Forming and editing contigs in GeneStudio

Before we run analyses on the sequences that we just produced we must first combine the forward and reverse sequences. This is called forming a contig. If you remember we used forward and reverse primers in the PCR reactions. We also sequenced the PCR products with the forward and reverse primers separately. In producing these two sequences we can align them (contig) and be sure that the base calls are correct. This tutorial will allow you to form and edit a contig from the sequences you produced. We can then go on to analyze these sequences.

Step 1: Open GeneStudio

Step 2: On the right side of the window click the Contig editor button



Step 3: Click the File tab and then click "Import sequences...."

• Navigate to the file that contains your raw sequences. In this case they should be in the file folder "Raw_Sequences". Select the forward and reverse sequence of one sample (example: 1511L_A03 and 1511H_A01). In this case the specimen is 1511 and the forward product is 1511L and the reverse is 1511H. Once selected click ok to import. Click yes when this window pops up.



Then ensure that the following are checked in the next pop up window

🤞 Trim ends	? <mark>×</mark>
 Trim 5' ends Automatic: Low stringency Medium stringency High stringency 	Fixed position:
 Trim 3' ends Automatic: Low stringency Medium stringency High stringency 	Fixed position: 450
	OK Cancel

Click Yes and then click "Edit Ambiguities"

Step 4: Ensure that your alignment is not reversed

• Looking at the window you can see the consensus sequence and the two original sequence. Make sure that your forward primer is correct



In this case our reverse sequence (1511H) is going forward (green arrow) and the forward is the reverse (red arrow). To correct this go up to the Contig menu and click and then click reverse.



Now our sequences are aligned correctly. It is very important to make sure you know what the forward and reverse sequences should be.



Step 5: Editing contig

• Click show all, this will display the chromatograms of each sequence.





• We can make the chromatograms easier to see by adjusting the trace parameters.

• We begin by editing the beginning of the sequence. It usually takes a few base calls for the sequencing read to stabilize.



• Let us delete the first 15-20 bp of the forward sequenc by selecting those bases in the consensus.



• Once selected hit delete.

• Select the ambiguity tab and then click "next"



• Select the red C and type another c, this changes the c from red to blue. Click next, sometimes an additional base is called that is not really there. Notice the red c in the image below. Notice how wide the peak is compared to the surrounding sharp peaks.



• Now the peaks are even. Continue to the next ambiguity until you are near the end of the sequence. Once at the end we can delete the unclear base calls

53			
show all all Quality			
	1130	1140	1150
састадаат	GAGCTC	САТАБАБАТА	GCACTGGC
	<u> </u>		X
снсіАСААТ	6 H 6 C U U	C A I A G A G A I A	6 6 8 6 7 6 6 6

• Scroll through the whole consensus alignment to double check you have corrected all ambiguities.

Step 6: Renaming and exporting consensus sequence.

- Go to the contig tab and select properties
 - Change the name to the specimen number (in this case 1511)
- Go to File and select "Export current consensus"
 - Navigate to your edited sequences folder within your project files and set the file name as the specimen number (1511) and save the file as a FASTA file. Here is what that file will look like:

гие сан гоглал view пер
>1511
AGCCTACGAAAAACACATCCCCTCATTAAAATCGCTAACGACGCACTAGTCGACCTACCA
GCACCATCCAACATCTCAGCATGATGAAACTTTGGATCCCTCCTAGGACTATGTTTAATT
ACCCAAATCCTAACCGGGCTCTTCCTGGCTATACACTACACCTCAGACATCTCAACCGCA
TTCTCATCAGTAACCCACATCTGCCGTGATGTAAACTACGGCTGACTAATTCGTAATATA
TTCCTATTACCATTTATTATCGCCGCCGCAACCATCCTTCACCTCCTATTCCTTCACGAA
ACAGGATCTAATAATCCAATTGGATTAAATTCAGACGCAGACAAAATCTCCTTCCATCCG
TACTTTACATACAAAGATCTCCTTGGATTTGTAATCATATTATTAGCCCTTACACTACTA
GCACTATTTTCCCCAAACCTATTAGGAGACCCAGAAAACTTCACCCCTGCAAACCCCCTA
GTAACCCCTCCACATATTAAACCAGAATGATACTTCCTATTTGCCTATGCCATCTTACGA
TCAATTCCTAACAAACTAGGAGGAGTTCTTGCACTACTATTCTCAATTCTAGTACTAATA
GTAGTACCCCTTCTACACACTTCAAAACAACGAGGATTAACATTCCGCCCAATCACTCAA

Step 7: BLASTing your sequence with NCBI genbank.

