Guide Finder Draft Genome Example Exercise:

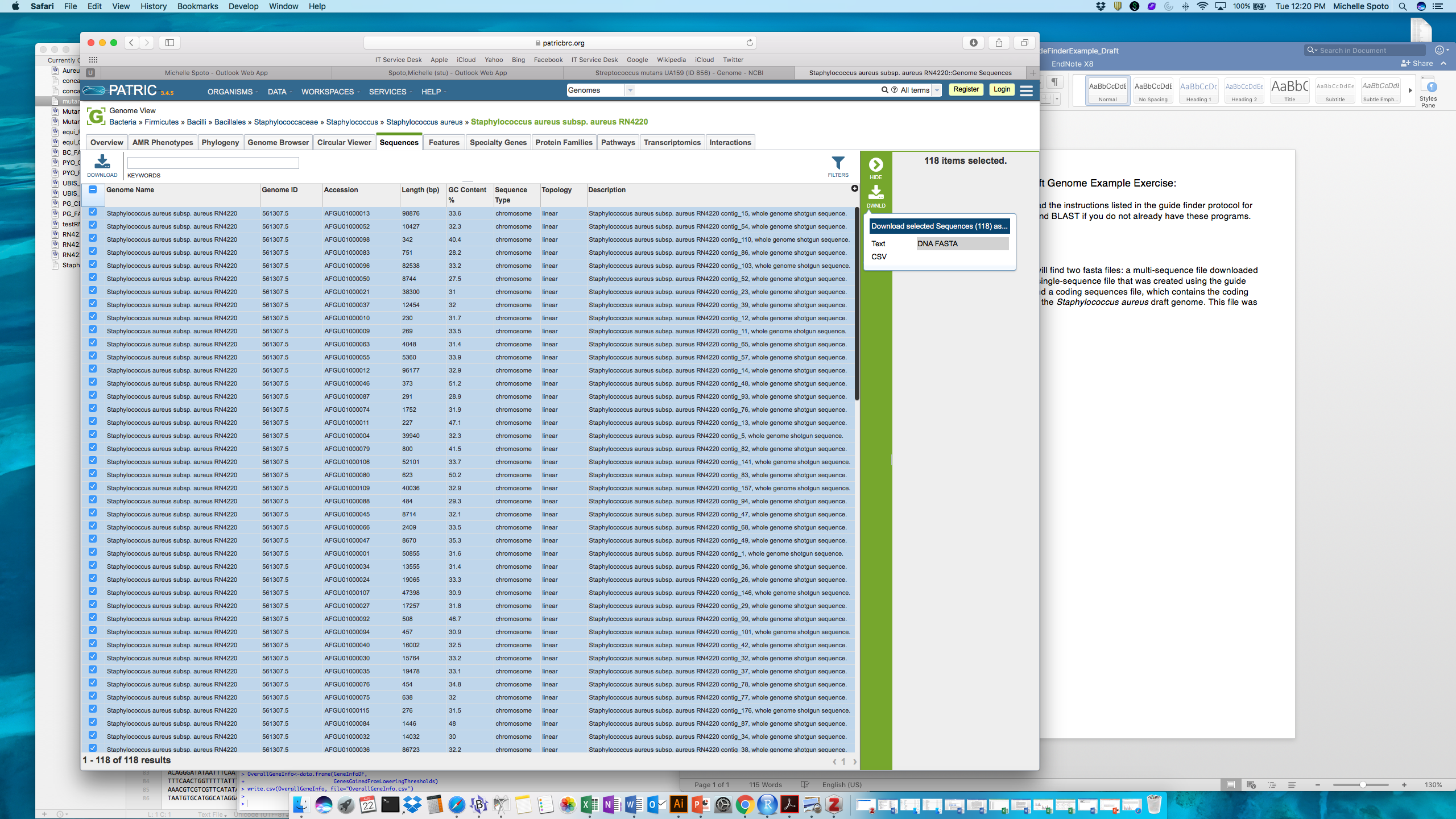
Prior to beginning the example exercise, read the instructions listed in the guide finder protocol for draft genomes. Download the R, RStudio, and BLAST if you do not already have these programs.

