it.
- 9.1 Look at the front of the assembly. Light blue is very trusted (both the forward and reverse raw read, or just one of them if it's longer than the other as long as it's clear) support the location of that nucleotide. The darker the blue, the less trusted that data is.



You can see how much confidence there is in the placement of a certain nucleotide by the colorful lines above the bottom two raw reads. In some cases, you can see clear peaks:



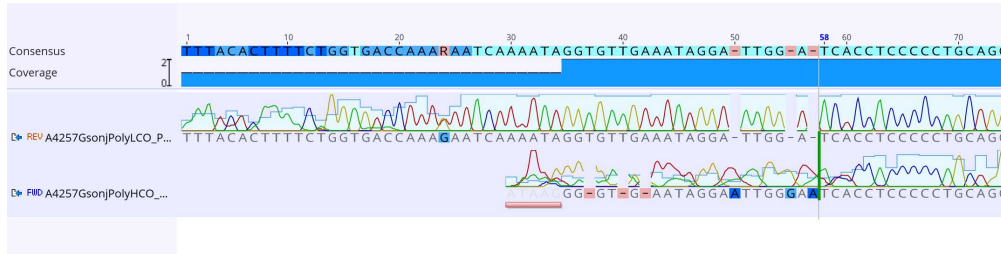
And in other cases, it's really messy:



The clearer and taller the peak, the better.

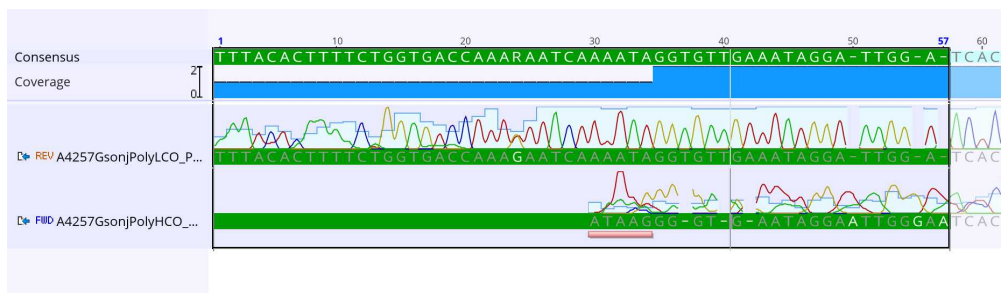
- 9.2 Decide where you want to cut away from the front. You CANNOT cut away parts from the middle, you have to cut everything from the beginning up until a certain point at the front. You want to get rid of as much dark and medium blue as possible, and also get rid of ambiguities if possible without sacrificing too much length (i.e. you don't want to have only 100 base pairs of a gene that you could usually get 700 base pairs from - in that case, it's better to leave an ambiguity or a medium blue somewhere). It's not as essential to also get rid of gaps, but if you can, it's usually advisable.

In my case, I'm choosing to get rid of everything from the beginning including the three gaps (to where the green vertical line is):

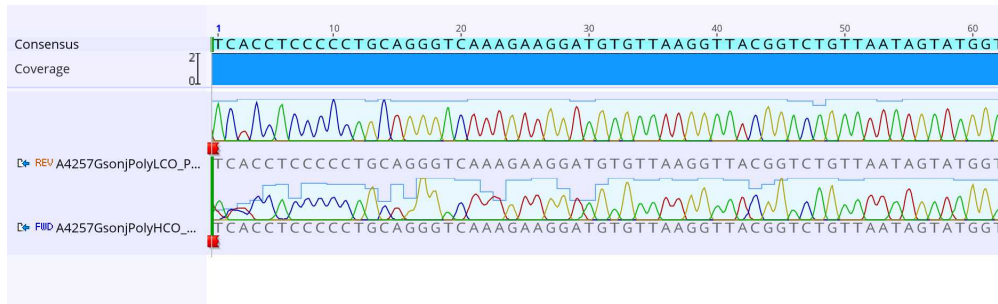


You don't have to, but for my personal records, I like to put my cursor up to where I intend to cut away the less trustworthy assembly, and take a screen shot. Geneious does keep a record of what was done, but sometimes it's just easier to go back and look at a screenshot that's kept in an organized way to know what happened, in case you ever need to check.

- 9.3 Click in the consensus sequence to after the nucleotide you want to cut, and then drag the mouse to the beginning of the alignment but at the bottom left corner - so you make sure two (or six if you're editing 18S) raw reads AND the consensus sequence are highlighted:

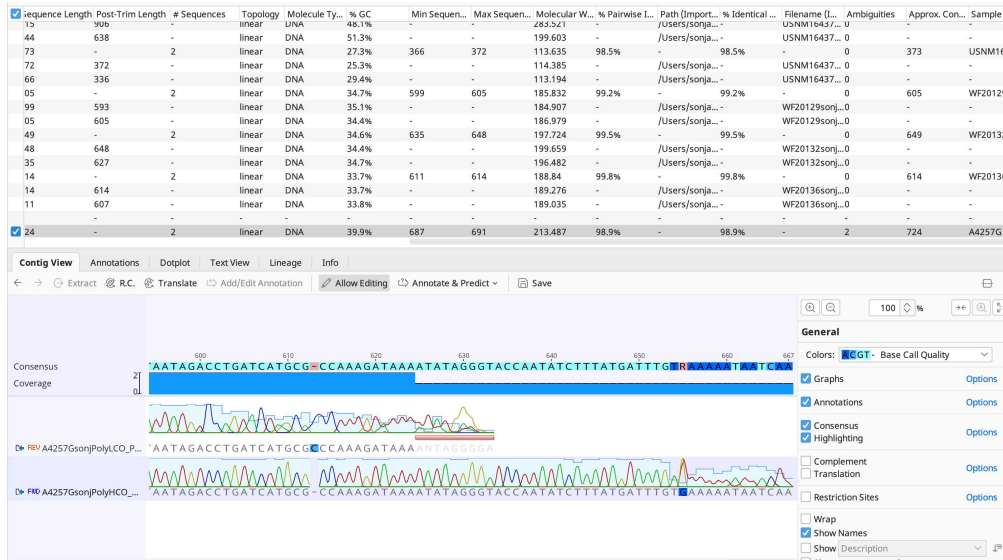


- 9.4 Click your "delete" button! If a pop-up window asks you to allow changes, yes, allow changes. Your sequence should look like this:

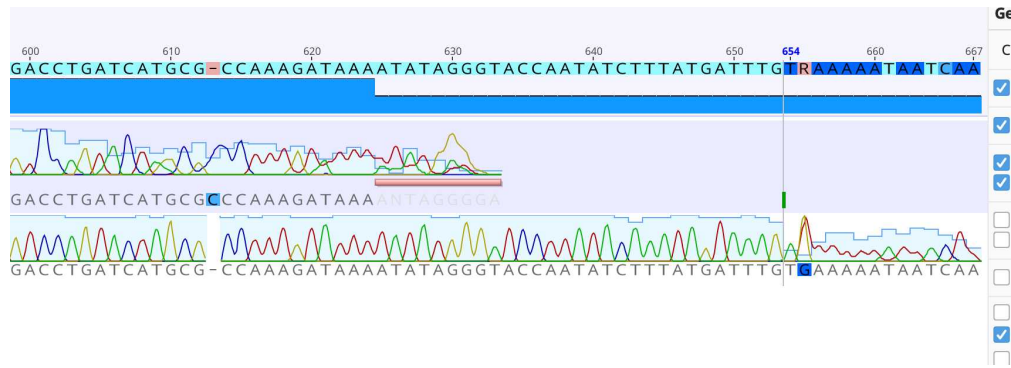


If you hover over the red tags, it will tell you what you deleted.

- 9.5 Then, look at the end of your assembly, and decide from where onward you want to cut the consensus sequence. Again, you CANNOT cut away parts from the middle, you have to cut everything from the nucleotide you want to cut from all the way to the end of the sequence, or it will interfere with your phylogenetic analyses.

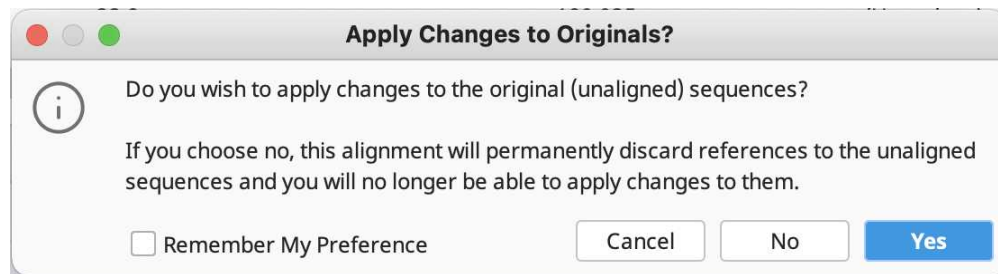


In this case, I am choosing to cut from nucleotide 654:



