**16S QC and Pipeline Run Validation Checklist**

(Version: Feb 10, 2020)

Confirm directory consistency:

1. Are inputs and outputs in the standard location (<project directory>/Sequences)?

Confirm raw sequence data:

1. Are the input fastq individual files in the Sequences/Raw directory?
2. Does each sample have 3 files: index primer “I1”, Forward “R1” and Reverse Read “R2”?
3. Are the R1/R2 file sizes generally greater than 500kB and of variable sizes?
4. How many samples are there? (count number of R1 or R2 files)

Confirm QC Results:

1. Is there a directory for each sample in the QV\_\* directory?
2. Are there fastq\_qc\_log’s and timing\_log’s?
3. Spot check some fastq\_qc\_log’s
   1. Do you see entries for both Forward, Reverse for the steps specified in run?
   2. Do you see a final paired/merged entry?
4. Across all the fastq\_qc\_log’s “grep” out the “raw” entries and “paired merged” entries. Do the number of entries match? Was there a substantial drop in the number of reads between the raw and QC’d samples?
5. Concatentate the fastq\_qc\_log’s and attach to email.

Confirm Merge:

1. In the QV\_\*/MergeMates/screened\_fasta directory, do you see a single file for each sample?
2. Are the file sizes generally greater than 100kB and of variable sizes?

Confirm Mothur Run:

1. Are the Mothur results in the <project directory>/Mothur\_\* directory?
2. In Mothur\_\*/Results directory, do you see steps 01\_unique.seqs through 13\_classify.otu?
3. Do you see the Summary\_Tables directory?
4. Do you see summary tables for the 6 taxonomic levels (kingdom down to genus) and the OTUs?
5. Were the mitochrondria/chloroplasts filtered?
6. Look at the first two columns from the genus.summary\_table.tsv (cut -f 1,2) for the filtered file. Do the number of samples match the original fastq files? E-mail file with these two columns.