Instructions for using MOTHUR

* MOTHUR has to be downloaded for your specific from github
* For help in regards to commands: https://www.mothur.org/wiki/Main\_Page
* Save all files obtained in the Mothur folder generated after download. Since there are a lot of files after the analysis is done move them to a unique folder.

When writing command line do not use spaces unless there is a \_ present. The names have to be highly specific.

* The first item is to generate a **distance matrix file**. This file will calculate how similar the sequences are from one another. For this you need to have an alignment file from MEGA saved as a .fasta
  + Open up the terminal
  + Type dist.seqs(fasta=\*thenameofyourfile\*,\_ output=lt)
    - Lt it a lower triangle matrix
  + What will be saved is a phylp.dis file in your Mothur folder
* Next is to generate a **cluster file**.
  + In the terminal type cluster(phylp=\*thenameofyourfile\*,\_precision=1000,\_cutoff=0.05)
    - The precision is how many decimals it will go to. For sequences highly 1000 is the best because it will give values to 3 decimal places.
    - The cutoff are the % differences between the clusters. 0.05 is 5% difference.
      * Ex. For carP it is estimated by 5% differences there will be only 1 group because they are highly similar.
* This is will generate a **.list file**. This file is in a basic .txt file that can be copied and pasted into an excel file.
  + The .list file has all of the name separated into their individual clusters. The file will tell you how many clusters there are per division.
    - Ex. For 0.05% differences in sequences you could have 100 clusters, however for 1% differences you may only have 10 clusters. At one % difference you might see only 1 cluster because all sequences are within a certain % difference.
    - We will need to look at this file and determine what % difference we will use.
    - Once we decide which cluster to use copy this line from the .txt file not the excel file and paste into another .txt file. \*\*\*Save this as a unique name with the clustering %
* Now that we have the clusters and have decided which ones to move forward with we will have to retrieve all the sequences and put them into clusters.
  + In the command window type get.seqs(accnos=\*thenameofyourfile.txt\*,\_\*thenameofyourfile.fast\*)
    - The first file name is the .txt file we generated in the previous step and the second name is the .fasta file for your alignment.
  + This will give you a pick.fasta file that can be used in MEGA to generate large trees.
* In order to have readable trees, we might want to only align 1 representative from each cluster and can have Mothur pull these sequences. To generate a **representative file** 
  + In the command line type getoturep(phylip=\*thenameofyourdistancefile\*,\_list=\*thenameofyourfile.txt\*,\_fasta=\*thenameofyourfile.fasta\*)
  + The last two files are the ones used in the previous commands.
  + This can be opened in MEGA and used to generate trees