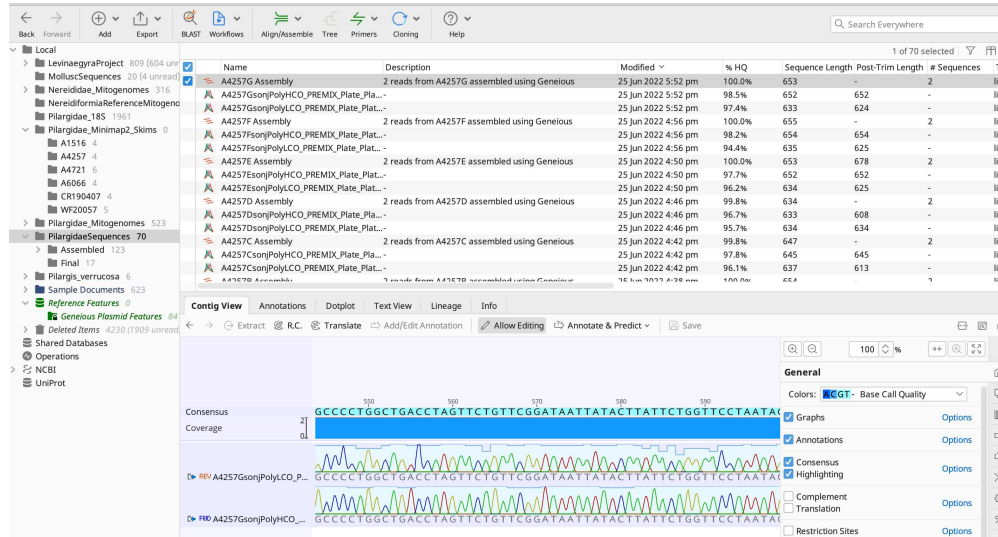
This is because the support for the rest of the spaces after the gap seems pretty good to me. I could however, just as well delete everything including the gap onward. Your call.

- 9.6 After deleting both ends, press "command + S" to save the changes. This popup will appear:

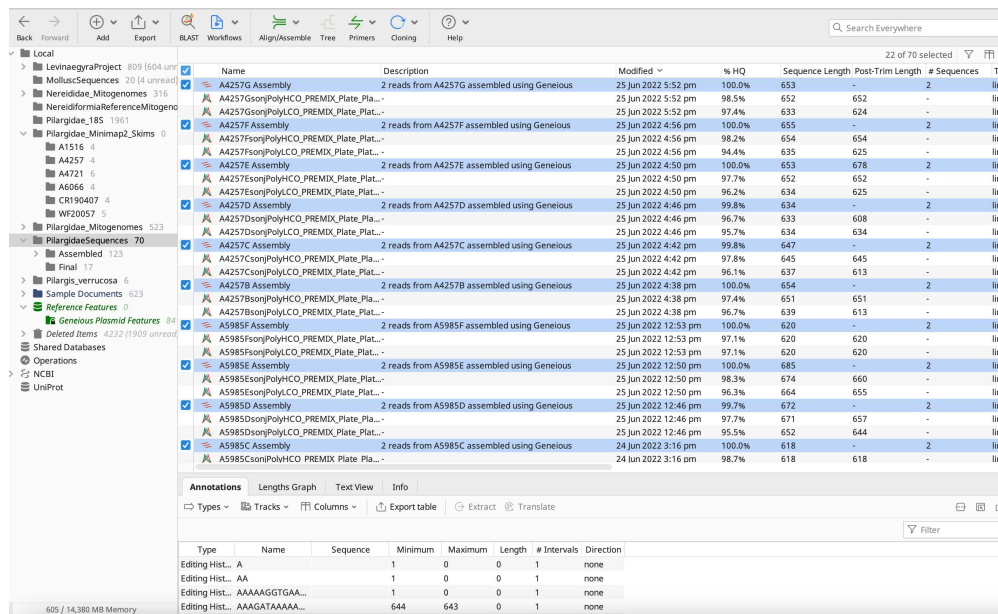


Select "Yes".

If you have Geneious set to sort files by "Modified" (which I recommend), this will put the just modified assembly and its corresponding raw reads at the top of your file list. It will look like this:



- 10 Look at your edited assembly, and note what the HQ% is now after you've edited, what the Sequence Length is, how many ambiguities there are, and count how many gaps there are, like before, and write this down in your excel spreadsheet file.
- 11 Repeat from step 8 to 10 (noting down stats and editing) for every assembly you have.
- 12 Once you're done, select all of the edited assemblies in Geneious.



