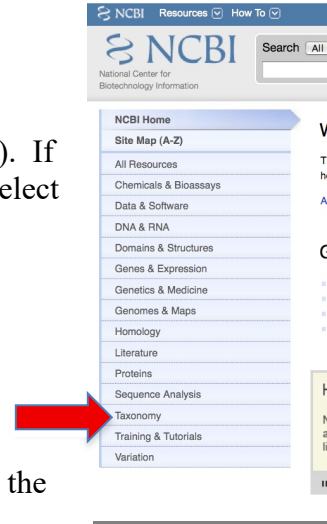


# Bioinformatics Lab (Neu 2026)

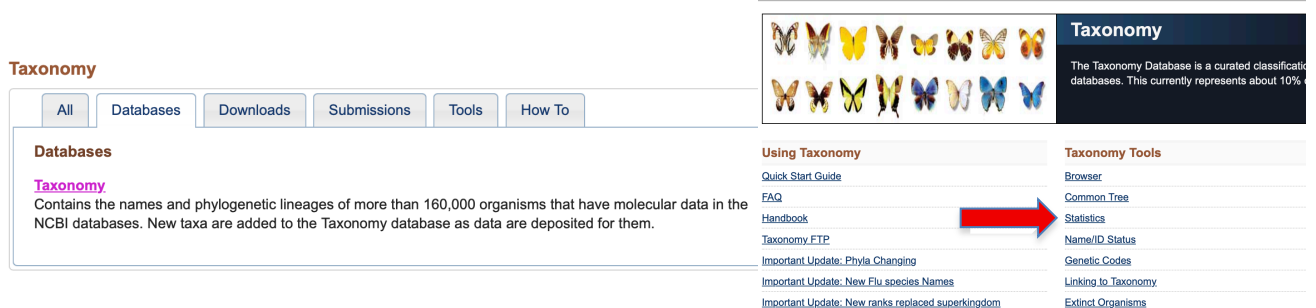
## MODULE 1: Sequence Taxonomy

Objective: The goal of this module is to introduce you to the number and diversity of nucleotide sequences in the NCBI database.

Begin by linking to the NCBI homepage ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). If you ever get lost, always return to this page as a starting point. Select '[Taxonomy](#)' at the bottom of the left menu bar.



The NCBI Taxonomy database contains the names of more than 160,000 organisms whose sequences have been deposited in the NCBI databases. Only a small fraction of the millions of species estimated to exist on earth is represented! Select the '[Taxonomy](#)' link under DATABASES.

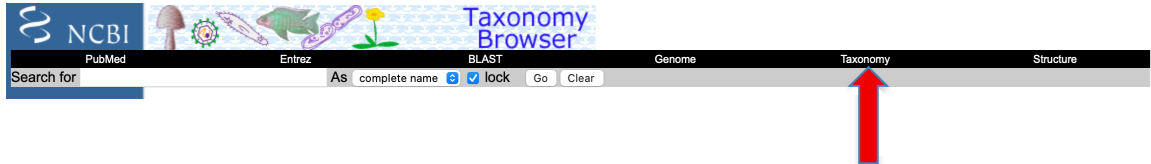


Then select the option '[Taxonomy Statistics](#)' in the middle of the left-side navigation bar.

1. For the 'Taxonomy Nodes (all dates)' column, how many Bacterial Species were in the sequence database? \_\_\_\_\_

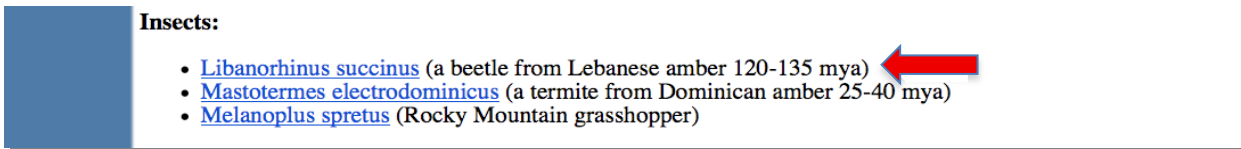
Interestingly, the sequence data from extinct organisms are even listed in the GenBank database. Let's look for a gene sequence from a 120 Mya old insect preserved in amber! From your last website,

