Note to instructors: This protocol uses programs SSTAR v1.1.01, Blast v2.6.0, and MEGA6 which will need to be downloaded from the sources below:
https://github.com/tomdeman-bio/Sequence-Search-Tool-for-Antimicrobial-Resistance-SSTAR-
https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download
http://www.megasoftware.net/

If you wish to use different versions of the software, please run through the protocol prior to the module to ensure version compatibility. Other programs used like RNAmmer and ITOL are web-based and only require internet access.

Final Project Protocol

1) Go to Blackboard. Under the Assignments tab you will find a folder for your group. Download the assembly file (.fna) the SNP file (.fasta) and save them to your desktop
    a. It is important to save the files to your desktop (rather than leaving them in the download folder) so that SSTAR can access them in Step 3.
    b. Also be sure your file names do not include any spaces. For example it should look like "Outbreak1_assembly.fna" NOT "Outbreak1_assembly.fna (1)"
2) First, you will want to identify which bacterial species is involved in your outbreak.
    a. Go to http://www.cbs.dtu.dk/services/RNAmmer/ at the bottom of the webpage there is a submission box.
    b. Click the 'Browse' button and select your assembly file (.fna) and then click 'submit'.
    c. Once your job is finished running, you can click the option to DOWNLOAD PREDICTION RESULTS. Click the FASTA option
    d. This will present you with sequences for a number of genes. Scroll to the bottom of the page and select the sequence for the 16s gene. The 16s gene ID should look something like this:

```
>rRNA_NODE_49_length_1573_cov_102.24_ID_3990_21-1548_DIR+
/molecule=16s_rRNA /score=1946.2
```

    e. Copy/paste the sequence ID and nucleotide sequence into a new text file and save
    f. Go to https://blast.ncbi.nlm.nih.gov/Blast.cgi where you will perform a BLAST search to see what species of bacteria your 16s sequence matches
    g. Click the "Nucleotide BLAST" option and paste your 16s sequence in the query sequence box, then click "BLAST" at the bottom of the page. Do not change any of the settings.
    h. Once your BLAST search concludes, if you scroll down the page you will see the species of bacteria your sequence matches. Write down the species name for your report
    i. You have successfully ID'd your bacteria!
3) Now you will need to identify resistance genes in your outbreak.
    a. Open the program SSTAR_windows
    b. Click "upload your assembly in FASTA format here!" and upload your assembly (.fna)
    c. Click "upload your enzymes in FASTA format here!" and upload the ARG-ANNOT file which should be located in the same folder as the SSTAR program

d. For “Specify cut-off sequence similarity value for new variants” enter “100”
e. Click ”Identify resistance genes!”
f. Once the program is finished, click “Export output files!”. The program automatically exports your outfile to the location of your assembly file.
g. Open the file “yourfilename.txt.genes_tab_separated” in excel (right click ‘Open with’ excel)
h. Examine the genes with 100% sequence similarity. Google the gene name plus the word gene, for example “TetA gene” to learn about the antibiotic resistant genes.
i. Record the names of the genes, what type of antibiotic they are resistant to, and what mechanism of resistance they use (i.e. efflux pump, etc.).
j. Now you know your bacteria ID, the type of antibiotics your bacteria is resistant to, and its mechanism of resistance! You’re already halfway done with the project.

4) You’re done with the genome assembly, time to move on to the SNP matrix to build your tree
   a. Open the program MEGA on your computer (or download at http://www.megasoftware.net/)
   b. Click File -> Open a File/Session then select your SNP matrix file (.fasta)
   c. The program will ask if you want to “analyze” or “align”. Select “analyze” because your file is already aligned.
   d. Your data type is “nucleotide sequences”. Leave the remaining defaults alone and click “ok”. However when it asks if you have “protein-coding nucleotide data” select NO. You are dealing with SNPs so these are not codons which lead to proteins.
   e. Now click the Phylogeny icon and select “Construct/Test Maximum Likelihood Tree”
      i. Yes, you want to use the active dataset
   f. All the defaults are acceptable. Just click “compute” and wait!
   g. Check out your tree. This is your first look at your outbreak. On the tree pop-out window, go to File -> Export Current Tree (Newick) and save your file. This will save your tree in Newick format so you can use it in the next program.
5) You’ve made your phylogenetic tree, time to make it pretty using ITOL.
   a. Go to http://itol.embl.de/
   b. At the top of the webpage you’ll see “Upload” on the toolbar. Click here.
   c. Now you can name your tree. Then under “Tree file”, upload your tree (newick format). Then click “upload”
   d. From this page you can color/edit your tree how you’d like
      i. One option is to left- click a sample or end node, go to Color -> New color range
         1. Now you can select a color and label it. For example labeling it “Transmission” may be helpful
      ii. You can also edit your sample names to correspond with your meta data
         1. Left-click, Label -> Edit
   e. Once you are happy with your tree, go to “Export” on the control panel.
      i. For Format, select an export format (.png is probably easiest)
      ii. For Resolution, change the DPI to 300
      iii. For Export area, select “Full image”

iv. Choose “Colored ranges legend” ON to create a legend
    1. You can also drag and move your tree if it is in the way of your legend
v. Change the Legend title if you’d like
vi. Click “export”

6) Congratulations! You’ve completed the bioinformatics portion of the assignment.