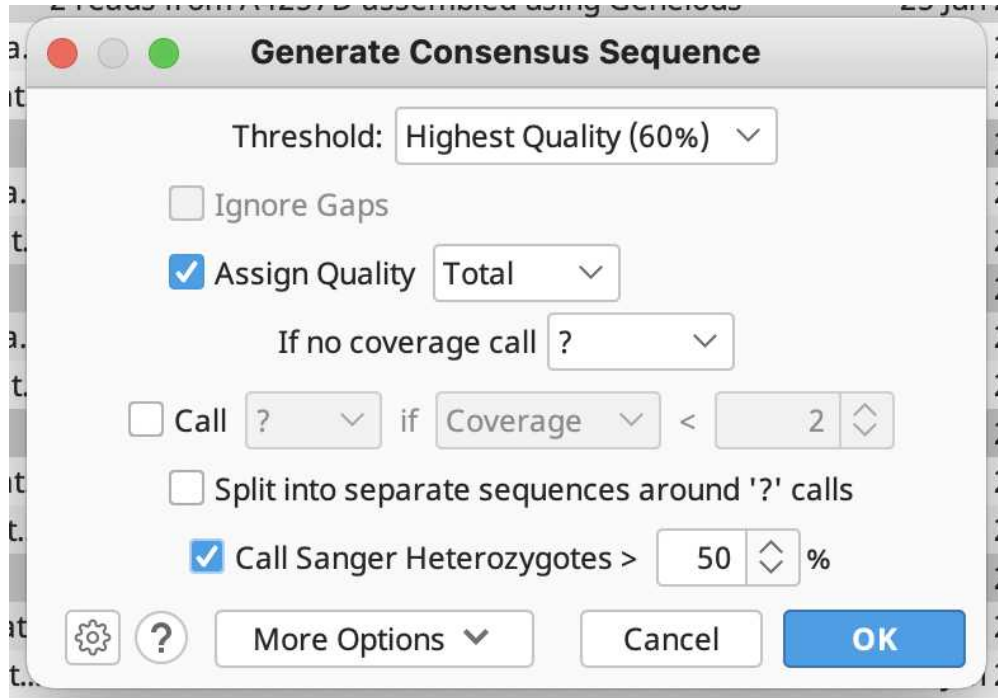


13 Then in the tool bar, click on "Tools" and then "Generate Consensus Sequence..."

The screenshot shows the Geneious Prime application window. The top menu bar includes Apple logo, Geneious Prime, File, Edit, and View. Below the menu bar is a toolbar with icons for Back, Forward, Add, Export, and BLAST. The main area displays a file browser view of a project named 'PilargidaeSequences'.

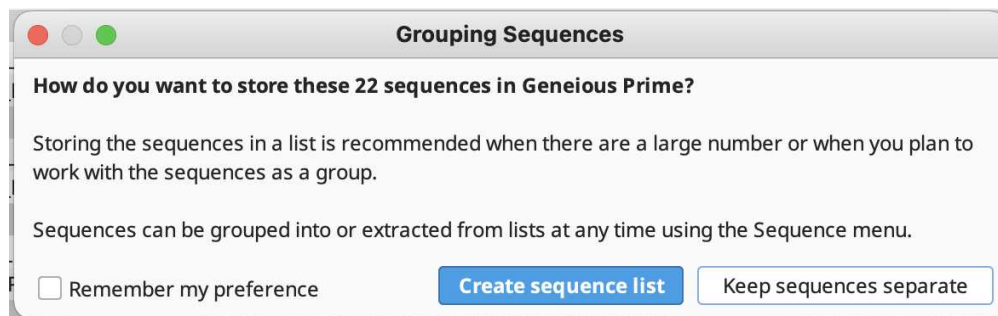
Folder Name	Count	Checkmark	Icon
Local			
LevinaegyraProject	809 (604 unr)	☒	
MolluscSequences	20 (4 unread)	☒	
Nereididae_Mitogenomes	316		
NereidiformiaReferenceMitogeno			
Pilargidae_18S	1961	☒	
Pilargidae_Minimap2_Skims	0		
A1516	4		
A4257	4	☒	
A4721	6		
A6066	4		
CR190407	4	☒	
WF20057	5		
Pilargidae_Mitogenomes	523		
PilargidaeSequences	70	☒	
Assembled	123		
Final	17		

14 You will get this popup:



Generally, everything should be left as default, but check to make sure yours looks like this, too. Click "OK".

