Microbiome Bioinformatics for High Schools

_____

_____

INTRODUCTION

What can 16S rRNA from the gut tell us about one's health?

The human microbiota consists of the 10-100 trillion symbiotic microbial cells harbored by each person, primarily bacteria in the gut; the human microbiome consists of the genes these cells harbor. Microbiome projects worldwide have been launched with the goal of understanding the roles that these symbionts play and their impacts on human health (PMCID: PMC3426293).

The National Center for Biotechnology Information (NCBI) has a lot of bioinformatics resources, including the Basic Local Alignment Search Tool (BLAST). This program lets you very quickly compare a DNA or protein sequence to many (millions) known DNA or protein sequences. The top hit has the highest similarity to the input sequence.

After profiling the guts of Romeo and Juliet by amplifying and sequencing 16S rRNA marker genes, we found lots of different sequences but chose one very common one in each sample. One of them is healthy but the other has lots of abdominal pain. Which person might need to see a doctor?

CLASSROOM HANDS-ON EXERCISE
Requirements: A computer with internet access

1. To start, go to the NCBI page: https://www.ncbi.nlm.nih.gov/
2. Under "Popular Resources" (top right), click on "BLAST".
3. Since we will be comparing DNA sequences to DNA sequences, click on the "Nucleotide BLAST" box.
4. In the box under "Enter accession number(s), gi(s), or FASTA sequence(s)", copy and paste all lines (including the lines beginning with >) in the file "16S_rRNA_sequences_for_Microbiome_Bioinformatics.txt". Note that we have two sequences, one for 16S rRNA from ROMEO's gut and one for 16S_rRNA from JULIET's gut.
   - Sequence file can be found here: https://tinyurl.com/y7vm3dql

>JULIET_2_gut_16S_rRNA
CCTACGGGAGGCAGCAGTGAGGAATATTGGTCAATGGGCGCTAGCCTGAACCAGCCAAGTAGCGTGAAGGATGA
AGGCTCTATGGGGCGTAAACTTCTTTTATATAAGAATAAAGTGCAGTATGTATACTGTTTTGTATGTATTATATGAA
TAAGGATCGGCTAACTCCGTGCCACACCGCGGTAATACGGAAGATCCGAGCGTTATCCGGATTTATTGGGTTTAAA
GGGAGCGTAGGTG
GACTGGTAAGTCAGTTGTGAAAGTTTGCGGCTCAACCGTAAAATTGCAGTTGATACTGTCAGTCTTGAGTACAGTA
AAGGTGGGCGGAATTCGTGGTGTAGCGGTGAAATGCTTAGATATCACGAAAAACTCCGATTGCGAAGGCAGCTC
ACTGGACTGCAACTGACACTGATGCTCGAAAGTGTGGGTATCAAACAGGATTAGATACCCT

>ROMEO_gut_16S_rRNA
CCTACGGGAGGCAGCAGTAGGGAATATTGCTCAATGGGGGAAACCCTGAAGCAGCAACGCCGCGTGGAGGATG
AAGGTTTTAGGATTGTAAACTCCTTTTGTTAGAGAAGATAATGACGGTATCTAACGAATAAGCACCGGCTAACTCC
GTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTACTCGGAATCACTGGGCGTAAAGAGCGCGTAGGCGG
GATAGTCAGTCAGGTGTGAAATCCTATGGCTTAACCATAGAACTGCATTTGAAACTACTATTCTAGAGTGTGGGAG
AGGTAGGTGGAATTCTTGGTGTAGGGGTAAAATCCGTAGAGATCAAGAGGAATACTCATTGCGAAGGCGACCTG
CTGGAACATTACTGACGCTGATTGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCAC

5. Next to "Database ", note that we're going to compare Romeo and Juliet's RNA to a "Nucleotide collection" of many genes in many species.
6. Click on the BLAST button at the bottom left and wait 15-60 seconds for the results to return.
7. When the results page appears, next to "Results for", select Juliet (if not already selected).
8. Scroll below the "Graphic Summary" section and look at the first entry in the "Descriptions" table.
9. What species is the first sequence from?
10. What do you know about that species? Search the Web to find out more.
11. Does Juliet appear to be healthy?

12. Go back to "Results for" at the top of the page and select Romeo.
13. What species is the first sequence from for Romeo?
14. What do you know about that species? Search the Web to find out more.
15. Does Romeo appear to be healthy?