- Go to the BLAST website http://blast.ncbi.nlm.nih.gov
- Enter your sequence in the sequence query window. You can either upload your file or copy and paste it in the window. Make sure that the Nucleotide database is selected and hit BLAST
- You should see something like the following

← → C D blast.ncbi.nlm.nih.gov/Blast.cgi	☆ =
nome recent results Savea strategies nelp	
ICBI/BLAST/blastn suite/Formatting Results - 32PEP55401R	
Edit and Resubmit Save Search Strategies > Formatting options. > Download You Imp3 How to read this page Blast report de:	acription
1511 (1134 letters)	
Query ID Id[233803 Database Name nr Description 1511 Description Nucleotide collection (nt) Molecule type nucleic acid Program BLASTN 2.2.28+ > <u>Citation</u> Other reports: > Search Summary [Taxonomy reports] [Distance tree of results] Other sports: > Search Summary [Taxonomy reports] [Distance tree of results]	
Distribution of 100 Blast Hits on the Query Sequence 😡 Mouse over to see the define, click to show alignments	
Color key for alignment scores </td <td></td>	
Show all de	wnloads ×

• Scroll down and you will see a list of the most closely related sequences in the database.

blast.ncbi.nlm.nih.gov/Blast.cgi		
crintions		
· · · · · · · · · ·		
Sequences producing significant alignments:		
Select: All None Selected:0		5
	Max Total Query E	÷.
Description	score score cover value Ident Accession	
Labeo parvus isolate CTOL3162 cytochrome b (Cytb) gene, complete cds; mitochondrial	2073 2073 100% 0.0 99% JX074292.1	
Labeo parvus isolate CTOL3957 cytochrome b (Cytb) gene, complete cds; mitochondrial	2073 2073 100% 0.0 99% <u>JX074285.1</u>	
Labeo parvus isolate CTOL3958 cytochrome b (Cytb) gene, complete cds; mitochondrial	2067 2067 100% 0.0 99% <u>JX074286.1</u>	
Labeo forskalii Isolate CTOL3959 cytochrome b (Cytb) gene, complete cds; mitochondrial	1701 1701 100% 0.0 94% <u>JX074287.1</u>	
Labeo forskalii isolate T10 cytochrome b (cytb) gene, partial cds; mitochondrial	1661 1661 97% 0.0 94% FJ196832.1	
Labeo forskalii isolate T9 cvtochrome b (cvtb) gene, partial cds; mitochondrial	1661 1661 97% 0.0 94% <u>FJ196831.1</u>	
Labeo forskalii isolate T4 cvtochrome b (cvtb) gene, partial cds; mitochondrial	1655 1655 97% 0.0 94% <u>FJ196833.1</u>	
Labeo dyocheilus isolate CTOL01922 cytochrome b (Cytb) gene, complete cds; mitochondrial	1618 1618 100% 0.0 92% JX074262.1	
Labeo senegalensis mitochondrial DNA, complete genome	1585 1585 100% 0.0 92% AB238968.1	
Labeo altivelis isolate CTOL3912 cytochrome b (Cytb) gene, complete cds; mitochondrial	1580 1580 100% 0.0 92% <u>JX074294.1</u>	
Labeo horie isolate CTOL3960 cytochrome b (Cytb) gene, complete cds; mitochondrial	1574 1574 100% 0.0 92% <u>JX074288.1</u>	
Labeo sp. CTOL3978 crtochrome b (Crtb) gene, complete cds; mitochondrial	1568 1568 100% 0.0 92% <u>JX074230.1</u>	
Labeo calbasu mitochondrion, complete genome	1541 1541 100% 0.0 91% <u>JQ231113.1</u>	
.zip 🗸	- Show all	down
MB / mins left	* <u>Show an</u>	

- We see that *Labeo parvus* CTOL3162 is close to our query sequence. Since we have identified our specimen as *Labeo cylindricus* it seems like we have done everything correctly.
- Form and edit your remaining contigs and verify them in BLAST. Once completed with that we can combine all sequences in an alignment. We will work on that next.