15 With this popup, choose "Keep sequences separate".



16 The folder in Geneious that you are working in should look something like this:

Back Forward Add Export BLAST Workflows Align/Assemble Tree Primers Cloning Help						
Local LevinayegraProject 809 (604 unread) MolluscSequences 20 (4 unread) Nereididae_Mitogenomes 316 NereidiformiaReferenceMitogenomes 1 Pilargidae_185 1961 Pilargidae_Minimap2_Skims 0 A1516 4 A4257 6 A4721 6 A6066 4 CR190407 4 WF20057 5 Pilargidae_Mitogenomes 523 PilargidaeSequences 92 (22 unread) Assembled 123 Final 17 Pilargis_verrucosa 6 Sample Documents 623 Reference Features 0 Genieious Plasmid Features 84 Deleted Items 4232 (1909 unread) Shared Databases Operations NCBI UniProt						
U		Name	Description	Modified		%
☒	☐	WF20136 Assembly consensus sequence	2 reads from WF2013...	25 Jun 2022 6:14 pm		100
☒	☐	WF20132 Assembly consensus sequence	2 reads from WF2013...	25 Jun 2022 6:14 pm		99
☒	☐	WF20129 Assembly consensus sequence	2 reads from WF2012...	25 Jun 2022 6:14 pm		99
☒	☐	USNM1643733 Assembly Contig 2 consensus sequence	2 reads from USNM1...	25 Jun 2022 6:14 pm		95
☒	☐	USNM1643733 Assembly Contig 1 consensus sequence	6 reads from USNM1...	25 Jun 2022 6:14 pm		100
☒	☐	USNM1499416 Assembly Contig 3 consensus sequence	2 reads from USNM1...	25 Jun 2022 6:14 pm		100
☒	☐	USNM1499416 Assembly Contig 2 consensus sequence	2 reads from USNM1...	25 Jun 2022 6:14 pm		100
☒	☐	USNM1499416 Assembly Contig 1 consensus sequence	6 reads from USNM1...	25 Jun 2022 6:14 pm		100
☒	☐	A13955 Assembly Contig 3 consensus sequence	2 reads from A13955 ...	25 Jun 2022 6:14 pm		98
☒	☐	A13955 Assembly Contig 2 consensus sequence	2 reads from A13955 ...	25 Jun 2022 6:14 pm		99
☒	☐	A13955 Assembly Contig 1 consensus sequence	6 reads from A13955 ...	25 Jun 2022 6:14 pm		100
☒	☐	A5985E Assembly consensus sequence	2 reads from A5985E ...	25 Jun 2022 6:14 pm		100
☒	☐	A5985D Assembly consensus sequence	2 reads from A5985D ...	25 Jun 2022 6:14 pm		100
☒	☐	A5985C Assembly consensus sequence	2 reads from A5985C ...	25 Jun 2022 6:14 pm		100
☒	☐	A5985B Assembly consensus sequence	2 reads from A5985B ...	25 Jun 2022 6:14 pm		100
☒	☐	A5985F Assembly consensus sequence	2 reads from A5985F ...	25 Jun 2022 6:14 pm		100
☒	☐	A4257F Assembly consensus sequence	2 reads from A4257F ...	25 Jun 2022 6:14 pm		100
☒	☐	A4257E Assembly consensus sequence	2 reads from A4257E ...	25 Jun 2022 6:14 pm		100
☒	☐	A4257D Assembly consensus sequence	2 reads from A4257D ...	25 Jun 2022 6:14 pm		99
☒	☐	A4257C Assembly consensus sequence	2 reads from A4257C ...	25 Jun 2022 6:14 pm		99
☒	☐	A4257B Assembly consensus sequence	2 reads from A4257B ...	25 Jun 2022 6:14 pm		100
☒	☐	A4257G Assembly consensus sequence	2 reads from A4257G ...	25 Jun 2022 6:14 pm		100
☒	☐	A4257G Assembly	2 reads from A4257G ...	25 Jun 2022 5:52 pm		100
☒	☐	A4257GsonjPolyHCO_PREMIX_Plate_Plate GR1623_A04.ab1	-	25 Jun 2022 5:52 pm		98
☒	☐	A4257GsonjPolyHCO_PREMIX_Plate_Plate GR1623_C02.ab1	-	25 Jun 2022 5:52 pm		97
☒	☐	A4257F Assembly	2 reads from A4257F a...	25 Jun 2022 4:56 pm		100
☒	☐	A4257FsonjPolyHCO_PREMIX_Plate_Plate GR1623_H03.ab1	-	25 Jun 2022 4:56 pm		98
☒	☐	A4257FsonjPolyHCO_PREMIX_Plate_Plate GR1623_B02.ab1	-	25 Jun 2022 4:56 pm		94
☒	☐	A4257E Assembly	2 reads from A4257E a...	25 Jun 2022 4:50 pm		100

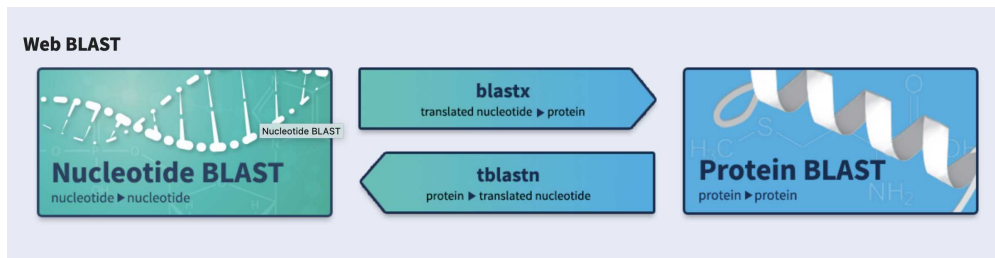
Drag the these sequences into the "Final" folder you created at the beginning, and drag the rest of the assemblies and raw reads into the "Assembled" folder you also created at the beginning. You can choose to have extra folders within the "Assembled" folder for each of the specimens you sequence, to aid in extra organization.

This should leave your working folder empty for the next time you'll De Novo assemble raw reads from Eurofins.