1. Identify required files as information

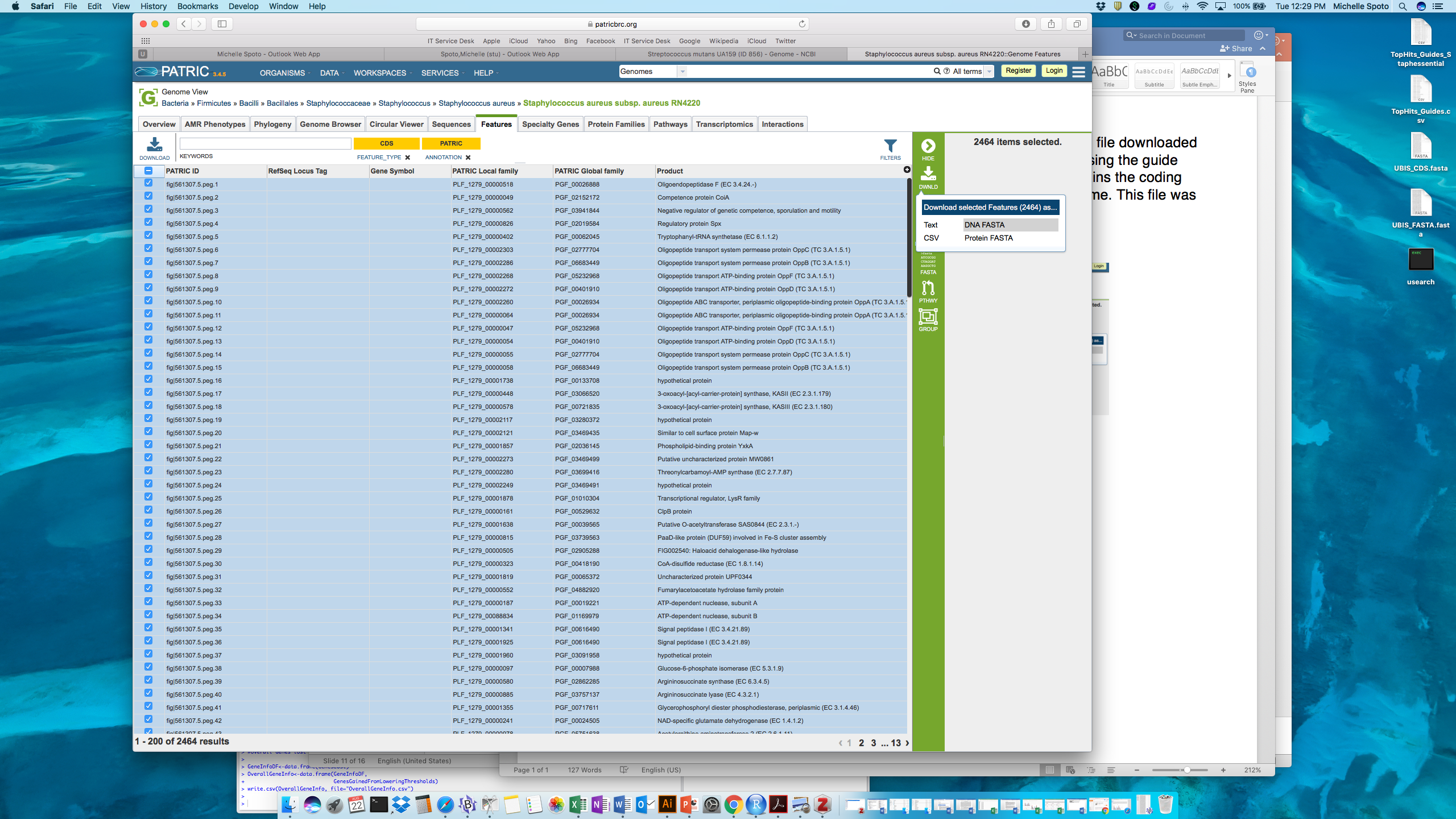
* Multi-sequence fasta file
* Coding sequences file

In the “Draft Genome Example” folder you will find two fasta files: a multi-sequence file downloaded directly from PATRIC and a concatenated, single-sequence file that was created using the guide finder pre-processing script. You will also find a coding sequences file, which contains the coding sequence of each gene (5’🡪 3’ direction) in the *Staphylococcus aureus* draft genome. This file was downloaded directly from PATRIC. The processing for downloading these two files is shown below, for reference.

