|  |
| --- |
| To analyze your PhaNGS data you will need to...   1. Write out the H3 sequences of the phage clones you're looking for, then append sufficient post-H3 sequence or trim your H3 sequence to make a 35bp string of nucleotides. This is your "reference library". 2. Arrange your files and folders for the main script and run.   Assemble your H3 reference library  https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQVVKS1pUAmpkCyi_eo-MVrGqa2zdhFG-5RX-Xs5FPnaOP_PUTPiZW5p0sMost trouble with this step comes from the smallest mistakes in folder names, file names, or script changes. Be very careful when following the instruction and naming things.  Make a folder for processing your data called, for example, "Processing\_1". Then within Processing\_1, create a folder called "library\_reference" and a folder called "sequencing\_runs":  Open an excel document and add the name of your phage in column 1 and the H3 sequence of the phage in column 2 as in this example (no headers). Make sure there is no empty space in either column (this can happen when pasting from online spreadsheets especially):    Save the excel file in "windows formatted text" format. You may get a warning that some formatting might be left out. Just hit continue. It should look like this if opened in a text editor:    Move the exported file (Data\_Step1\_excel\_export) to the library\_reference folder, along with formatting script "0-format-indices.pl". Then, navigate to the folder using Terminal, or similar program.  Type "vi 0-format-indices.pl" to open up the formatting perl script. Press "i", navigate to line 11 using the arrow keys, and replace the file name there with your file name (in this example case, Data\_Step1\_excel\_export.txt).  Press "Esc + : + w + q + Enter" to save the changes. Then, in Terminal, type "./0-format-indices.pl" to run the script. You should see all results get printed in Terminal. A file called "indices.txt" will also appear in the "library\_reference" folder and should be formatted like so:    Your indices are now prepared, and you're ready to move on to actually running the counting script.  Arrange your files and folders for the main script and run  Move the file "1-count-parallel.pl" into the "Processing\_1" folder. Navigate to the folder in terminal, then open up the script by typing "vi 1-count-parallel.pl"    Change the location and name of the indices, sequencing files, processors, and output directory to match your preferences (lines 9-12). Here, we're going to make the output directory "test\_counts". Make sure EVERYTHING in "1-count-parallel.pl" matches all file names. Then press "Esc + : + w + q + Enter" to save the changes.  Move your sequencing files (in .fastq.gz format) to Processing\_1>sequencing\_runs. Our sample sequencing file is "SP05-502...". Here's what everything should look like in the folder before you run:  Navigate terminal to Processing\_1 and type "./1-count-parallel.pl". You should see a message that the file is processing and your output folder has been created. The example file at 150MB should take just a few seconds. With more indices and larger sequencing files this can take several minutes.  Each counts file will be named the same as the ".fastq.gz" sequencing file, but with ".counts" appended to the end. Each counts file should look like this:    Move the script "2-format-counts.pl" into the "test\_counts" folder. Navigate Terminal to the same folder and type "./2-format-counts.pl". This will output several files.  Ignore the genes.txt file. The counts\_table.txt file can be opened in Excel to provide antigens vs. samples in spreadsheet format:    In this case we only have one sample so there's only one column of data. How to work up this data to give something that makes sense is explained in supplementary spreadsheet 1. |