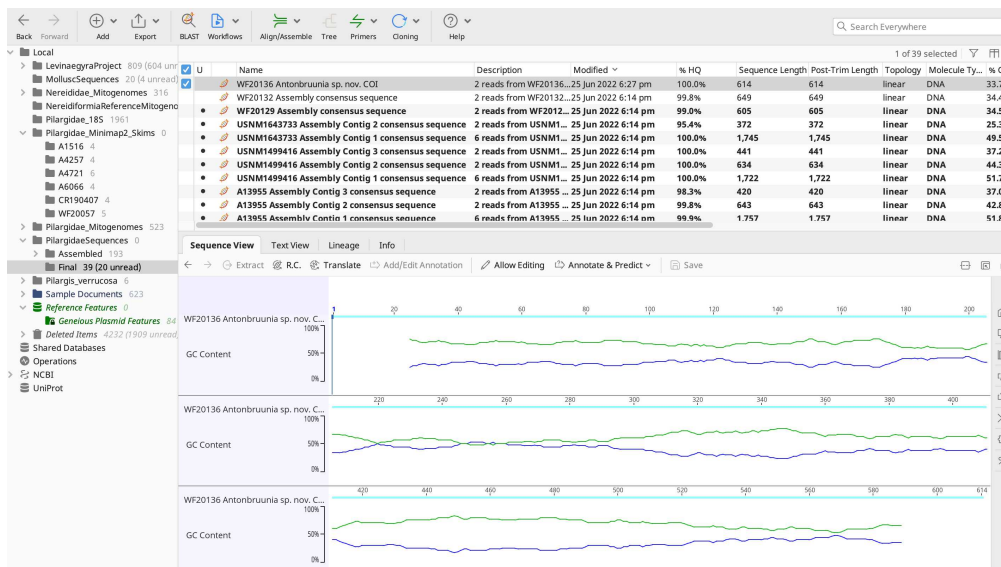
17 Blast: **This next step is VERY IMPORTANT even if it seems like you're done.** Now it's time to blast the sequences, a) to make sure they're what they're supposed to be, and not some weird contamination either on our part or Eurofins' part, and b) to make sure your assembly does not include errors (predicted stop codons or frameshifts). ***These errors are easily fixable now, so do not risk having to redo your analyses and rework your thesis/paper.***

Even if the contamination check is ok, **skipping the quality control step can come back to bite you during the GenBank submission step at the end of your project.** GenBank will automatically block your entire submission if it detects errors in any sequence. Do this now, while your .ab1 files and Geneious are easily accessible. Not years later when you have lost track of them and may not have easy access to Geneious. Reproducibility is critical in science, so the sequences you put on GenBank should reflect what you actually used in your analyses.

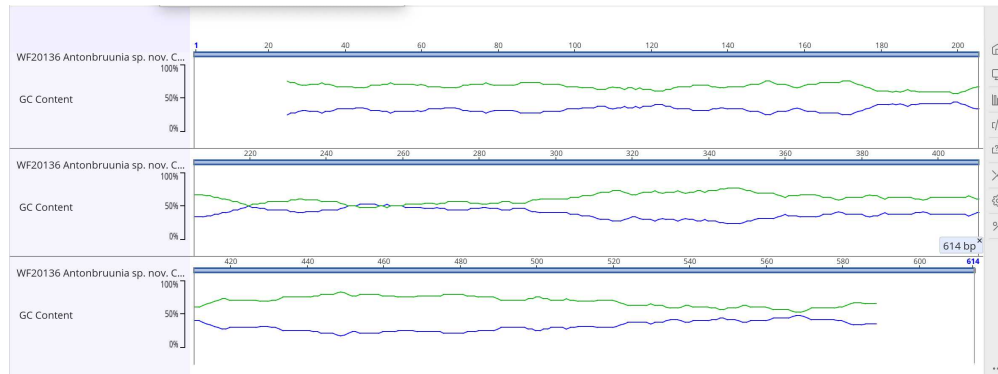
- 17.1 Go to <https://blast.ncbi.nlm.nih.gov/Blast.cgi> and click on "Nucleotide BLAST" (on the left).



- 17.2 Go to the "Final" folder in Geneious where you just dragged the generated sequences. You can rename them to something that tells you more about what they actually are (i.e. add a "COI" or "18S" at the end of whatever is there already in the name, or just keep the sample ID and add what gene it is, whatever helps you know what it is). Select the first one, and it should look something like this:



- 17.3 Click at the beginning of the light blue sequence, and drag to the end of it, so that it turns a darker blue (it's highlighted, selected). Then click "command + C" to copy the sequence.



- 17.4 Go back to the NCBI blastn web page, and paste the sequence into the empty box in the "Enter Query Sequence" section.


National Library of Medicine
National Center for Biotechnology Information

[Log in](#)

BLAST® » blastn suite
[Home](#)
[Recent Results](#)
[Saved Strategies](#)
[Help](#)

blastn
blastp
blastx
tblastn
tblastx

Standard Nucleotide BLAST

BLASTn programs search nucleotide databases using a nucleotide query. more...
[Reset page](#)
[Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

CACC AATT AAAAT GGGT ATTACT AAAAAAAAAA TTTAT TAAAAACGCATGTGCAG
TAACAATCACATTATAAATTTGATCATTACCTAAAAAGAACCGGTTGACCTA
ATTCAACCCGAATAATCACCCTTATAGATATCCCTACCAACCTGCCCAACC
CCTAAATAAAATATAAGTACC

Query subrange [?](#)
From
To

Or, upload file No file selected. [?](#)

Job Title
Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database
☒ Standard databases (nr etc.):
☐ rRNA/ITS databases
☐ Genomic + transcript databases
☐ Betacoronavirus

Nucleotide collection (nr/nt) [?](#)

Organism [Optional](#)
 Enter organism name or id--completions will be suggested ☐ exclude [Add organism](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude [Optional](#)
☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to [Optional](#)
☐ Sequences from type material

Entrez Query [Optional](#)
 [YouTube](#) [Create custom database](#)
Enter an Entrez query to limit search [?](#)

Program Selection

Optimize for
☒ Highly similar sequences (megablast)
☐ More dissimilar sequences (discontiguous megablast)
☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)
☐ Show results in a new window

+ Algorithm parameters

- 18 Select "Somewhat similar sequences (blastn)" under the "Program Selection", and then press "BLAST".



