

Protocol for using Sequencher and BLASTing Results

After doing this exercise, you should be able to

- Tell the difference between a distinct peak and an ambiguous one in your chromatogram
- Edit bases in Sequencher
- Assemble a contig
- Use the Basic Local Alignment Search Tool (BLAST) to compare your sequence to those published in GenBank

Preparation

Now that your samples are sequenced, they must be analyzed, aligned, and compared to previously discovered sequences. After cycle-sequencing all samples in the forward and reverse direction, obtain sequence files from Core Lab. Unzip and import folder of sequences into Sequencher program.

Trimming/Editing Sequences

1. Open each sequence separately to view the nucleotide sequence. Once this window opens select Show Chromatogram, which will open another window.
2. Visually analyze the chromatogram to make sure that the peaks are distinct with little overlap and background.
 - a. Verify that Sequencher has correctly identified each base.
 - b. Use the slider (located in the top left of the window) to optimize the peak height if needed.
 - c. To view secondary peaks better, click on a respective base (located just below the slider) to “turn it off.”
 - d. If there is an incorrect read, correct it by highlighting the incorrect base and changing it to the correct one. For example, if it reads A but it should be T, just highlight the A and type T. If an N is incorrectly labeled and should be replaced by a base, then do the same for the N.
3. You can also pull down the sequencer tab and select call secondary peak, but this will call all of them using the ambiguous DNA codes)
4. If there is appears to be generally bad quality repeated at the beginning and/or end of the sequence, delete these bases.

Assembly

5. After editing each sequence, close all chromatogram windows. Check the quality (third column) of each sequence.
6. Set the Assembly Parameters as follows:
 - a. Assembly Algorithm: Select Dirty Data
 - b. Optimize Gap placement: Check Use ReAligner and Prefer 3' Gap Placement
 - c. Minimum Match Percentage: Set to 85
 - d. Minimum Overlap: Set to 45
 - e. Assemble By Name: Optional to enable or not enable (convenient if you have a lot of sequences)
7. Highlight the forward and reverse sequence for a sample.
 - a. Click the Assemble Automatically button to form a contig.

8. A window reading Assembly Completed will appear, noting the number of contigs created.
 - a. Close this window and then click on the contig to open it. A new window will appear with a graphical view of the contig.
9. Click on the Bases tab to view a consensus sequence of these two directions.
 - a. Note that the consensus sequence is along the baseline and the two sequences are above. Mismatches are denoted by a along the consensus sequence.
 - b. To view and edit a particular section, click on a selected base of the consensus (the above bases will highlight as well) and click Show Chromatograms.

BLASTing

1. Open your browser and go to the NCBI BLAST website (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Select Nucleotide BLAST.
2. In the Sequencher program, highlight the entire consensus sequence and copy it into the box labeled Enter Query Sequence on the blast webpage.
3. To mass BLAST a collection of contigs:
 - a. Highlight all contigs
 - b. Export as Consensus
 - c. Select FASTA format
 - d. Export and then upload into NCBI blast webpage.
 - e. Change the page default settings.
 - f. Under Choose Search Set, click on Others
 - g. Under Program Selection, click on Highly Similar Sequences.
4. Now, click on the BLAST button.
5. A new window will appear in a few seconds with a graphic summary, description and alignment of the “best” blasts or matched database sequences to your submitted sequence.
 - a. Scroll down to the Descriptions tab to check the statistics on these matches. The score represents the similarity between your queried sequence and the found sequence. Meanwhile, the E value represents the amount of alignments you would expect to find by chance that have the same score as that respective alignment. Therefore, a valid and significant e-value should be 10^{-5} or less.
 - b. The query coverage is the percent of the query length that is included in the aligned segments.
6. Sort by max score or E value, and ensure that all of the significant sequences are segment 4 of the HA gene and that their H subtypes all match.
7. Record this subtype as the H subtype of your sample in the respective results folder.