- Select the '[Taxonomy](#)' option in the right of the top menu bar

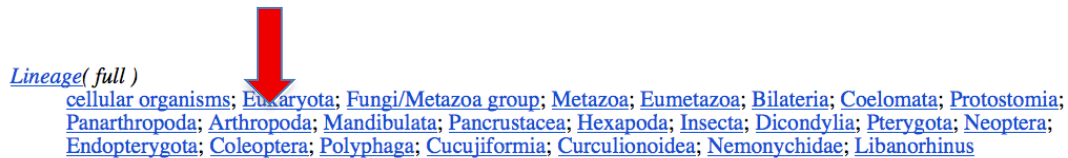


- Select '[Extinct organisms](#)' in the bottom of the middle Column "Taxonomy Tools"

- Scroll down to *Insects* on the main page and select '[Libanorhinus succinus \(a beetle from Lebanese amber 120-135 Mya\)](#)'.



- This page gives you very specific information about the ancestry of this organism. Select the option '[Arthropoda](#)' under *Lineage*.



2. What are some other organisms that belong to this phylum of animals?

\_\_\_\_\_

3. Can you think of any body traits that these organisms have in common?

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

4. Go back one page. How many 'Nucleotide' sequences have been deposited into the Entrez Records from this organism? Look at the box on the top right labeled 'Entrez records' \_\_\_\_\_
5. What is the name of the gene that was sequenced for this organism (to find out, click on the [number 1 next to nucleotide](#))? \_\_\_\_\_ (how does this relate to 16S rRNA) \_\_\_\_\_
6. How many nucleotide base pairs does this DNA entry contain? (the answer is in the first line of the flatfile after you select the Identification link L08072)  
\_\_\_\_\_

Scroll through the complete reference report on this sequence by clicking the blue link with "Leabanorhinus succinus 18S...". A lot of information may seem confusing, but it is all there to provide scientists with as much information as possible about this sequence. This data is formatted into what is called a "flatfile". At the bottom of the screen, you will find the nucleotide sequence (all of the A,T,G,C base pairs) of this gene. Click on the **PUBMED '8505978'** to directly link to the title, authors, and abstract of the published paper! Amazing, now you can read the research article that discovered this nucleotide sequence.

7. What is the title of the research article that published this gene sequence?  
\_\_\_\_\_  
\_\_\_\_\_
8. Go two pages back. Select the '[NCBI link](#)' in the top left corner of the screen (next to the DNA symbol) to return to the NCBI home page. Great! That's where we started with Module 1.

# Bioinformatics Lab

## MODULE 2: Sequence Searching and BLAST

**Objective:** The goal of this module 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/](http://www.ncbi.nlm.nih.gov/)). Select ‘**BLAST**’ in right menu bar under “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



### Popular Resources

PubMed

Bookshelf

PubMed Central

**BLAST** ←

Nucleotide

Genome

SNP

Gene

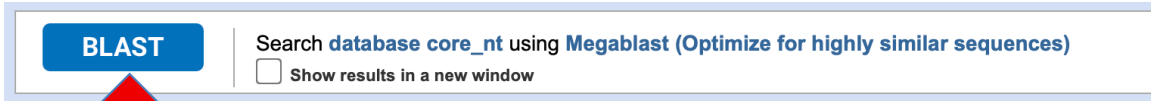
Protein

PubChem

**Input your own, random nucleotides (A,T,G,C) that fill one complete line** in the blank box at the top under “Enter Query Sequence”. Your sequence is referred to as the query sequence.