BLAST® » blastn suite Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTn programs search nucleotide databases using a nucleotide query. more... [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [?](#)

`CACCAATTAAATGGGTATTACTAAAAAATAATTATTAACGCGATGTGCAG
TAACAATCAGATTATAATTTGATCATTACCTAAAAAGAACCGGGTTGACCTA
ATTCAACCCGATATACCCCTTATAGATATCCCTACCAACCTGCCCAAAAC
CCTAAATATAAATAAGTACC` From To

Or, upload file No file selected. [?](#)

Job Title
Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus
Nucleotide collection (nr/nt) [?](#)

Organism [Optional](#)
 Enter organism name or id—completions will be suggested ☐ exclude [Add organism](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude [Optional](#)
☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to [Optional](#)
☐ Sequences from type material

Entrez Query [Optional](#)
 Enter an Entrez query to limit search [?](#) [YouTube](#) [Create custom database](#)

Program Selection

Optimize for ☐ Highly similar sequences (megablast)
☐ More dissimilar sequences (discontiguous megablast)
☒ Somewhat similar sequences (blastn) [?](#)
Choose a BLAST algorithm [?](#)

BLAST Search database Nucleotide collection (nr/nt) using Blastn (Optimize for somewhat similar sequences)
☐ Show results in a new window

+ Algorithm parameters

- 19 Check that the overall results make sense given the known identification of your sequence (e.g., not bacterial contamination).
- 20 Click on the 'Description' link for the the top GenBank record that your sequence aligns with.

Under the 'Strand' heading, it should say 'Plus/Plus'. This indicates that your sequence is oriented in the correct forward 5' direction. If it says 'Plus/Minus', then reverse complement the alignment generated in step 1.
- 21 **Additional quality control steps for protein-coding genes such as COI**
 - 21.1 Check the box for 'CDS feature' to display the expected amino acids for your sequence compared to those of the BLAST hit. Mismatches will be shown in pink. Overall the

majority of amino acids should match. Some variation is expected, of course.

Descriptions

Graphic Summary

Alignments

Taxonomy

Alignment view

Pairwise

☒ CDS feature
 [Restore defaults](#)

100 sequences selected

[Download](#)
[GenBank](#)
[Graphics](#)

Amyntas aspergillus isolate M01 mitochondrion, complete genome

Sequence ID: [NC_025292.1](#)
 Length: 15115
 Number of Matches: 1
 [See 1 more title\(s\)](#)
[See all Identical Proteins\(IPG\)](#)

Range 1: 88 to 845
 [GenBank](#)
[Graphics](#)

[Next Match](#)
[Previous Match](#)

Score	Expect	Identities	Gaps	Strand
633 bits(701)	2e-176	595/758(78%)	0/758(0%)	Plus/Plus
CDS: Putative 1	1	M S L I I R V E L G Q P G A F L G S D Q		
Query	15	ATAAGACTAATTATCCGCGTTGAACCTTGGTCAACAGGTGCCTTCCTTGGAAGTGATCAA		74
Sbjct	88	ATAAGACTTCTTATTTCGTATTGAATTAAGACAACCTGGATCCTTCCTTGGAAGAGATCAG		147
CDS:cytochrome c oxi	30	M S L L I R I E L S Q P G S F L G S D Q		
CDS: Putative 1	21	L Y N V V V T A H A F L M I F F L V M P		
Query	75	CTCTACAACGTTGTAGTCACAGCTCACGCATTTTAAATAATTTCTTTCTAGTTATACCA		134
Sbjct	148	CTATAACAACAATTGTAACAGCACACGATTTCTAATAATTTCTTTCTAGTGATGCCA		207
CDS:cytochrome c oxi	50	L Y N T I V T A H A F L M I F F L V M P		

Example BLAST alignment showing an acceptable CDS

If you see a long string of mismatches, there is likely a frameshift in your sequence. A nonzero number of gaps would also suggest a frameshift.

