Bean Beetle Microbiome Identification of Picked Colony 16S rRNA DNA Sequence Student Handout

Sanger sequencing is a biochemical method of determining the complete nucleotide sequence of a piece of DNA (see the following video for a detailed description of this process: https://www.youtube.com/watch?v=-QIMkQ4E_wE). The DNA that each of you amplified using PCR with 16S rRNA gene specific primers was sent to a commercial laboratory where they conducted Sanger sequencing on each sample. In 2020, the Sanger sequencing was conducted with a 24-hour turn-around at a cost of $9.00 per sample. Most but not all the samples were successfully sequenced. A data file (called a "FASTA" file, pronounced “fast-A”) for each successfully sequenced sample was sent back to us. The FASTA file contains a string of letters corresponding to the nitrogenous bases in the 16S rRNA gene (DNA) of the bacterium you picked from a culture plate. Conducting a Nucleotide BLAST (BLASTn) (basic local alignment search tool) will permit you to compare your bacterium 16S rRNA gene with sequences in a library of DNA sequences maintained by the National Center for Biotechnology Information, in the National Library of Medicine at the US National Institutes of Health.

BLASTing Sanger Sequences to Determine Identity

1. Open PDF file of your sequence. Assess sequence quality. Determine which part of the sequence is high enough quality to use for downstream analysis.
2. Open the resulting FASTA file in any text editor. This file has the .seq extension. Your computer probably will not recognize to open a .seq file with a text editor, so you can open up the text editor and then open up the file.
3. Copy the part of the sequence that you determined is high quality.
    a. You want to avoid most “N”s, for example, at the start or end of a sequence.
4. Use Nucleotide BLAST (BLASTn) at https://blast.ncbi.nlm.nih.gov/Blast.cgi to compare your sequence to the database of 16s rRNA sequences:
    a. Paste sequence into text box for query sequence.
    b. Under database, select 16S ribosomal RNA database.
    c. Hit BLAST.
5. Look carefully at the top 5-10 hits:
    a. What is the % identity of your sequence and the database sequences?
    b. What % of your sequence aligns with the database sequences?
    c. What is the taxonomic identification of the top 5-10 hits?
    d. Can you assign your sample confidentially to family? To genus? To species?

Generally, an identification to genus is reliable using the 16S V4 region that each of you amplified by PCR. The primers we used are specific for the 16S V4 region of the rRNA DNA of bacteria.