This FAQ is geared toward data analysis using <https://github.com/KyleLevi/bam_scripts>. The repository contains a makefile so that students only need to type “make \_\_\_\_\_\_\_” to analyze their data.

**What are these ARG names? What is AAG07064.1?**

You will ] need to look up your ARG names in an online protein database, such as Uniprot.org.

**I did a Protein scan and have lots of .m8 files in Input/RAP\_Results/ what do I do?**

There are two different commands at your disposal:

1. make protein\_hits - This will read each .m8 file and count the number of times each protein gets a match in each file. The resulting .csv file will be in Output/protein\_hits.csv
2. make protein\_stats – This will read every .m8 file and group together hits by protein. Hits for each protein are grouped and, for each protein, a file will be written in in the output folder as a CSV file. This will give you a detailed look at the variation of individual amino acids within a protein.

**I did a Nucleotide scan and have lots of .bam files in Input/BAM\_Files/ what do I do?**

There are two different commands at your disposal:

1. make sam\_hits - This will read each .bam file and count the number of times each organism has a match in each file. The resulting .csv file will be in Output/BAM\_hits.csv
2. make sam\_stats – This will read every .bam file and group together hits by organism. IF THERE ARE MULTIPLE ORGANISMS PRESENT IT WILL ASK FOR AN ORGANISM NAME. It will display available names; type the name of the organism you are interested in and press Enter. It will create a detailed CSV file of matches to each position in the genome and write a file to Output/genome\_name.csv

**How do I download all of these CSV files so I can open them in a spreadsheet program?**

FROM A TERMINAL ON YOUR COMPUTER (before you ssh into your Jetstream IP):

scp username@jet.stream.ip:~/bam\_scripts/Output/\* path/on/your/computer/

There are a few things to know about this command:

1. It is case sensitive, EX: bam\_scripts vs BAM\_Scripts
2. path/on/your/computer will vary by user. If you are on windows using ubuntu, it is probably something like: /mnt/c/Users/kylelevi/Desktop/

**My scan is slow/failed, what should I do**

Start multiple instances and have each of them scan for a portion of your SRA Run IDs (in the file SraAccList.txt). To setup a fresh instance, follow these steps after ssh username@jetstreamip:

1. git clone https://github.com/kylelevi/bam\_scripts
2. cd bam\_scripts
3. make setup
4. (If you are doing a protein scan) make diamond\_install
5. Add your .fasta files to the Input/Genomes/ or Input/Proteins/ folder depending on your type of scan
6. Add your SRA Run IDS to Input/SraAccList.txt

**Help!**

Email me