Descriptions
Graphic Summary
Alignments
Taxonomy

Alignment view
Pairwise
CDS feature
Restore defaults

100 sequences selected

Download
GenBank
Graphics

Amyntas aspergillus isolate M01 mitochondrion, complete genome

Sequence ID: [NC_025292.1](#) Length: 15115 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 89 to 845 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
619 bits(686)	1e-172	592/757(78%)	2/757(0%)	Plus/Plus

CDS: Putative 1
Query

14
73

W L I I R V E L G Q P G A F L G S D Q
TATGACTAATTATCCGCGTTGAACCTGGTCAACCAGGTGCCTTCCTTGAAGTGATCAAC

Sbjct
CDS:cytochrome c oxi

89
30

TAA GACTTCTTATTCGATTGAATTAAGACAACCTGGATCCTTCCTTGAAGAGATCAGC
M S L L I R I E L S Q P G S F L G S D Q

CDS: Putative 1
Query

20
74

L Y N V V V T A H A F N N F L S S Y T S
TCTACAACGTTGTAGTCACAGCTCAGCATTT--AATAATTTCTTTCTAGTTATACCAG

Sbjct
CDS:cytochrome c oxi

149
50

TATACAACACAATTGTAACAGCACACGATTCTAATAATTTCTTTCTAGTGATGCCAG
L Y N T I V T A H A F L M I F F L V M P

CDS: Putative 1
Query

40
132

I Y R S T S K L I S P S Y T S S T W Y S
TATTTATCGGAGGACTAGGAACTGATTAGTCCCTCTTACTAGGAGCACCTGATATAG

Sbjct
CDS:cytochrome c oxi

209
70

TATTTATTGGAGGTTTGGAACTGACTGCTCCCACTTACTAGGAACCCCGACATAG
V F I G G F G N W L L P L M L G T P D M

Example BLAST alignment showing a frameshift

If you see a frameshift, revisit that region of the sequence in your .ab1 files in Geneious and check if any base calls need to be edited.

21.2 Also check that there are no stop codons (marked by *) in your coding sequence.

CDS: Putative 1
Query

119
372

G G L F N F S R Y Q L H F Y S D * H T N
CGGGGGTCTCTTCAATTTTAGGCGCTATCAACTTCATTCTACAGTGATTAACATACGAA

Sbjct
CDS:cytochrome c oxi

449
150

CGGGTGCCTCATCAATTTTAGGTGCCATTAACCTTATCACTACAGTAATTAACATACGAT
A G A S S I L G A I N F I T T V I N M R

Example BLAST alignment showing a frameshift (pink text) and stop codon (*)

Revisit your .ab1 files if needed.

21.3 Now check the rest of your sequence for any remaining frameshifts and stop codons. (The BLAST step above was important to check for contamination and tell you the correct reading frame, but it only checked the parts of your sequence that were aligned with BLAST hits.)

For many sequences you may wish to use Mesquite, but if you are only working with a few sequences, you might find it easiest to use a reputable online translation tool such as Expasy (<https://web.expasy.org/translate/>). Choose the appropriate genetic code for your organism, such as "invertebrate mitochondrial". Make sure that one of the 5'3' (forward) reading frames produces an uninterrupted amino acid sequence with no stop



codons (-). This amino acid sequence should overlap with the CDS result from your sequence during the BLAST check.

5'3' Frame 1
Y-EPL-D-LSALNLVNQVPSLEVINSTTL-SQLTHF--FSF-LYQYLSED-ETD-SLLY-EHLM-HSPE-TT-ASDFS LHP SHY-LYPLLWKVWGQAEPTTPHYQ
MM-LMQALLLTLLQFSPT-RGSLQF-ALSTSFLQWLTYEQKVIV-NVFHYLYGPKLSLFFYSYLYQC-RELLLYYLQTEI-MHLSSILQEEGTQSYSTYFDSSA
IQKYM-FFQDLG-SPML-PTMOES-NLSVL-E-PTQC-ELES-D

5'3' Frame 2
ISNLYKTNYPRTWSTCLPFWKWTQLQRCSHSSRIFNNFLSSYTSIYRSTSKLISPSYTSSTWYSIPFNKQHKLLTSPSIPHTTSYIRCCGKWGGDSLNLRPPIIS
-YSSCSPFCWPCNPLPSFGGLFNFSTRYQLHFYSD-HTNKSSTFSTYSTICMVR-NHCHSTSLISTSVSGSYTYTYSKPKFYIFLRSCSSSGPNPIPTLILILRP
SSSMYFNSSSIWGNLPYCNPLCSKVSTFRYFSNNLRNASNWNPSI

5'3' Frame 3
LGTSMSLIIRVELGQPGAFGLGSDQLYNNVVTAHAFLMIFFLVMPVFIGGLGNWLVPMLGAPDMAFFPMNMSFWLLPPSLTLLVMSAAVESGVGTGWTVPPLSD
NMAHAGPSVDLAIFSLHLAGVSSILGAINFISTVINMRTKGLRLERIPFVWSVKITAILLLSLPVLAGAITMLLTDRLNTSFFDPAGGGDPILFQHLFWFFGH
PEVYIILPGFGVISHIVTHYAGKLEFFGTGLGMIYAMLGIGILG

Example translation in Expasy

Revisit your .ab1 files as needed to resolve stop codons or frameshifts.