The image shows a screenshot of the BLAST query input form. At the top, there is a tab labeled "Enter Query Sequence". Below this, there is a text input field with the placeholder text "Enter accession number(s), gi(s), or FASTA sequence(s) ? Clear". A red arrow points to this input field. To the right of the input field, there is a "Query subrange ?" section with "From" and "To" input boxes.

Page Down and Select 'BLAST!' at end of page. A new window appears.



- Wait for the results page to automatically launch. The wait time depends on the type of search you are doing and how many other researchers are using the NCBI website at the same time you are.

1. Did your sequence have any significant similarity to anything in the NCBI databases? How do you determine significance (hint: a significant hit has an E value below E-5 or E raised to the negative 5...a very small number)? If there was no significant similarity, can you offer an explanation why?

---

---

---

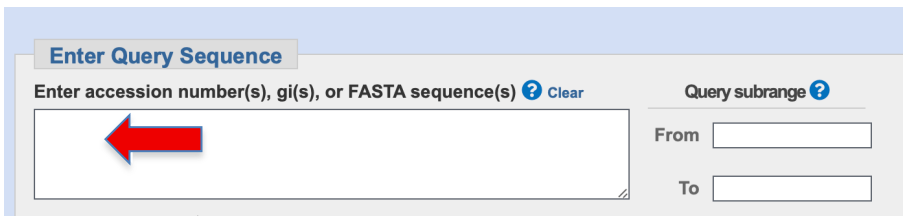
2. What was your E-Value? \_\_\_\_\_

Select Home at the top left of the BLAST page.

Select 'nucleotide BLAST' under the Basic BLAST category



Enter your *Wolbachia* sequence below into the Search box. (At this point in the lab, if students generated their own *Wolbachia* sequences, they could BLAST their own sequence. Here everyone will BLAST the same sequence provided to you below)



Your *Wolbachia* Sequence (either use the new *Wolbachia* sequences from the insects that the class discovered infections in, or if you have no sequences yet, its OK! Use the one provided to you here (*Wolbachia* strain):

```
AGAGTTTGATCCTGGCTCAGAATGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGAGTTAT
ATTGTAGCCTGCTATGGTATAACTTAGTGGCAGACGGGTGAGTAATGTATAGGAATCTACCTAGTAGT
ACGGAATAATTGTTGAAACGGCAACTAATACCGTATACGCCCTACGGGGGAAAAATTTATTGCTATT
AGATGAGCCTATATTAGATTAGCTAGTTGGTGGAGTAATAGCCTACCAAGGCAATGATCTATAGCTGA
TCTGAGAGGATGATCAGCCACACTGGAAGTACGATACGGTCCAGACTCCTACGGGAGGCAGCAGTGGG
GAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCATGAGTGAAGAAGGCCTTTGGGTTGT
AAAGCTCTTTTAGTGAGGAAGATAATGACGGTACTCACAGAAGAAGTCTTGGCTAACTCCGTGCCAGC
AGCCGCGGTAATACGGAGAGGGCTAGCGTTATTCGGAATTATTGGGCGTAAAGGGTGCCTAGGCGGAT
TAGTAAGTTAAAAGTAAAATCCCAAGGCTCAACCTTGGAAATTGCTTTTAAAAC TGCTAATCTAGAGAT
TGAAAGAGGATAGAGGAATTCCTAGTGTAGAGGTGAAATTCGTAAATATTAGGAGGAACACCAGTGGC
GAAGGCGTCTATCTGGTTCAAATCTGACGCTGAGGCGCGAAGGCGTGGGGAGCAAACAGGATTAGATA
CCCTGGTAGTCCACGCTGTAAACGATGAATGTTAAATATGGGAAGTTTACTTTCTGTATTACAGCTA
ACGCGTTAAACATTCGCGCTGGGGACTACGGTCGCAAGATTA AAAACTCAAAGGAATTGACGGGGACCC
GCACAAGCGGTGGAGCATGTGGTTAATTCGATGCAACGCGAAAAACCTTACCACTCCTTGACATGGA
AATTATACCTATTCGAAGGGATAGGGTCGGTTCGGCCGGGTTTCACACAGGTGTTGCATGGCTGTTCGT
CAGCTCGTGTTCGTGAGATGTTGGGTTAAGTCCCGCAACGATCGCAACCCATCCTTAGTTACCATCA
GGTAATGCTGGGGACTTTAAGGAAACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGATGTCAAGTC
ATCATGGCCCTTATGGAGTGGGCTACACACGTGCTACAATGGTGGCTACAATGGGCTGCAAAGTCGCG
AGGCTAAGCTAATCCCTTAAAAGCCATCTCAGTTCGGATTGTACTCTGCAACTCGAGTGCATGAAGTT
GGAATCGCTAGTAATCGTGGATTAGCACGCCACGGTGAATACGTTCTCGGGTCTTGTACACACTGCCC
GTCACGCCATGGGAATTGGTTTCACTCGAAGCTAACGACCTAACCGCAAGGAGGGAGTTATTTAAAGT
GGGATCGGTGACTGGGGTGAAGTCGTAACAAGGTAGCAGTAGGGGAATCTGCAGCTGGATTACCT
```