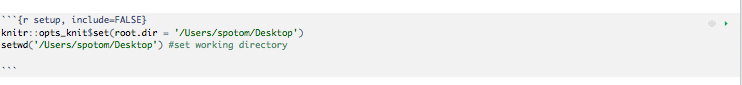
Downloading multi-sequence fasta file from PATRIC:



Downloading coding sequence file from PATRIC:

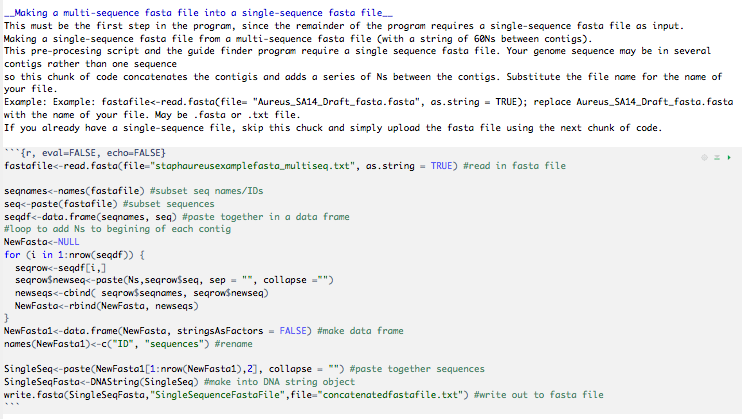


2. Open the pre-processing script with the RStudio program and set the working directory   
The working directory is where all of the files you input into this program should be kept and where all files output by this program will save to. For ease, you can set your working directory to your Desktop, similar to the example shown below, if desired. Set the working directory by identifying the file path to this location. Press the green arrow to run this chunk of code



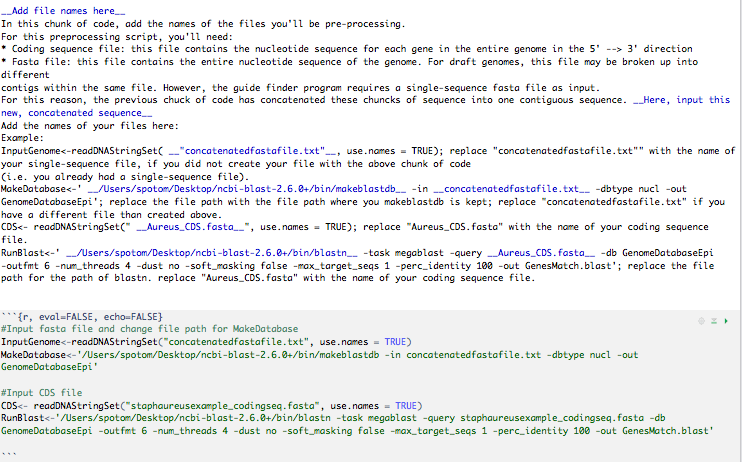
3. Run the next three chunks of code (“Install Packages”, “Load Packages”, and “Functions Created”). You can do this by pressing the green arrow at the top of each chunk. You do not need to make any edits to any of these sections of code. Note that you only need to install the packages once and you only need to load the packages each time you run the pre-processing finder script.

4. Input the mult-sequence fasta file and run the chunk of code to make into a single-sequence file.   
Do this by including the name of your fasta file (staphaureusexamplefasta\_multiseq.txt) in the quotation marks within the read.fasta function (as shown below), if the file name is not already there. Run this chunk of code by pressing the green arrow. Wait until the chunk is entirely done running before proceeding with the next step. The output of this chunk is a file (saved to your working director) called “concatenatedfastafile.txt”, which contains the nucleotide sequence of the input genome concatenated into one sequence, with a series of 60 N’s between each contig. **This single-sequence file will be the working fasta file for the remainder of the pre-processing script and the main guide finder script. It will be saved to your working directory after completion of this step and then re-input into the program in Step 5.**



5. Input your concatenated sequence file (made in Step 4) and the coding sequence file and set the path to BLAST.  
**A**. As shown in the red boxes, set the file path for the location of the makeblastdb file within the BLAST folder downloaded on your machine. In this example, the BLAST folder is located on the Desktop and within that folder is a folder called bin, where both the makeblastdb and blastn files are located. Also, input your fasta file by including the name of the single-sequence fasta file (concatenatedfastafile.txt) in the quotation marks within the readDNAStringSet function and within MakeDatabase (as shown)  
**B.** As shown in the red boxes, set the file path for the location of the blastn file. Also, input the coding sequences file (staphaureusexample\_codingseq.fasta) in the quotation marks within the readDNAStringSet function and within RunBlast (as shown).

Run this chunk of code by pressing the green arrow



6.Run the rest of the code chunks!   
Run each of the code chunks by pressing the green arrow.

7. Identify the output file  
The output file, DraftGenome\_NewCoordinates.csv, contains a list of each individual gene ID, the start and end coordinates for each gene, and the strand on which the gene is coded (+/-). Check this output file against the file provided (aureusDraftExample\_DraftGenome\_NewCoordinates.csv), to make sure that the preprocessing script was run correctly; they should be the same. Also, check the concatenated file (created in Step 4) and check this against the example concatenated file (EXAMPLEconcatenatedfastafile.txt); they should be the same.