## Bioinformatics Lab 2

**Sequence Searching and BLAST**

Objective: The goal of this lab is to retrieve genetic sequence data from the NCBI database that identifies the ‘Wolbachia Sequence’ you generated. The Basic Local Alignment Search Tool (BLAST) is an essential tool for comparing a DNA or protein sequence to other sequences in various organisms. Two of the most common uses are to a) determine the identity of a particular sequence and b) identify closely related organisms that also contain this particular DNA sequence.

Using BLAST to identify a fake sequence and your ‘Wolbachia Sequence’: Begin by linking to the NCBI homepage (www.ncbi.nlm.nih.gov/). Select ‘BLAST’ in Popular Resources. With your new knowledge of Sequence Searching and BLAST, let’s begin with a sequence you make up and then your Wolbachia sequence.

• Select ‘nucleotide BLAST’ under the Basic BLAST category

• Input your own nucleotides (A,T,G,C) that fill one complete line into the Search Box. This is referred to as the query sequence.

• Click on the circle for ‘Others (nr etc.) under Choose Search Set

• Select ‘BLAST!’ at end of page. A new window appears.

• Wait for the results page to automatically launch.

1. Did your fake sequence produce a significant alignment (probably not since a significant hit is below E-10 usually)\_\_\_\_\_\_\_\_\_\_\_\_\_ If yes, how many\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. How many sequences did it search in the database?\_\_\_\_\_\_\_\_\_\_\_\_\_
3. How many nucleotide letters did it search in the database?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

• Select Home at the top of the BLAST page.

• Select ‘nucleotide BLAST’ under the Basic BLAST category

• Enter your Wolbachia sequence below into the Search box.

>Your Wolbachia Sequence

GTTGCAGCAATGGTAGACTCAACGGTAGCAATAACTGCAGGACCTAGAGGAAAAACAGTAGGGATTAATAAGCCCTATGGAGCACCAGAAATTACAAAAGATGGTTATAAGGTGATGAAGGGTATCAAGCCTGAAAAACCATTAAACGCTGCGATAGCAAGCATCTTTGCACAGAGTTGTTCTCAATGTAACGATAAAGTTGGTGATGGTACAACAACGTGCTCAATACTAACTAGCAACATGATAATGGAAGCTTCAAAATCAATTGCTGCTGGAAACGATCGTGTTGGTATTAAAAACGGAATACAGAAGGCAAAAGATGTAATATTAAAGGAAATTGCGTCAATGTCTCGTACAATTTCTCTAGAGAAAATAGACGAAGTGGCACAAGTTGCAATAATCTCTGCAAATGGTGATAAGGATATAGGTAACAGTATCGCTGATTCCGTGAAAAAAGTTGGAAAAGAGGGTGTAATAACTGTTGAAGAGAGTAAAGGTTCAAAAGAGTTAGAAGTTGAGCTGACTACTGGCATGCAATTTGATCGCGGTTATCTCTCTCCGTATTTTATTACAAATAATGAAAAAATGATCGTGGAGCTTGATAATCCTTATCTATTAATTACAGAGAAAAAATTAAATATTATTCAACCTTTACTTCCTATTCTTGAAGCTATTGTTAAATCTGGTAAACCTTTGGTTATTATTGCAGAGGATATCGAAGGTGAAGCATTAAGCACTTTAGTTATCAATAAATTGCGTGGTGGTTTAAAAGTTGCTGCAGTAAAAGCTCCAGGTTTTGGTGACAGAAGAAAGGAGATGCTCGAAGACATAGCAACTTTAACTGGTGCTAAGTACGTCATAAAAGATGAACTT

• Select ‘BLAST!’ A new window appears

1. How long (query length) is the Wolbachia sequence that you used to search the database?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. What is the E-value and Max score of the best hit (in this case, the first matching sequence)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. What is the most likely identity of this sequence? (click on the blue link to the left of the top hit) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What is the title of the scientific publication that reported this sequence (click on the PUBMED 16267140 link)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

• Go back twice when you’re done.

• Select Home at the top of the BLAST page.

• Select ‘nucleotide BLAST’ under the Basic BLAST category

• Now enter only the first 135 base pairs of your Wolbachia sequence below into the Search box.

>Your Wolbachia Sequence

GTTGCAGCAATGGTAGACTCAACGGTAGCAATAACTGCAGGACCTAGAGGAAAAACAGTAGGGATTAATAAGCCCTATGGAGCACCAGAAATTACAAAAGATGGTTATAAGGTGATGAAGGGTATCAAGCCTGAA

• As you did before, select ‘BLAST!’ A new window appears

1. What is the E-value and Max score of the best hit (the first matching sequence)?\_\_\_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_. What do you observe about the E-values? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. Is the identity of the best hit different from when you used the complete nucleotide sequence?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Is it the same gene as identified before?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. From the two BLAST searches, what can you deduce about how the length of a query sequence affects your confidence in the sequence search? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_• Close all web windows. There is a lot of information on the NCBI website. Feel free to explore the website and you can find more tutorials at:

http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html