Select **‘BLAST!’** A new window appears

4. How long (query length) is the *Wolbachia* sequence that you used to search the database? \_\_\_\_\_
5. What is the E-value and Maximum Identity (%) of the best hit (in this case, the first matching sequence)? \_\_\_\_\_ and  
\_\_\_\_\_
6. What is the most likely identity of this sequence? (click on the blue ‘Accession’ link to the left of the top hit, CP003884.1)  
\_\_\_\_\_
7. What is the title of the scientific publication that reported this sequence (click on the PUBMED 23593012 link). Briefly describe in one sentence what this article reported \_\_\_\_\_  
\_\_\_\_\_

---

Go back when you’re done.

Select **Distance tree of results**. This will open a separate page with a phylogenetic tree that includes your sequences (highlighted in yellow with blue dot)

**Print the phylogenetic tree** (if you can print) **and discuss** what the tree tells you about the evolutionary relatedness of your *Wolbachia* strain to other strains in the database. The class might want to create a portfolio of their trees along with a picture and general information on their insects. In particular, what insects are the closely related *Wolbachia* from and are they the same as yours or different? What does this tell you about Horizontal Transmission of *Wolbachia*?

8. What does a phylogenetic tree show? For instance, what does the length and order of the branches tell you about evolutionary relatedness?

---

---

---

---

---

9. What is your strain most closely related to in the phylogenetic tree?

---

---

---

**Return** to the page with your BLAST results that should still be open in another tab.

**Select Home** at the top of the BLAST page.

**Select ‘nucleotide BLAST’** under the Basic BLAST category

**Now enter only the first 25 base pairs of your *Wolbachia* sequence below into the Search box.**

>Your *Wolbachia* Sequence  
AGAGTTTGATCCTGGCTCAGAATGA

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

10. What is the E-value and Maximum Identity (%) of the best hit (the first matching sequence)? \_\_\_\_\_ and \_\_\_\_\_. Is the E-value more or less significant than when you BLASTED the longer *Wolbachia* sequence in step 3? \_\_\_\_\_
11. Is the identity of the best hit different from when you used the complete nucleotide sequence? \_\_\_\_\_
12. From the two BLAST searches you performed, what can you deduce about how the length of a query sequence affects your confidence in the sequence search? \_\_\_\_\_

---

---

## **WEBSITES – Wolbachia Bioinformatics**

NCBI (BLAST)

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Alignment Tool

[https://www.phylogeny.fr/one\\_task.cgi?task\\_type=muscle](https://www.phylogeny.fr/one_task.cgi?task_type=muscle)

Phylogenetic Tree Tool

[https://www.phylogeny.fr/simple\\_phylogeny.cgi](https://www.phylogeny.fr/simple_phylogeny.cgi)